systembiologie.de

DAS MAGAZIN FÜR SYSTEMBIOLOGISCHE FORSCHUNG IN DEUTSCHLAND

AUSGABE 10 APRIL 2016

spezial:
bioinformatik als schlüsseltechnologie

das deutsche bioinformatik-netzwerk de.NBI stellt sich vor
ab Seite 8

ethische und rechtliche fragen in der genomforschung
Seite 40

woran man sich erinnert
Seite 72

interviews mit Niklas Blomberg, Heyo K. Kroemer, Peter Jansen und Olaf Wolkenhauer
Seite 28, 60, 69 und 66

ISSN 2191-2505

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

HELMHOLTZ
GEMEINSCHAFT

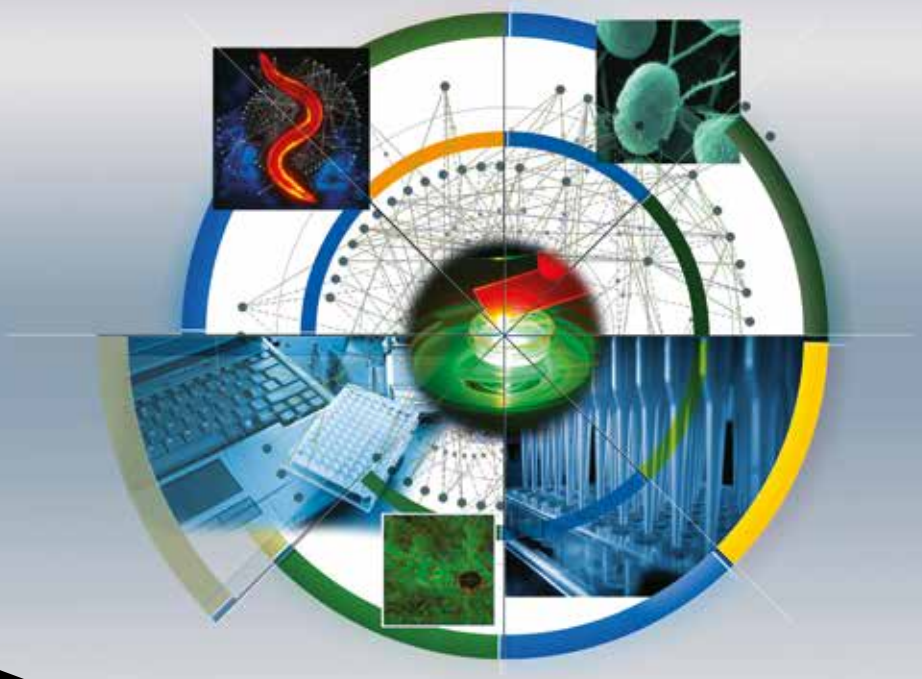


Foto: Derichs Kommunikation GmbH, Jülich, unter Verwendung von Fotos von Prof. Dr. Ralf Baumeister

systembiologie.de

Die Systembiologie ist eine junge und dynamische Disziplin mit dem Blick fürs Ganze. Als Teil der molekularen Lebenswissenschaften schlägt sie die Brücke zwischen ausgeklügeltem Laborexperiment und mathematischer Modellierung, zwischen hoch technisierter Erfassung von Messdaten und computergestützter Datenauswertung. Ihr Forschungsgegenstand sind die netzwerkartig verwobenen Abläufe der Signalübertragung und Stoffumwandlung in Zellen, Geweben, Organen und Organismen. Die systembiologische Forschung stellt sich dieser Komplexität, indem sie sich in fächerübergreifenden Netzwerken organisiert. Erfahren Sie im Magazin systembiologie.de, wie dieser faszinierende und aufstrebende Wissenschaftszweig arbeitet und welche Antworten er auf die bislang ungelösten Fragen des menschlichen Lebens findet.

Titelbild: Sergey Nivens – Fotolia.com

grußwort

Liebe Leserinnen und Leser,

die modernen Informationstechnologien haben die Welt der Biowissenschaft in den vergangenen Jahren rasant verändert. Heute können in wenigen Tagen ganze Genome sequenziert werden. Ebenso lässt sich eine Vielzahl von Proteinen mit neuesten Technologien analysieren. Für die Auswertung der dabei entstehenden Datenmengen hat sich die Bioinformatik als Schlüsseltechnologie etabliert. Sie liefert die Basis, um Daten effizient zu nutzen sowie systembiologische und systemmedizinische Modelle zu erstellen.

Das Bundesministerium für Bildung und Forschung fördert deshalb den Aufbau eines „Deutschen Netzwerkes für Bioinformatik-Infrastruktur". Dafür stellen wir in einem Zeitraum von fünf Jahren etwa 22 Millionen Euro zur Verfügung. Ziel ist es, bioinformatische Auswertungsprogramme und dazu passende Beratungen durch Expertinnen und Experten bundesweit als Dienstleistung zur Verfügung zu stellen – damit Forscherinnen und Forscher in den Lebenswissenschaften und der Biomedizin von den neuesten bioinformatischen Verfahren profitieren.

Durch das Zusammenspiel von Experiment, bioinformatischer Analyse, mathematischer Modellierung und Computersimulation entsteht ein immer detaillierteres und vollständigeres Bild von biologischen Prozessen. Die Modellierung und Simulation ganzer Zellen rückt in erreichbare Nähe. Und damit eröffnet sich beispielsweise die Möglichkeit, die Wirkung von Medikamenten im systembiologischen Gesamtkontext vorherzusagen.

