

Special DNA- & RNA-Sequenzierung – Deep Sequencing in Freiburg

Epigenom-Analyse

■ Die Bioinformatics & Deep-Sequencing-Gruppe um Thomas Manke sequenziert Millionen verschiedener DNA-Moleküle gleichzeitig, analysiert die Ergebnisse und interpretiert sie für die Wissenschaftlerkollegen – ein schöner Service!

In einer Ära von personalisierter Medizin und synthetischen Genomen sind hocheffiziente Sequenzieretechniken vonnöten, mit welchen sich schnell und unkompliziert biologische Funktionen und phänotypische Unterschiede analysieren lassen. Dabei gilt es nicht nur, einfach eine Nukleotidsequenz abzulesen. Die heute üblichen Verfahren sind meist Weiterentwicklungen der Didesoxy-Methode nach Sanger und werden Next Generation Sequencing oder NextGen genannt.

Versteckte Informationen

Auch das Max-Planck-Institut für Immunologie und Epigenetik (MPI-IE) in Freiburg betreibt seit 2010 eine „Bioinformatics and Deep Sequencing-Unit“. Leiter der Einrichtung ist Thomas Manke, der nach der Promotion in Theoretischer Physik seine Begeisterung für die Genforschung und Bioinformatik entdeckte. Schon als Postdoc am MPI für Molekulare Genetik (MPI-MG) in Berlin entwickelte er numerische Methoden, um versteckte Informationen in Genomsequenzen zu entziffern.

Die moderne Forschung zeigt allerdings, dass zelluläre Informationsflüsse nicht nur durch die DNA kontrolliert werden – Umwelteinflüsse können biologische Programme modulieren und die

Organisation und Funktion des Genoms dauerhaft verändern. Die Erforschung dieser Prozesse ist Aufgabe der Epigenetik. Sie untersucht die molekularen Veränderungen der DNA – das Epigenom –, die mit den neuen Sequenziermethoden bestimmt werden können. Dazu zählen die Methylierung von DNA, die Modifizierung von Histonen und die Expression nicht-kodierender RNA.

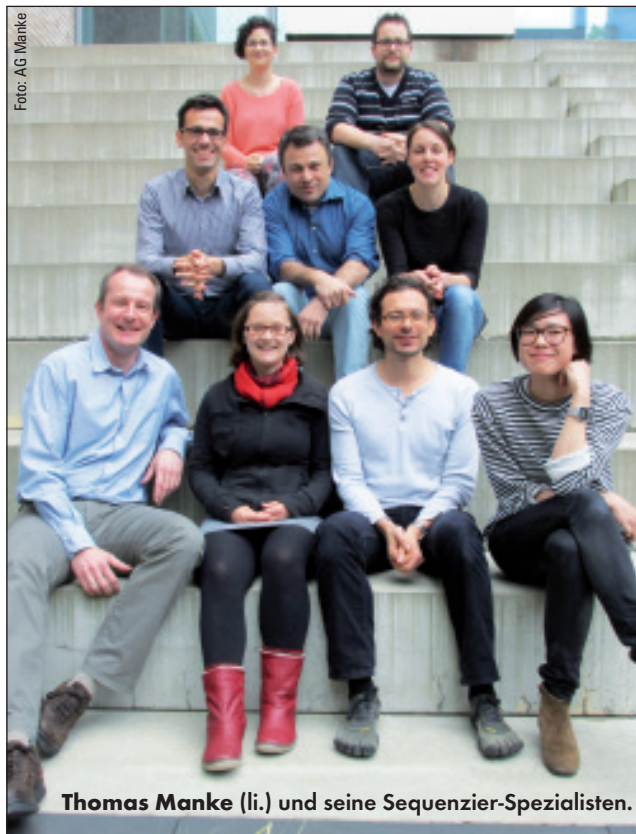
Die Entzifferung der dynamischen Epigenome ist um ein Vielfaches komplexer als die des statischen Genoms, denn durch die ständigen Veränderungen muss eine Sequenz wesentlich öfter abgelesen werden, um eine vergleichbare Abdeckung (Sequenzier-Tiefe) zu erhalten. Dies ist bislang nur in großen internationalen Forschungsverbänden möglich. Das International Human Epigenome Consortium (IHEC) etwa hat vor, in den nächsten Jahren 1.000 Epigenome für verschiedene

menschliche Zelltypen zu kartieren. Das Freiburger MPI und die Gruppe um Manke sind an diesem Großprojekt im Rahmen des Deutschen Epigenom-Programms (DEEP) beteiligt. Der Name ist Programm und soll die Bedeutung der Tiefensequenzierung für ein systematisches Verständnis von epigenetischen Prozessen unterstreichen. In einem weiteren Großprojekt, dem Freiburger Sonderforschungsbereich „Medical Epigenetics“ wird auch die langfristige Vision der Forschung am MPI-IE deutlich: Das bessere Verständnis der epigenetischen Mechanismen soll letztlich in bessere Prognosen und Therapien umgesetzt werden.

