

Gen-Sequenzierung und Expressionsprofile in Diagnostik und Prognose

H. Lehrach

MPI für molekulare Genetik, Berlin

Korrespondenz an: Prof. Dr. Hans Lehrach, Max-Planck-Institut für molekulare Genetik, Ihnestraße 73, D-14195 Berlin

Eingegangen: 20. September 2007
Online First 20 February 2008

Schlüsselwörter: Gen-Sequenzierung, Expressionsprofile, Diagnostik.

Alle Lebensprozesse können als eine Art ‚Rechenprozess‘ aufgefasst werden, in dem die Information im Genom unter dem Einfluss der Umwelt in den Phänotyp des Organismus umgesetzt wird. Dieser Prozess ist zu einem vielleicht überraschenden Grad deterministisch, wie man zum Beispiel an der großen Ähnlichkeit zwischen eineiigen Zwillingen oder Mäusen aus Inzuchtstämmen sehen kann.

Krankheiten sind Störungen dieses ‚Rechenprozesses‘. Um viele Krankheiten besser vermeiden oder – nach Ausbruch der Krankheit – besser therapieren zu können, wird es notwendig sein, das Verhalten solcher komplexen Netzwerke, zum Beispiel unter bestimmten Therapieansätzen, besser voraussagen zu können. Dazu wird es unserer Ansicht nach notwendig sein, prädiktive Computermodelle zu entwickeln, die imstande sind, individuelle Informationen über den Kranken und seine Krankheit zu liefern, wie wir sie zum Beispiel durch Analysen des Genoms, des Epigenoms und des Transkriptoms des Patienten beziehungsweise des erkrankten Gewebes erhalten können.

Um solche prädiktiven Modelle entwickeln zu können, brauchen wir einerseits weitgehende Informationen über die Komponenten dieser Netzwerke, ihre Expression, ihre Interaktionen und die Funktion dieser Komponenten innerhalb der Struktur dieser Netzwerke.

Die Sequenzierung des menschlichen Genoms sowie der Genome vieler Modellorganismen hat uns erstmals eine Übersicht über die möglichen Komponenten solcher Netzwerke geliefert. Neue Sequenzierverfahren (‚second generation sequencing techniques‘) versetzen uns dabei immer mehr in die Lage, auch große Teile des Genoms einzelner Individuen zu sequenzieren, eine Möglichkeit, die in Zukunft enorme Auswirkungen auf die medizinische Diagnostik haben könnte.

Durch die Entwicklung der neuen Sequenzierertechnologie, wie z. B. dem Illumina Genome Analyzer System (Solexa.) oder dem 454 Genome Sequencer FLX System (Roche Applied Sci-

ence), ist es möglich, in relativ kurzem Zeitraum mit hoher Empfindlichkeit und zu sehr geringen Kosten z. B. klinische Proben von Tumoren oder einer ganzen Reihe anderer genetisch bedingter Krankheiten in sehr großen Patientenzahlen zu sequenzieren und zu analysieren. Beide Instrumente stehen am MPI für molekulare Genetik zur Verfügung und werden für verschiedene Forschungsprojekte eingesetzt. Das Roche/454 GS FLX System erzeugt pro Lauf etwa 100–120 MB an Rohdaten, das entspricht ca. 500.000 „Reads“ mit einer Leselänge von etwa 200–250 Basen, wobei die Lesegenauigkeit bei ca. 99,5% liegt. Das GS FLX System ist hervorragend geeignet zur Analyse vollständiger Transkriptome, Neusequenzierung von bakteriellen und kleinen eukaryotischen Genomen. Das Illumina Genome Analyzer System erlaubt die Sequenzierung zu erheblich günstigeren Kosten, ein Lauf dauert etwa 3 Tage und erzeugt 5 Mio. „Reads“ von einer Leselänge von ca. 25–35 Basen. Der gesamte Durchsatz des Systems ist ungefähr ein Gigabase an Rohdaten. Das Solexa System ist zurzeit stark in Benutzung für unterschiedliche Anwendungen, wie Re-sequenzierung, Transkriptom Analyse, Genotypisierung oder Analyse von MicroRNAs.

Da Transkripte und Proteine nur wirken können, wenn sie auch aus dem Genom ausgelesen werden, spielen Verfahren zur Analyse der Muster der Genexpression eine wichtige Rolle, um die Funktion der einzelnen Gene abschätzen zu können. Dazu können einerseits (Chip oder Sequenz-basierte) Verfahren eingesetzt werden, um die Expression aller Gene in einem bestimmten Gewebe zu verfolgen. Diese Information wird komplementiert durch hochauflösende, *in situ* Hybridisierungstechniken, wie sie zum Beispiel von uns im Rahmen des EUrepress Projektes eingesetzt werden.