Die vorliegende Ausgabe von systembiologie.de gibt einen faszinierenden Einblick in die Chancen, die sich aus der Verschmelzung von Informatik und Lebenswissenschaften für die Systembiologie und die biomedizinische Forschung im 21. Jahrhundert ergeben.

Prof. Dr. Johanna Wanka
Bundesministerin für Bildung und Forschung

Foto: Bundesregierung / Steffen Kugler

grußwort

Liebe Leserinnen und Leser,

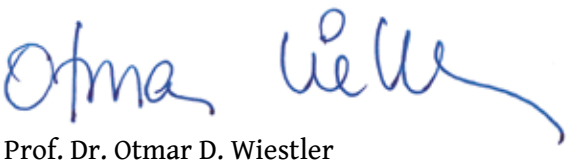
Computer sind aus dem heutigen Leben nicht mehr weg zu denken. Genauso wie für die meisten Menschen der tägliche Umgang mit PC, Smartphone und Digitalkamera heute ganz normal ist, so ist die wissenschaftliche Forschung ohne die Unterstützung von Computern heute gänzlich unvorstellbar geworden.

Während der Einsatz von Supercomputern und Computermodellen im letzten Jahrtausend zunächst vor allem Disziplinen wie der Teilchenphysik oder Klimaforschung vorbehalten war, sind in den letzten fünfzehn Jahren solche Methoden auch ein wesentlicher Bestandteil der Lebenswissenschaften geworden. Ein Paradebeispiel hierfür ist die Systembiologie, wo eine ganze Disziplin auf der Nachbildung von biologischen Systemen in mathematischen Computermodellen aufbaut. Der Vergleich von computergenerierten Modellvorhersagen und experimentellen Ergebnissen bietet die Grundlage, um die theoretischen Modelle weiter der biologischen Realität anzupassen. Auch für die Bioinformatik, deren vielfältige Aktivitäten rund um das Deutsche Netzwerk für Bioinformatik-Infrastruktur in diesem Heft vorgestellt werden, ist eine hochmoderne IT-Infrastruktur unerlässlich. Um diesen Anforderungen gerecht zu werden, betreibt das Forschungszentrum Jülich am Jülich Supercomputing Centre das Simulation Laboratory Biology (s. S. 56).

Aber nicht nur in der Grundlagenforschung ist der Computer aus dem Alltag nicht mehr wegzudenken. Auch die Medizin arbeitet in zunehmendem Ausmaß digital. Die Datenmengen, die heutzutage bereits in der Routineversorgung u. a. durch bildgebende Verfahren oder Langzeituntersuchungen entstehen, sind enorm. Hinzu kommen neue Technologien wie die Genomsequenzierung, die sich immer mehr als neue Methode zur differenzierten Diagnostik erweist. Größte Herausforderung beim produktiven Einsatz dieser Datenschätze ist allerdings nicht mehr die Erhebung der Daten, sondern viel mehr deren Verknüpfung und integrierte Auswertung. Hier besteht sowohl in der medizinischen Grundlagenforschung wie auch in der Patientenversorgung noch erhebliches Entwicklungspotential. Einen großartigen Ansatz in Richtung Datenintegration haben die Forscher des internationalen Projekts *Pan-Cancer Analysis of Whole Genomes (PCAWG)* (s. S. 34) geleistet: Durch das Zusammenführen von über 2.800 in verschiedensten Projekten gewonnenen Tumorsequenzen können hochrelevante Fragen zum Zusammenhang von Mutationen und der Krebsentstehung beantwortet werden.

Um solche Forschungsprojekte auch innerhalb Deutschlands häufiger durchführen zu können hat das BMBF eine Medizininformatik-Initiative initiiert (s. Interview mit Prof. Kroemer, S. 60), mit der in den nächsten Jahren auch die Grundlagen für eine weitere Digitalisierung der Medizin geschaffen werden können.

Die aktuelle Ausgabe von systembiologie.de vermittelt interessante Einblicke in das große Potential der Informationstechnologie im Bereich der Lebenswissenschaften. Ich wünsche Ihnen eine spannende Lektüre!



Prof. Dr. Otmar D. Wiestler
Präsident der Helmholtz-Gemeinschaft

Foto: Steffen Jänicke / Helmholtz-Gemeinschaft

vorwort

Wer hätte das gedacht...

meine lieben Leserinnen und Leser, vor Ihnen liegt eine spezielle Ausgabe zur Bioinformatik in einem Magazin, das sich sonst der Systembiologie widmet. Was ist geschehen? Die Lebenswissenschaften und die Gesundheitsforschung sind von Entwicklungen überrollt worden, die so noch vor zehn Jahren fast niemand voraus gesagt hätte. Als Anfang der 2000er Jahre die erste Blaupause des menschlichen Genoms veröffentlicht wurde, hatte dies noch die Anstrengung aller Sequenzierlabors dieser Welt erfordert und dabei Kosten in der Größenordnung von drei Milliarden Euro verschlungen. Nur ein Jahrzehnt später kann ein modernes Genomzentrum täglich Dutzende dieser Humangenome zu einem Preis um Tausend Euro sequenzieren. Und die Preise fallen weiter. Die Gewinnung der Genomdaten ist kein limitierender Faktor mehr, jedoch erweist sich die Analyse dieser Daten zunehmend als Hemmschuh des lebenswissenschaftlichen Fortschritts.