Günstig: Sequenzieren mit Antikörpern

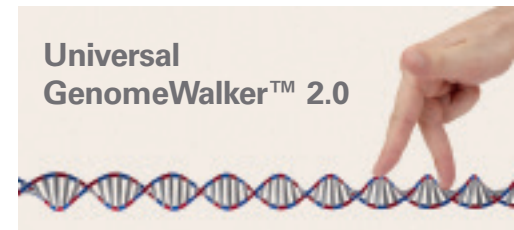
Die erste Aufgabe der Gruppe um Manke im Deutschen Epigenom-Programm besteht in der Generierung von Milliarden kurzer Sequenzfragmente aus einem gegebenen Pool von DNA-Proben. Hierzu hat die Gruppe eine standardisierte Pipeline aufgebaut. Nicht nur die Tiefensequenzierung, sondern auch die Vorbereitung der Proben ist hochgradig automatisiert – das steigert den Durchsatz und gewährleistet die Reproduzierbarkeit. Die analysierten Sequenzpools können dabei aus sehr unterschiedlichen Experimenten entspringen, doch die Gruppe interessiert sich in erster Linie für sogenannte ChIP-seq-Proben.

Bei der Chromatin-Immunpräzipitation (ChIP) inkubiert man fixierte und fragmentierte chromosomale DNA mit Antikörpern gegen Transkriptionsfaktoren oder Histon-Modifikationen und reichert sie an. Solcherart vorbereitet geht die doppelsträngige Präparation an die NGS-Unit um Manke. Sie ligieren spezielle Adapter an den gesamten Pool von DNA-Fragmenten (library generation), vervielfältigen diese zu



Thomas Manke (li.) und seine Sequenzier-Spezialisten.

Discover Unknown Sequences



“Walk” your DNA to identify

- Promoters or regulatory sequences
- Intron/exon junctions
- Integration sites for transposons and viruses
- Genome modifications performed using zinc finger nucleases, TAL effector nucleases, or other methods

Find out more at



www.clontech.com/GenomeWalker

Clustern aus klonalen Molekülen und führen schließlich eine möglichst tiefe Sequenzierung – mit möglichst hoher Abdeckung – durch.

Hierbei kommen im Freiburger Max-Planck-Institut eine HiSeq-2500 und die kleinere Benchtop-Variante “MiSeq“ von Illumina zum Einsatz. Diese arbeiten relativ kostengünstig nach dem *Sequencing by Synthesis*-Prinzip. Allerdings liefern sie die Sequenzen nur in der Form von kurzen *Reads* (DNA-Schnipsel von 50 bis 250 Basen). Ein typisches ChIP-seq-Experiment liefert etwa 50 Millionen *Reads* pro Probe. Genau das ist der Kern von *Deep Sequencing*: man kann – ganz im Gegensatz zu traditionellen Methoden wie der von Sanger – Millionen von verschiedenen DNA-Molekülen gleichzeitig sequenzieren. In einer durchschnittlichen Woche werden so mehrere hundert Milliarden Basenpaare generiert, was mehreren Dutzend menschlichen Genomen entspricht.

Teuer: Analyse der Reads

Aus diesen gewaltigen Datenmengen ergibt sich die zweite große Aufgabe der Gruppe: die primäre Analyse, wobei die Sequenzdaten zu sinnvollen Einheiten kombiniert werden. Am MPI-IE stehen dafür mehr als 100 TB Speicherkapazität in hunderten leistungsstarken Prozessoren zur Verfügung. „Auch dies reicht nur, wenn die Datenvolumina schnell genug reduziert werden“, erklärt Manke. Durch die rasanten Entwicklungen der letzten Jahre ist die Generierung von Sequenzdaten schon jetzt kostengünstiger als deren Speicherung. Damit kommt der effizienten Verkleinerung der Datensätze besondere Bedeutung zu.

