

Alexander Leichtle¹

Neue «-omics»-Technologien und die Herausforderungen der Dateninterpretation

Die Labormedizin hatte sich in den letzten Jahrzehnten einem rasanten Automatisierungs- und Konsolidierungsprozess unterworfen: Die bekannten Analyten konnten immer günstiger und einfacher, aber auch immer schneller und genauer – manchmal vielleicht sogar über den klinischen Bedarf hinaus – gemessen werden. Die Menge an neu hinzukommenden Parametern war überschaubar, jeder Analyt konnte mit einer Handvoll analytischer Techniken gemessen werden und musste im Rahmen der klinischen Einführung gründlichst auf seine diagnostischen Leistungsdaten hin evaluiert werden – eine grosse Fülle an entsprechenden Publikationen zeigt das bis heute. Die Übermittlung von Resultaten erfolgte und erfolgt üblicherweise immer noch als Konzentration oder Aktivität mit einem entsprechenden Referenzbereich, der z.B. im Rahmen einer Akkreditierung auf Lückenlosigkeit sorgfältig geprüft wird. Unsere klinisch tätigen Kollegen haben im Laufe der Jahre mehr oder weniger gelernt, mit diesen Werten umzugehen, sie ggf. sinnvoll zu kombinieren und Diagnosen daraus abzuleiten – der Labormediziner hat sich häufig auf eine Rolle als «Messknecht» zurückgezogen. Diese zugegebenermassen bequeme Situation hat zu einem schleichenden Bedeutungsverlust der Labormedizin insbesondere im klinischen Umfeld geführt. Mit den neuen «-omics»-Technologien besteht die Chance, verlorenes Terrain zurückzugewinnen.

Drei Hauptaspekte bedingen die besonderen Herausforderungen im Umgang mit «-omics»-Daten. Die erste Novität ist die enorme Multiparametrisität. Während wir auch heute noch im klinischen Alltag bei 1000 Patienten pro Tag durchschnittlich 5 bis 8 Messwerte generieren, hat sich das Verhält-

noch kaum berechenbarer Weise mit z.B. lebensstilassoziierten Kovariaten kombinieren und kumulieren. So ist gegenwärtig eine Prädiktion z.B. eines Risikos allein aufgrund einer genomweiten Untersuchung – abgesehen von monogenetischen Erkrankungen – extrem schwierig und eine besondere Expertise erforderlich. In diesem Lichte sind auch kommerzielle Angebote, bei denen potentielle Patienten selbst Proben nehmen und zu Untersuchungen einsenden können, zumindest kritisch zu betrachten. Die Beantwortung der Frage «welches Risiko bedeutet es für mich, wenn ich den SNP xyz in meinem Genom trage?» sollte in kompetenten Händen, z.B. eines Labormediziners oder Humangenetikers liegen. In Zukunft wer-

den solche Fragen durch den sinkenden Preis und einfachere Verfügbarkeit immer drängender werden.

Messung von Replikaten

Die nächste, eher bioinformatische Herausforderung stellt die Messunsicherheit in Verbindung mit der Messung von Replikaten dar. Während in der «klassischen» Labormedizin die Methode für jeden Parameter kalibriert wird, entsprechende Kontrollen gemessen werden und letztendlich die Messwerte ohne weiteren Kommentar dazu auf den Befunden erscheinen, war man sich in den «-omics»-Technologien von Beginn an bewusst, dass die analytischen Methoden bestenfalls näherungsweise den «wahren» Wert messen. Man hat

Die besondere Schwierigkeit liegt in der Tatsache, dass Zusammenhänge häufig weder direkt noch linear sind und damit unser Vorstellungsvermögen übersteigen.

nis z.B. bei genomweiten Untersuchungen häufig umgekehrt. «Signifikante» Ergebnisse entstehen allein schon aus der Unmenge an gemessenen Parametern, so dass eine Korrektur für multiples Testen zwingend erforderlich und inzwischen auch Standard geworden ist. Dass nach einer solchen Korrektur nicht selten kaum noch signifikante Effekte übrig bleiben, mag nicht nur an der beliebten, aber sehr konservativen Bonferroni-Korrektur liegen. Vielmehr lassen sich, insbesondere bei multifaktoriellen Erkrankungen wie z.B. der koronaren Herzkrankheit oder der Entwicklung des Diabetes, sehr schwache genetische Einzeleffekte sich in derzeit