Wo werden wir also in weiteren zehn Jahren stehen? Die Beantwortung dieser Frage verlangt prophetische Fähigkeiten würden die einen sagen. „Die beste Methode die Zukunft vorherzusagen besteht darin, sie zu erfinden“ erwidert Alan Curtis Kay, der in den 1960er Jahren Mathematik und Molekularbiologie studierte und als Pionier der objektorientierten Programmierung und Architekt der fensterbasierten grafischen Benutzeroberflächen wie kaum ein anderer aktiv die Zukunft der Informationsgesellschaft mitgestaltet hat.

Die Bioinformatik hat die große Chance die Zukunft der Lebenswissenschaften selbst zu gestalten. Machbar erscheint alles, was informationstechnisch möglich und gesellschaftlich gewollt ist. In zehn Jahren werden wir ein menschliches Genom für einen Bruchteil der heutigen Kosten sequenzieren können. Die Genomsequenzierung wird damit für jedermann jederzeit möglich sein. Dasselbe gilt für die Erfassung des menschlichen Epigenoms, das als Spiegelbild der Interaktion des Genoms mit seiner Umwelt gilt. In zehn Jahren werden wir vermutlich lebenslang – und zwar von Geburt an – regelmäßig Bioproben wie Blut oder Speichel sammeln und sequenzieren. Wir werden damit in der Lage sein, die Entwicklung bestimmter Erkrankungen später im Leben vorherzusagen. Gewissermaßen ein (epi-)genetisches Gesundheits-Frühwarnsystem. Zudem werden wir rückblickend bei der Entstehung einer Erkrankung detailliert beschreiben können, zu welchem Zeitpunkt eine (epi-)genetische Fehlregulation stattgefunden hat, was eine noch präzisere Behandlung der Krankheit ermöglicht.

Ist dies die schöne neue Welt der (Epi-)Genomik? Neben all den rechtlichen und ethischen Herausforderungen, die mit einer alles umfassenden genetischen Überwachung verbunden sind, sei daran erinnert, dass diese kontinuierliche Erfassung unseres (epi-)genetischen Zustands enorme Datenmengen erzeugen wird. Schon heute stellt uns die Speicherung großer Sequenzdaten vor riesige Herausforderungen. Wer hier darauf hofft, dass sich das Problem angesichts des raschen Preisverfalls in der IT Branche von selber lösen wird, der sollte wissen, dass der Preisverfall in der Sequenzierung noch schneller verläuft. Oder mit anderen Worten gesagt, die Zeit spielt gegen uns: wir werden in immer kürzeren Zyklen mehr Daten in den Lebenswissenschaften produzieren als wir speichern können. Soll die Bioinformatik diese Datenflut bändigen, muss sie nach den wichtigsten Mustern in den Daten suchen, und das möglichst in Echtzeit. In anderen Disziplinen wie der Teilchenphysik ist dies heute schon Realität. Hier werden immens große Datenmengen erzeugt, um die Gesetzmäßigkeiten der Physik mit bislang unerreichter Präzision zu erforschen. In Zukunft werden wir in den Lebenswissenschaften Vorbilder wie dem LHC, dem Large Hadron Collider am CERN, folgen und nur einen Bruchteil der Daten aufheben, während die große Masse der Daten in Echtzeit verworfen wird. Hierbei wird die Bioinformatik die Standards setzen, den Rahmen des Möglichen definieren und zu heute noch nicht erahnten Einblicken in Gesundheit und Krankheit verhelfen.

Im vorliegenden Magazin wird die Bioinformatik aus verschiedensten Blickrichtungen beleuchtet, Bioinformatiker und Systembiologen kommen wie auch Nutzer und Förderer dieser Disziplin zu Wort. Allen gemein ist der große Enthusiasmus, den sie der Bioinformatik entgegen bringen.

Lassen Sie sich davon anstecken bei einer gewohnt unterhaltsamen Lektüre des systembiologie.de-Magazins.

Ihr Roland Eils

Chefredakteur

Foto: Tobias Schwerdt / DKFZ

inhalt

grußwort 3
Prof. Dr. Johanna Wanka, Bundesministerin für Bildung und Forschung

grußwort 4
Prof. Dr. Otmar D. Wiestler, Präsident der Helmholtz-Gemeinschaft

vorwort 5
Prof. Dr. Roland Eils, Chefredakteur

deutsches netzwerk für bioinformatik-infrastruktur – de.NBI 8
Eine BMBF Infrastrukturmaßnahme zur Lösung des Big-Data-Problems in den Lebenswissenschaften
von Alfred Pühler

rna-bioinformatik unter einem dach – das rna bioinformatik service zentrum 14
Herausforderungen und Lösungsansätze für eine einfach zugängliche Forschungsinfrastruktur
von Björn Grüning

BRENDA 18
Von der Datenbank zum Kompetenzzentrum
von Dietmar Schomburg und Ida Schomburg

NBI-SysBio – der de.NBI datenmanagement knoten 22
Vom spezialisierten Daten-Silo zu vernetzten Daten
von Wolfgang Müller

entwicklung von software-lösungen für mikrobielle bioinformatik 25
Institutsportrait: Bioinformatik Zentrum der Justus-Liebig-Universität Gießen
von Alexander Goesmann