Um die Muster der zell- und entwicklungspezifischen Genexpression besser zu verstehen, setzen wir eine Reihe von Methoden ein, um systematisch die Mechanismen der Genregulation entschlüsseln zu können. Solche Informationen können aus Daten aus der Literatur beziehungsweise den daraus entwickelten Signalweg-Datenbanken gewonnen werden. Darüber hinaus entwickeln wir Ansätze, um syste-

matisch Informationen über Regulationsprozesse in Vertebraten zu gewinnen. So verwenden wir unter anderem eine Kombination verschiedener molekularer und genetischer Verfahren zur Entschlüsselung der Genregulationsnetzwerke, die in bestimmten menschlichen Zelllinien (HEK 293 und andere) aktiv sind. Dabei setzen wir die selektive Inhibition von Transkriptionsfaktorgenen durch RNAi ein, um systematisch Informationen über die direkten und indirekten Zielgene dieser Faktoren zu erhalten. Diese Informationen werden komplementiert durch die direkte Analyse der Bindungsstellen der Transkriptionsfaktoren durch ChIP-chip und ChIP-seq Verfahren (Analyse der Transkriptionsfaktorbindungsstellen nach Chromatinimmunpräzipitation durch Chip oder Sequenzierungsanalysen), durch die Analyse der Evolution der Promotersequenzen in verschiedenen Affenspezies (,evolutionary shadowing') sowie die direkte Analyse der Aktivität von Promoterkonstrukten in ,cell arrays'.

Wichtige Informationen zur Entschlüsselung der regulatorischen Netzwerke des Menschen können jedoch auch über genetische Untersuchungen in Modellorganismen, besonders der Maus, erhalten werden. So verwenden wir ein neues genetisches System, einen Satz von Chromosomenaustauschstämmen, die Jiri Forejt (Institute of Molecular Genetics, AS CR, Prag, Tschechische Republik) durch Austausch einzelner Chromosomen des (*Mus musculus domesticus*) Inzuchtstammes C57/Bl-6 gegen die entsprechenden Chromosomen des (*Mus musculus musculus*) Stammes PWD konstruiert hat. Änderungen in der Methylierung, der Transkription, der Prozessierung, der Proteinmenge oder der Proteinmodifikation von Genen / Genprodukten, die nicht auf dem ausgetauschten Chromosom liegen, beweisen, dass auf dem ausgetauschten Chromosom Gene liegen, die diese Prozesse in ,trans' steuern. Falls es sich hier um ein einzelnes Gen handelt, sollten wir imstande sein, durch entsprechende Kreuzungen dieses Gen auf dem Chromosom zu lokalisieren und so in vielen Fällen das für die Änderung verantwortliche Gen zu identifizieren.

Aufbauend auf dem Verständnis der regulatorischen Netzwerke im gesunden Organismus wird es wichtig sein, über eine detaillierte Analyse und Modellierung der Störungen dieser Netzwerke bei Erkrankungen die Basis für eine

Systembiologie dieses Krankheitsbereiches zu entwickeln, um so die Grundlage für verbesserte Präventions- und Therapiestrategien zu schaffen. Systembiologische Methoden und Computerprogramme für die Modellierung und Vorhersagen von physiologischen Prozessen in komplexen biologischen Systemen und Vorhersagen zu möglichen Folgen können bei der Prognose, Diagnose und Therapie vieler Erkrankungen helfen. Am MPI für molekulare Genetik haben wir das Modellierungs- und Simulationssystem PyBioS entwickelt. Es bietet eine Web-basierte Umgebung für weit reichende Anwendungen in der Systembiologie. Es unterstützt die Konstruktion, Simulation, Visualisierung, Analyse und Speicherung großer Computermodelle zellulärer Systeme. Für die Erstellung derartiger Modelle ist eine Schnittstelle zu externen Datenbanken wie KEGG und Reactome vorhanden sowie für die Integration experimenteller Daten aus Hochdurchsatz-Experimenten, z.B. von Mikroarrays. Ferner verfügt PyBioS über weitere Werkzeuge zur Modellanalyse, Identifizierung von stationären Zuständen einschließlich einer Stabilitätsanalyse sowie Algorithmen zum "Parameter scanning" und für die metabolische Kontrollanalyse. Mit Hilfe von PyBioS haben wir Computermodelle Krebs-relevanter Signalwege bzw. der Somitenbildung während der Embryogenese von Vertebraten entwickelt. Neben der Modelltopologie werden für die Entwicklung eines detaillierten kinetischen Computermodells auch Informationen über die Reaktionspartner und deren Kinetiken benötigt. Da solche Informationen oft nicht zur Verfügung stehen, verwenden wir einen so genannten Monte-Carlo Ansatz, bei dem verschiedene biologische Szenarien simuliert werden. Durch mehrfache Wiederholung der Simulation mit veränderten Werten für unbekannte kinetische Parameter und Startkonzentrationen kann das Verhalten der Modelle, das verschiedene biologische Szenarien reflektiert, untersucht werden. Ferner wird es notwendig sein, in Patientenproben (z. B. Biopsien aus Tumor, Fettgewebe etc.) ausreichende Informationen zu gewinnen, um die Modelle an die Situation im Patienten anzupassen und einen wichtigen Beitrag für die Prognose / Diagnose, Therapieansätze und Medikamentenentwicklung liefern zu können.

To access this journal online:
<http://www.birkhauser.ch/JVL>