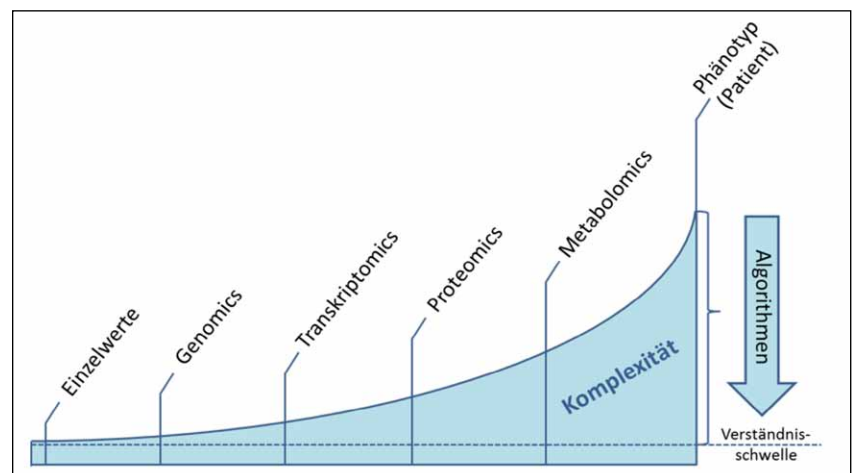


Abb. 1: Die Komplexität der Resultate der neuen «-omics»-Technologien nimmt stetig zu. Geeignete Algorithmen und bioinformatische Modelle können helfen, die Datenmenge auf die für die Fragestellung relevanten Informationen zu reduzieren.

¹ Dr. med. Alexander Leichtle, Universitätsinstitut für Klinische Chemie, Inselspital Bern

Les nouvelles technologies «-omiques» et les enjeux de l'interprétation des données

L'évolution des technologies «-omiques» et leur utilisation en médecine ont suscité d'importants espoirs en ce qui concerne les possibilités diagnostiques, l'évaluation des risques, le contrôle thérapeutique et le pronostic de maladies jusqu'à présent difficiles à diagnostiquer. Pourtant, avec chaque nouveau membre dans la famille sans cesse en expansion des «-omiques» s'accroissent également les exigences relatives au traitement de données toujours plus complexes. Alors qu'au début de l'engouement pour les «-omiques», les difficultés qui occupaient le devant de la scène étaient avant tout d'ordre technique et analytique, ces dernières années, l'attention s'est davantage focalisée sur la gestion de quantités de données jusqu'à présent inconnues. Et le développement ne s'arrête pas là: du matériel informatique qui devient toujours plus performant, les grilles et les nuages informatiques qui, entre-temps, viennent à bout des gigaoctets, lesquels peuvent par exemple être produits pour chaque échantillon au cours d'une expérience métabolomique. Le prochain obstacle à surmonter est encore bien plus important: trouver l'essence de l'information qui soit pertinente pour le questionnement diagnostique sous-jacent au sein de la profusion existante de résultats de mesures, qui ne constituent généralement qu'un instantané de processus interactifs de construction, de transformation et de dégradation extrêmement complexes.

deshalb versucht, dieses Problem durch die Messung von biologischen und technischen Replikaten anzugehen. Während aber Replikate selbst nicht schwerer zu messen sind als Einzelproben, stellt ihre adäquate Implementierung in die Auswertungsalgorithmen eine nicht zu unterschätzende Hürde dar. Gern wird vergessen, dass Replikate keine unabhängigen Messungen darstellen und deshalb auch nicht so behandelt werden dürfen. Häufig behilft man sich damit, dass man Mittelwert oder Median bestimmt und damit weiterrechnet. Dabei geht aber ein erheblicher Anteil an Information verloren, der mit geeigneten Methoden erfasst werden könnte – wenn es sie denn gäbe!