ELIXIR – damit der datenfluss nicht ins stocken kommt 28
Interview mit Niklas Blomberg
von Marcus Garzón und Vera Grimm

omics-infrastrukturen für forschung und lehre 30
Ein Konzept der Leopoldina zum Umbruch in den Lebenswissenschaften
von Alfred Pühler

eine globale initiative zur erforschung von krebserkrankungen 34
Das Pan-Cancer Analysis of Whole Genomes (PCAWG) Projekt
von Jan O. Korbel, Sergei Yakneen, Sebastian M. Waszak, Matthias Schlesner, Roland Eils und Fruzsina Molnár-Gábor

sensible genomdaten 40
EURAT beantwortet ethische und rechtliche Fragen in der Genomforschung
von Sebastian Schuol und Eva C. Winkler

big data: perspektiven in der krebstherapie 44
Eindrücke aus der Industrieperspektive
von Ajay Kumar

fachgruppe bioinformatik – interessen mit einer stimme vertreten 49
FaBI – der Zusammenschluss von Fachgruppen aus fünf deutschen Fachgesellschaften aus den Lebenswissenschaften und der Informatik
von Matthias Rarey

neuigkeiten aus dem BMBF 52

neuigkeiten der helmholtz-gemeinschaft 56
Vorstellung der Simulation Laboratories am Jülich Supercomputing Centre (JSC)
von Olav Zimmermann

neue chancen für die medizin 60
Interview mit Heyo K. Kroemer
von Katja Nellissen, Marco Leuer und Bettina Koblenz

i:DSem – integrative datensemantik in der systemmedizin 63
Neue Initiative des Bundesministeriums für Bildung und Forschung für innovatives Datenmanagement in der biomedizinischen Forschung
von Christian Rückert

„projekte können auch auf eine gute art scheitern“ 66
Interview mit Olaf Wolkenhauer
von Melanie Bergs und Gesa Terstiege

„die leber ist erste wahl für wissenschaftler und unser netzwerk“ 69
Interview mit Peter Jansen
von Melanie Bergs und Gesa Terstiege

woran man sich erinnert 72
Die epigenetische Grundlage des Gedächtnisses aus systembiologischer Sicht
von Tonatiuh Pena Centeno, Ramon Vidal, Magali Hennion und Stefan Bonn

AptaBodies 76
DNA-Aptamere als Alternative für Antikörper beim Western Blotting
von Jasmin Dehnen und Frieda Anna Sorgenfrei

events 80

news 86

impressum 89

wir über uns 90

kontakt 91

deutsches netzwerk für bioinformatik-infrastruktur – de.NBI

Eine BMBF Infrastrukturmaßnahme zur Lösung des Big-Data-Problems in den Lebenswissenschaften

von Alfred Pühler

Das Big-Data-Problem in den Lebenswissenschaften ist die Folge eines Paradigmenwechsels. Mittels Omics-Technologien – und dazu zählen Genomik, Transkriptomik, Proteomik und Metabolomik – können heute zelluläre Bestandteile beliebiger Organismen ganzheitlich erfasst werden. Hierbei entstehen riesige Datenmengen, die nur über eine umfangreiche bioinformatische Infrastruktur gespeichert und analysiert werden können. Experimentell arbeitenden Gruppen steht ein solches Instrumentarium meist jedoch nicht zur Verfügung. Hier kommt nun das Deutsche Netzwerk für Bioinformatik-Infrastruktur (de.NBI) ins Spiel, das experimentell arbeitenden Gruppen in den Lebenswissenschaften bei der Auswertung großer Datenmengen aus Hochdurchsatzverfahren Hilfe anbietet.

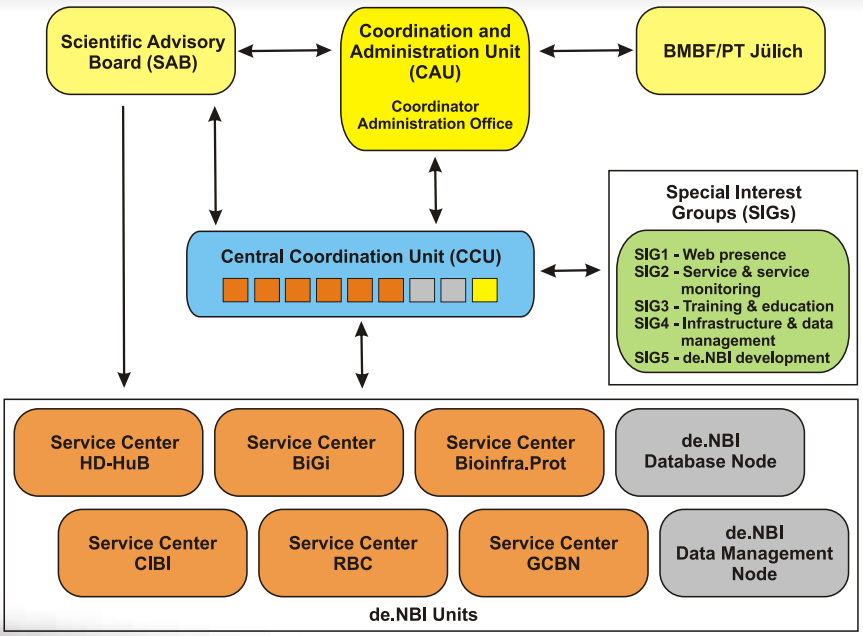
Etablierung und Aufgaben des de.NBI-Netzwerks

Die Etablierung eines Netzwerks für Bioinformatik-Infrastruktur geht auf eine Empfehlung des Bioökonomierates aus dem Jahr 2012 zurück. In einer Stellungnahme schlug er vor, ein Netz aus lokalen, gut ausgerüsteten und spezialisierten Zentren zu bilden, das von einem mit weitreichenden Befugnissen ausgestatteten Koordinationsgremium geleitet werden sollte [1].

de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

Dieser Vorschlag wurde vom Bundesministerium für Bildung und Forschung (BMBF) aufgenommen und im Mai 2013 durch eine Ausschreibung zur Bildung eines Deutschen Netzwerks für

Abbildung 1: de.NBI Organigramm



Grafik: de.NBI

Abbildung 2: Teilnehmer der ersten Sitzung der Zentralen Koordinationseinheit von de.NBI (Foto: de.NBI).