Mit einem schnellen Mapping-Programm wie Bowtie werden die Millionen Sequenzschnipsel nun mit dem Genom des betreffenden Organismus abgeglichen und lokalisiert. Ein großes Problem dabei sind die repetitiven Sequenzen in der genomischen DNA, die nicht immer eine eindeutige Zuordnung erlauben. Hier testet die Gruppe um Manke verschiedene experimentelle Methoden und statistische Verfahren, um die Analyse zu verbessern. Letztlich interessiert sie aber vor allem die Anreicherung von Sequenzfragmenten in bestimmten Regionen des Genoms. Der besondere Anreiz der Tiefensequenzierung besteht darin, diese Anreicherungen für das gesamte Genom quantifizieren zu können. Mit einem einzigen Experiment können somit alle genomischen Bindestellen von Transkriptionsfaktoren oder sämtliche Regionen mit spezifischen Histon-Modifi-

kationen bestimmt werden. Programme wie MACS können solche angereicherten Bereiche zwar bereits identifizieren. Allerdings ist bei der Analyse besondere Vorsicht geboten, denn die komplexen experimentellen Protokolle verursachen eine ganze Reihe von systematischen Fehlern.

Zeitintensiv: Resultate interpretieren

In einer dritten Phase, der iterativen und integrativen Analyse, überführen Manke *et al.* die gewonnenen Informationen in biologisch relevante Aussagen und interpretieren sie, was durchaus mehrere Monate in Anspruch nehmen kann. „Diese Phase wird von den beteiligten Wissenschaftlern zeitlich oft unterschätzt“, betont Manke. „Hier hilft nur die ständige Kommunikation über Fachgrenzen hinweg, was nicht immer einfach ist.“ Doch die Genugtuung überwiege, wenn man gemeinsame Projekte erfolgreich abschließen könne.

Derlei Erfolge kann das Freiburger MPI-IE schon diverse vorweisen. So wurden in der Fruchtfliege neuartige Mechanismen der Kontrolle von „housekeeping“-Genen identifiziert (Kin Chung Lam *et al.*, *PLoS Genet* 2012, 8(6):e1002736), und auch die epigenetische Festlegung von Centromeren hat man beleuchtet (Agata Olszak *et al.*, *Nat Cell Biol* 2011, 13(7):799-808). Interessanterweise spielen aber auch altbekannte Transkriptionsfaktoren und Sequenzinformation eine wichtige Rolle bei der Regulation von Heterochromatin (Aydan Bulut-Karslioglu *et al.*, *Nat Struct Mol Biol* 2012, 19(10):1023-30).

Interdisziplinarität

Nicht alle Forschungsprojekte am MPI-IE können so intensiv unterstützt werden. Deshalb stellt die Bioinformatics-Gruppe hunderte getesteter und teilweise selbstentwickelter Programme und Pipelines als Webservice zur Verfügung. In Trainingsprogrammen und Workshops werden Studenten und interessierte Forscher auf die anfallenden Datenmengen vorbereitet.

Manke ist überzeugt, dass die Ausbildung von interdisziplinären Wissenschaftlern die größte Herausforderung der genomischen Forschung darstelle. Aber gerade darin sieht er auch die Attraktivität seiner Gruppe: „Unsere Stärke ist die wunderbare Mischung aus Informatikern, Biologen, Chemikern und Physikern. Denn trotz aller Automatisierungen und standardisierten Pipelines war der Bedarf an unterschiedlichen und ideenreichen Wissenschaftlern nie größer.“ **THOMAS SCHMITT**



MEET US AT

BIOTECHNICA 2013
October 8 - 10
Hannover, Germany

Takara Bio Europe

www.clontech.com

Clontech is a Takara Bio Company

tech@takara-clontech.eu • orders@takara-clontech.eu

Europe: +33 (0)1 3904 6880 • Toll free tel: Austria: 0800 296 141

• Germany: 0800 182 5178 • Switzerland: 0800 563 629

• UK: 0800 234 8063

For Research Use Only. Not for use in diagnostic or therapeutic procedures. Not for resale. SMARTer, Clontech, the Clontech logo and all other trademarks are the property of Clontech Laboratories, Inc. unless noted otherwise. ©2013 Clontech Laboratories, Inc.



Clontech