Erste Statistikpakete sind inzwischen verfügbar, z.B. für die Berechnung von Korrelationskoeffizienten in Messungen mit Replikaten [1]. Da die Anforderungen in vielen «-omics»-Subdisziplinen häufig ähnlich sind, lassen sich solche Pakete oft auch «off label» verwenden, dann allerdings mit einem nicht unerheblichen Aufwand. Auf der anderen Seite kann auch die «klassische» Labormedizin von diesen Entwicklungen profitieren, den Umgang mit der in jeder Messung impliziten Unsicherheit neu definieren und dem Kliniker durch deren Mitteilung eine rationale(re) Entscheidung ermöglichen, als es bisher durch «exakte» Einzelwerte und «harte» Cut-offs möglich ist.

Definition unterschiedlicher Variablen und deren gegenseitige Abhängigkeit

Die bei weitem grösste Hürde und Herausforderung besteht jedoch in der Erschliessung der «mutual information», d.h. der Information, die im Zusammenhang unterschiedlicher Variablen liegt und deren gegenseitige Abhängigkeit definiert. Solche Abhängigkeiten sind in der Labormedizin seit langem bekannt, denkt man z.B. an die Regelkreise der Schilddrüsenfunktion. Mit dem Aufkommen der «-omics»-Technologien und der Untersuchung komplexer Systeme («-ome») «als Ganzes» hat die «mutual information» jedoch enorm an Bedeutung gewonnen, z.B. in der Entschlüsselung der Funktion von «non-coding regions» im Genom, der Expressionsregulation im Transkriptom, dem Nachweis posttranslationaler Modifikationen und Protease-vermittelten Decays im Proteo-/Peptidom und beim Verständnis von Metabolitveränderungen entlang von Stoffwechselfaden im Metabolom. Die besondere Schwierigkeit darin liegt in der Tatsache begründet, dass die Zusammenhänge häufig weder direkt noch linear sind und damit in der Regel unser Vorstellungsvermögen übersteigen. Wir müssen uns bewusst werden, dass wir uns Korrelationen über mehr als zwei Zwischenschritte oder Interaktionen mit mehr als einer Handvoll von Variablen nicht mehr vorstellen können. Jeder Versuch, derartige Informationen mit den üblichen «linearen» Verfahren darzustellen, ist

a priori unvollständig und damit unzulänglich. Auch hilft die Fokussierung auf «signifikante Unterschiede» hier nicht weiter, weil die eigentliche Information nicht in den absoluten Konzentrationsunterschieden, sondern in ihrer gegenseitigen und oft mittelbaren Bedingung liegt. Den häufig relativ einfachen diagnostischen Fragen (z.B. «wie hoch ist die Tumorstärke?») stehen also grosse Sets an miteinander in Beziehung stehenden Variablen gegenüber, aus denen die für die jeweilige Fragestellung relevanten Daten z. B. mittels Regressionsmodellen extrahiert werden müssen, wobei es – abhängig von der Anzahl der untersuchten Variablen – durchaus vorkommen kann, dass ganz unterschiedliche Kombinationen von Werten die gleichen diagnostischen Aussagen liefern können. In solchen Situationen kann die Kombination der Modellbildung mit bayesischer Statistik [2] dazu beitragen, geeignete Modellkombinationen zu berechnen, deren Aussagekraft sich ergänzt bzw. kumuliert [3]. Es ist im Prinzip nichts anderes als die medizinische Beurteilung von herkömmlichen Befundkonstellationen durch den erfahrenen (Labor-)Arzt – aber im grossen Massstab und weniger auf subjektive Einschätzung als auf die vorliegenden Daten gestützt. Die grosse Chance, die sich der Medizin hier bietet, ist die «Rückübertragung» dieser aus der Beschäftigung mit den «-omics»-Technologien gewonnenen Methoden und Erkenntnisse auf die tägliche Routine, damit eine verbesserte Interpretation auch ganz alltäglicher Befunde und letztendlich ein Mehrwert für den Patienten.

Korrespondenz:
Alexander.Leichtle@insel.ch

Literatur

- 1 Zhu D, Li Y, Li H, Zhu D, Li Y, Li H. Multivariate correlation estimator for inferring functional relationships from replicated genome-wide data. *Bioinformatics*. Oxford University Press; 2007;23(17):2298–2305.
- 2 Yeung KY, Bumgarner RE, Raftery AE. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics*. Oxford University Press; 2005May12;21(10):2394–2402.
- 3 Savoca R. Interpretation von Laborresultaten – Bayes oder Bauchgefühl? *Pipette* 2013 Jun.;(3):10–11.