Bioinformatik-Infrastruktur (de.NBI) umgesetzt. Die Etablierung des de.NBI-Netzwerks erfolgte in einem gestaffelten Verfahren. Zunächst wurden durch eine internationale Gutachterkommission aus einer Vielzahl von eingegangenen Bewerbungen acht Servicezentren ausgewählt, die zur Erstellung eines Gesamtantrags aufgerufen wurden, der die Zusammenarbeit dieser Zentren in einem Netzwerk sicherstellen sollte. Die Erstellung dieses Gesamtantrags wurde bereits von einem Koordinator mit Geschäftsstelle geleitet. Der de.NBI-Gesamtantrag wurde im Juli 2014 der etablierten Gutachterkommission zur Überprüfung vorgelegt. Nach Zustimmung zum Gesamtantrag erfolgte der offizielle Start des de.NBI-Netzwerks im März 2015.

Die Basis des de.NBI-Netzwerks besteht aus den acht Servicezentren, die ein breitgefächertes Spektrum an Serviceangeboten abdecken. Eine Übersicht über diese Servicezentren kann der **Tabelle 1** entnommen werden. Zunächst sind drei de.NBI-Zentren Organismen-orientiert ausgerichtet. So widmet sich das Heidelberger HD-HuB-Zentrum der humanen Bioinformatik, das Bielefeld-Gießener BiGi-Zentrum der mikrobiellen Bioinformatik und das GCBN-Zentrum in Gatersleben der Pflanzenbioinformatik. Im Weiteren gibt es Zentren, die methodenorientiert angelegt sind. Das Freiburger RBC-Zentrum ist auf RNA-Bioinformatik fokussiert, während sich das in Bochum angesiedelte BioInfra.Prot-Zentrum auf Proteombioinformatik spezialisiert. Methodenorientiert ist auch das Tübinger CIBI-Zentrum, das sich der integrativen Bioinformatik widmet. Schließlich verbleiben noch die datenorientierten Servicezentren. Hier gibt es zunächst das Datenbank-Servicezentrum in Bremen und schließlich noch das Datenmanagement-Servicezentrum NBI-SysBio in Heidelberg. Diese acht Servicezentren sind die Arbeitspferde des de.NBI-Netzwerks und letztendlich dafür verantwortlich, dass de.NBI die gesteckten Ziele erreicht.

Die Aufgaben des de.NBI-Netzwerks lassen sich mit den Stichworten Service, Training und Ausbildung sehr gut kennzeichnen. Unter Service versteht man dabei die Bereitschaft, experimentell arbeitenden Gruppen im Bereich der Lebenswissenschaften bei der Analyse größerer Datenpakete behilflich zu sein. Der Trainingsaspekt wird über Kurse bedient, die experimentellen Wissenschaftlern den Einsatz vorhandener Bioinformatik-Programme erläutern. Zur Ausbildung tragen schließlich de.NBI-spezifische Workshops und Sommerschulen bei. Zusätzlich hat das de.NBI-Konsortium noch die Aufgabe, die Integration von de.NBI in das europaweite Bioinformatik-Netzwerk ELIXIR voranzutreiben, die Einbeziehung von Industrieunternehmen in das de.NBI-Netzwerk zu ermöglichen und Strategien zur Verstetigung von de.NBI über den Förderzeitraum von fünf Jahren hinaus zu entwickeln.

Organisation und Arbeitsweise des de.NBI-Netzwerks

Der Zusammenschluss der acht ausgewählten Servicezentren zu einem Netzwerk wird über eine klar gegliederte Organisationsstruktur geregelt (Abb.1). Wichtigstes Element in diesem Organigramm ist die Zentrale Koordinationseinheit, die alle Entscheidungen in Netzwerkfragen trifft. In der Koordinationseinheit ist jedes der acht Servicezentren durch einen Delegierten vertreten, ein weiterer Sitz ist dem de.NBI-Koordinator vorbehalten. Die Koordinationseinheit hat themenspezifische Fachgruppen ins Leben gerufen, die der Koordinationseinheit zuarbeiten und

Tabelle 1: Liste der de.NBI-Partner

ZENTREN	BETEILIGTE PARTNER
Heidelberg Center for Human Bioinformatics – HD-HuB Zentrumskoordinator: Roland Eils, Heidelberg	• Universität Heidelberg • DKFZ Heidelberg • EMBL Heidelberg
Bielefeld-Gießen Resource Center for Microbial Bioinformatics – BiGi Zentrumskoordinator: Jens Stoye, Bielefeld	• Universität Bielefeld • Universität Gießen
Bioinformatics for Proteomics - BioInfra.Prot Zentrumskoordinator: Martin Eisenacher, Bochum	• Universität Bochum • Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V., Dortmund
Center for Integrative Bioinformatics – CIBI Zentrumskoordinator: Oliver Kohlbacher, Tübingen	• Universität Tübingen • Freie Universität Berlin • Universität Konstanz
RNA-Bioinformatics Center – RBC Rolf Backofen, Freiburg	• Universität Freiburg • Universität Leipzig • Max-Delbrück-Centrum für Molekulare Medizin, Berlin
German Crop BioGreenformatics Network – GCBN Zentrumskoordinator: Uwe Scholz, Gatersleben	• Leibniz-Institut für Pflanzengenetik und Kulturpflanzen forschung (IPK), Gatersleben • Helmholtz-Zentrum München • Forschungszentrum Jülich
Databases Zentrumskoordinator: Frank-Oliver Glöckner, Bremen	• Jacobs University Bremen gGmbH • Universität Bremen • Technische Universität Braunschweig • Leibniz-Institut DSMZ GmbH, Braunschweig
Data Management Node – NBI-SysBio Zentrumskoordinator: Wolfgang Müller, Heidelberg	• Heidelberger Institut für Theoretische Studien • Universität Rostock

vor allem anstehende Entscheidungsprozesse vorbereiten. Die de.NBI-Fachgruppen bearbeiten Themengebiete wie Web-Präsenz, Service und Service-Monitoring, Training und Ausbildung, Infrastruktur und Datenmanagement sowie de.NBI-Entwicklung. Das Netzwerk wird von dem bereits genannten Koordinator mit Geschäftsstelle gesteuert. Die Geschäftsstelle selbst besteht aus einem Geschäftsstellenleiter, einer wissenschaftlichen Mitarbeiterin, einem Web- und Service-Experten sowie einem Experten für Trainings- und Ausbildungsfragen.

Im ersten Jahr seines Bestehens bestand die vordringlichste Aufgabe des de.NBI-Netzwerks darin, die einzelnen Komponenten des gesamten Netzwerks abgestimmt zum Arbeiten zu bringen. Diese Aufgabe wurde in mehreren Sitzungen der

Zentralen Koordinationseinheit angegangen, wobei die de.NBI-Fachgruppen wesentliche Impulse beisteuerten. So bereitete die Fachgruppe **Web-Präsenz** speziell den Webauftritt von de.NBI vor, der als zentrales Element des gesamten Netzwerks angesehen werden muss. Eine de.NBI-Webpage, die unter dem Link www.denbi.de erreicht werden kann, gibt Auskunft über den bisherigen Entwicklungsstand. Die Fachgruppe **Service und Service-Monitoring** erstellte eine Liste aller in den einzelnen Servicezentren angesiedelten bioinformatischen Serviceangebote. Die Liste umfasst ca. 80 Serviceangebote und kann zukünftig ebenfalls über die de.NBI-Homepage eingesehen werden. Die Fachgruppe **Training und Ausbildung** koordinierte zusammen mit dem Trainingsexperten das Angebot der in den einzelnen Servicezentren entwickelten Trainingskurse. Bisher wurden bereits 16 Trainingskurse durchgeführt, die insgesamt von 329 Teilnehmern besucht wurden. Die Fachgruppe **Infrastruktur und Datenmanagement** widmete sich vor allem der im de.NBI-Netzwerk vorhandenen Rechnerkapazität, die derzeit als suboptimal eingeschätzt wird. Über den Umgang mit Daten, speziell aus dem medizinischen Bereich, wurden aus ethischer Sicht erste Vorgehensweisen entwickelt. Die Fachgruppe **de.NBI-Entwicklung** behandelte meist übergeordnete Fragen wie die Fortentwicklung des de.NBI-Netzwerks im Hinblick auf fachliche Ergänzungen, die internationale Kooperation, die Einbindung von Industrieunternehmen und die Verstetigung des Netzwerks. Bisher wurde wesentliche Arbeit bei der fachlichen Ergänzung des de.NBI-Netzwerks durch sogenannte Partnerprojekte und der Eingliederung von de.NBI als nationaler Knoten in das europaweite Bioinformatik-Netzwerk ELIXIR geleistet.

Die Zentrale Koordinationseinheit trifft sich alle drei Monate an den verschiedenen Standorten der einzelnen Servicezentren. Das erste Treffen, das bereits am 10. März 2015 in Berlin stattfand (Abb. 2), legte das administrative Regelwerk fest und versetzte so das de.NBI-Netzwerk in einen funktionsfähigen Zustand. Das Kick-off-Treffen des de.NBI-Netzwerks, verbunden mit einer ersten de.NBI-Vollversammlung, wurde dann am 26. März 2015 in Bielefeld abgehalten (Abb. 3). Seit dieser Zeit kooperieren die einzelnen Servicezentren intensiv und stellen sicher, dass die angestrebten Ziele im de.NBI-Netzwerk erreicht werden.

Eine weitere Aufgabe des de.NBI-Netzwerks bestand in der Etablierung eines wissenschaftlichen Beirats. Hier wurde dem BMBF vom de.NBI-Koordinator und der Zentralen Koordinationseinheit eine Liste von möglichen Kandidatinnen und Kandidaten vorgeschlagen. Nach Bestätigung dieser Liste konnten sechs Plätze im wissenschaftlichen Beirat besetzt werden. Die sechs Mitglieder sind ausgewiesene Bioinformatiker(innen) mit Expertenwissen auf dem Gebiet von Bioinformatik-Infrastrukturen. Ein erstes Zusammentreffen des wissenschaftlichen Beirats mit dem de.NBI-Netzwerk fand im November 2015 in Berlin statt. Das de.NBI-Netzwerk legte einen schriftlichen Report vor, der über die de.NBI-Etablierung und die geleistete Arbeit berichtete. Das Zusammentreffen zwischen Beirat und Netzwerk war in einen de.NBI-Workshop eingebettet, der die bisher erzielten Ergebnisse der de.NBI-Repräsentanten zum Inhalt hatte. Der Beirat verfasste einen ausführlichen Bericht, der die einzelnen Elemente von de.NBI beleuchtete. Dieser Bericht bescheinigt, dass das de.NBI-Netzwerk in den ersten neun Monaten seines Bestehens vorzügliche Arbeit geleistet hat. Darüber hinaus wurden Empfehlungen ausgesprochen, in welche Richtung sich das de.NBI-Netzwerk in Zukunft entwickeln sollte. Diese Stellungnahme des wissenschaftlichen Beirats ist für de.NBI von großer Bedeutung und wird nach intensiver Diskussion die Grundlage für einen de.NBI-Entwicklungsplan bilden.

Ein Entwicklungsplan zum Ausbau des de.NBI-Netzwerks

Nach Abschluss des ersten Förderjahres lässt sich festhalten, dass das de.NBI-Netzwerk zügig etabliert werden konnte und dass es seitdem mit Nachdruck an der Umsetzung der gesteckten Ziele arbeitet. Diskussionen mit dem wissenschaftlichen Beirat erbrachten, dass die Ziele des Netzwerkes ausgeweitet werden sollten. Als Konsequenz daraus wird innerhalb von de.NBI ein

Tabelle 2: Der de.NBI-Entwicklungsplan 2016

Integration von Partnerprojekten
Etablierung eines Forschungsprogramms mit Doktoranden
Integration von de.NBI als nationaler Knoten in ELIXIR
Entwicklung eines industriellen Zweiges des de.NBI-Netzwerks
Entwicklung einer de.NBI-Cloud
Verstetigung des de.NBI-Netzwerks

Entwicklungsplan diskutiert, der die im Folgenden skizzierten Punkte beinhaltet (Tab. 2).

Zunächst gilt es, thematische Lücken im de.NBI-Netzwerk durch die Integration von Partnerprojekten zu schließen. Anschließend sollte zur Etablierung einer Forschungskomponente im de.NBI-Netzwerk ein Forschungsprogramm mit Doktoranden installiert werden, dass natürlich auch dazu gedacht ist, Bioinformatiker auf dem Gebiet der Bioinformatik-Infrastruktur auszubilden. Des Weiteren gilt es, de.NBI als nationalen Knoten in das europaweite Bioinformatik-Infrastruktur-Netzwerk ELIXIR einzubinden. Hierzu müssen rechtliche und personelle Voraussetzungen für eine enge Kooperation geschaffen werden. Die Einbindung von Industrieunternehmen in das de.NBI-Netzwerk soll durch die Schaffung eines industriellen Zweiges von de.NBI vorangetrieben werden. Hier gilt es, vor allem zu klären, wie dieser industrielle Zweig mit dem akademisch ausgerichteten de.NBI-Netzwerk interagieren kann. Zur Lösung der fehlenden Rechnerkapazität innerhalb des de.NBI-Netzwerks soll eine de.NBI-eigene Cloud ins Leben gerufen werden. Eine solche de.NBI-Cloud ist natürlich mit weiteren Cloud-Aktivitäten in Deutschland und Europa abzustimmen. Schließlich hat de.NBI noch einen Plan vorzulegen, wie das Netzwerk nach Ablauf von fünf Jahren BMBF-Förderung verstetigt werden kann. Hierzu sollen Verstetigungspläne anderer europäischer Staaten, die als nationale Knoten in ELIXIR mitarbeiten, ausgewertet und auf Ihre Übertragbarkeit analysiert werden.

Steckbrief des deutschen Netzwerks für Bioinformatik-Infrastruktur (de.NBI):

Das de.NBI-Netzwerk, ein Projekt des Bundesministeriums für Bildung und Forschung, wurde am 1. März 2015 gestartet. Es gliedert sich in acht Servicezentren mit insgesamt 23 Partnern und wird von einem Koordinator mit Geschäftsstelle geleitet. Zur Steuerung des de.NBI-Netzwerks wurde eine Zentrale Koordinationseinheit eingerichtet, in der alle acht Servicezentren durch Verbundkoordinatoren vertreten sind.

Das de.NBI-Netzwerk wurde ins Leben gerufen, um experimentell arbeitende Gruppen im Bereich der Lebenswissenschaften bei der Analyse von großen Datenmengen behilflich zu sein. Die de.NBI-Hauptaufgaben lassen sich in Service, Training und Ausbildung untergliedern. Darüber hinaus soll de.NBI die Kooperation mit weiteren Bioinformatik-Infrastukturen in Europa in die Wege leiten, aber auch einschlägige Industrieunternehmen einbinden und Modelle für eine Verstetigung von de.NBI entwickeln.

Abbildung 3: Teilnehmer am Kick-off Treffen des de.NBI-Netzwerks in Bielefeld (Foto: de.NBI).

Beteiligte Partner:

Koordinator:

Prof. Dr. Alfred Pühler, Centrum für Biotechnologie (CeBiTec), Universität Bielefeld

Geschäftsstellenleiter:

Prof. Dr. Andreas Tauch, Centrum für Biotechnologie (CeBiTec), Universität Bielefeld

Leiter der beteiligten Servicezentren:

Prof. Dr. Rolf Backofen, Freiburg
Prof. Dr. Roland Eils, Heidelberg
PD Dr. Martin Eisenacher, Bochum
Prof. Dr. Frank Oliver Glöckner, Bremen
Prof. Dr. Oliver Kohlbacher, Tübingen
PD Dr. Wolfgang Müller, Heidelberg
Dr. Uwe Scholz, Gatersleben
Prof. Dr. Jens Stoye, Bielefeld

Referenz:

[1] Anforderungen an eine Bioinformatik-Infrastruktur in Deutschland zur Durchführung von Bioökonomierelevanter Forschung, Empfehlungen des Bioökonomierats (2012) http://www.biooekonomierat.de/publikationen.html?tx_rsmpublications_pi1[publication]=7&tx_rsmpublications_pi1[action]=show&tx_rsmpublications_pi1[controller]=Publication&cHash=d4769cfcee403d5ac8f14b01c3b3930d

Kontakt:

Prof. Dr. Alfred Pühler
Koordinator
puehler@cebitec.uni-bielefeld.de

Prof. Dr. Andreas Tauch
Co-Koordinator und Geschäftsstellenleiter
tauch@cebitec.uni-bielefeld.de

Centrum für Biotechnologie
Universität Bielefeld
Bielefeld
contact@denbi.de

www.denbi.de

de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

Foto Herr Pühler: Centrum für Biotechnologie, Bielefeld; Foto Herr Tauch: A. Tauch

rna-bioinformatik unter einem dach – das rna bioinformatik service zentrum

Herausforderungen und Lösungsansätze für eine einfach zugängliche Forschungsinfrastruktur

von Björn Grüning

Das RNA Bioinformatics Center (RBC) ist ein Leistungszentrum des Deutschen Netzwerks für Bioinformatik-Infrastruktur (de.NBI), das sich mit RNA-basierten Mechanismen der Genregulation beschäftigt. Die Aufgabe des RBC besteht darin, eine umfassende integrative Plattform zur RNA-Analyse zu entwickeln und dabei über die große Bedeutung der RNA bei der Genregulation aufzuklären. Der Service des RBC reicht von der Beratung zum experimentellen Studiendesign, über die Bereitstellung von Protokollen zur Datenauswertung und dazugehöriger Infrastruktur, bis hin zur Entwicklung von spezifischen Lösungsansätzen für individuelle wissenschaftliche Fragestellungen.

Die Bedeutung nicht-kodierender RNAs in der medizinischen Forschung

Nicht-kodierende RNAs (engl. *non-coding ribonucleic acids*, ncRNAs) und RNA-Protein-Interaktionen wurden in der Wissenschaft sehr lange ignoriert, weil der Fokus der Forschung bis vor einigen Jahren zumeist auf den Protein-kodierenden Regionen der DNA lag. Die genomweite Sequenzierung aber zeigte, dass der Großteil der DNA nicht für Proteine, sondern für ncRNAs kodiert. Zur Untersuchung von regulatorischen RNAs, z. B. mikro-RNAs (miRNA), und RNA-Protein-Interaktionen wurden neue Technologien entwickelt mit denen man zeigte, dass die Komplexität der Genregulation auf post-translationaler Ebene vergleichbar mit der transkriptionellen Genregulation ist. Das menschliche Genom enthält tausende miRNAs und mindestens 800 RNA-bindende Proteine. Mit diesem neuen Wissen konnte bereits nachgewiesen werden, dass viele Erkrankungen nicht nur durch Mutationen in bestimmten Genen ausgelöst werden, sondern ihre Ursache speziell in der post-transkriptionellen Genregulation liegen kann[1].

Ziele und Aufgaben des RBC

Das Leistungszentrum RNA Bioinformatics Center (RBC) vereint die international renommierten deutschen RNA-Bioinformatik-Arbeitsgruppen aus Freiburg (Rolf Backofen, Koordinator), Berlin (Nikolaus Rajewsky und Uwe Ohler, MDC) und Leipzig (Peter F. Stadler). Das RBC verfolgt drei Ziele:

1.) Die Etablierung einer einfach zugänglichen RNA-Workbench. Diese Analyse-Umgebung kann auf jedem PC verwendet oder alternativ in einer HPC-Umgebung (Universitätsrechenzentrum oder Cloudcomputing) zur Verfügung gestellt werden.

2.) Schaffung einer nachhaltigen Infrastruktur. Das RBC arbeitet mit einer Vielzahl von anderen Zentren zusammen, um zum einen die Interoperabilität zwischen den Zentren zu fördern, aber auch um die Wissenschaftsgemeinschaft auf ein breites Fundament zu stellen und die geschaffenen Lösungen nachhaltig zu gestalten.

3.) Die RNA-Workbench dient neben dem Produktiveinsatz als Experimentierplattform für Schulungen und Ausbildung. Um auch dem wissenschaftlichen Nachwuchs und Wissenschaftlern ohne Bioinformatikkenntnisse einen Zugang zu den RNA-basierten Analysetools zu ermöglichen und das Wissen über die Bedeutsamkeit der RNA-abhängigen Regulation weiterzugeben, soll es umfassende Trainings und Workshops zum Thema geben.

Das RBC benutzt zur Realisierung der genannten Ziele das weltweit verbreitete Workflow-Managementsystem Galaxy[2]. Diese Plattform ermöglicht es, Lebenswissenschaftlern auf transparente und reproduzierbare Weise selbstständig Sequenzdaten umfassend auszuwerten und mit anderen Wissenschaftlern auszutauschen. Bisher wurden in der RNA-Workbench mehr als 50 verschiedene Bioinformatik-Anwendungen integriert, die speziell zur Auswertung RNA-basierter Daten programmiert wurden. Um die verschiedenen Regulationsebenen (Transkription, Splicing, Translation und Degradierung) von RNA-Transkripten und deren gegenseitige Einflüsse umfassend zu analysieren und zu verstehen, bietet das RBC in seiner Workbench neben Tools zur Analyse von Transkriptom-Daten (z. B. RNA-seq) beispielsweise auch eine Vielzahl an Tools zur RNA-Struktur-Analyse

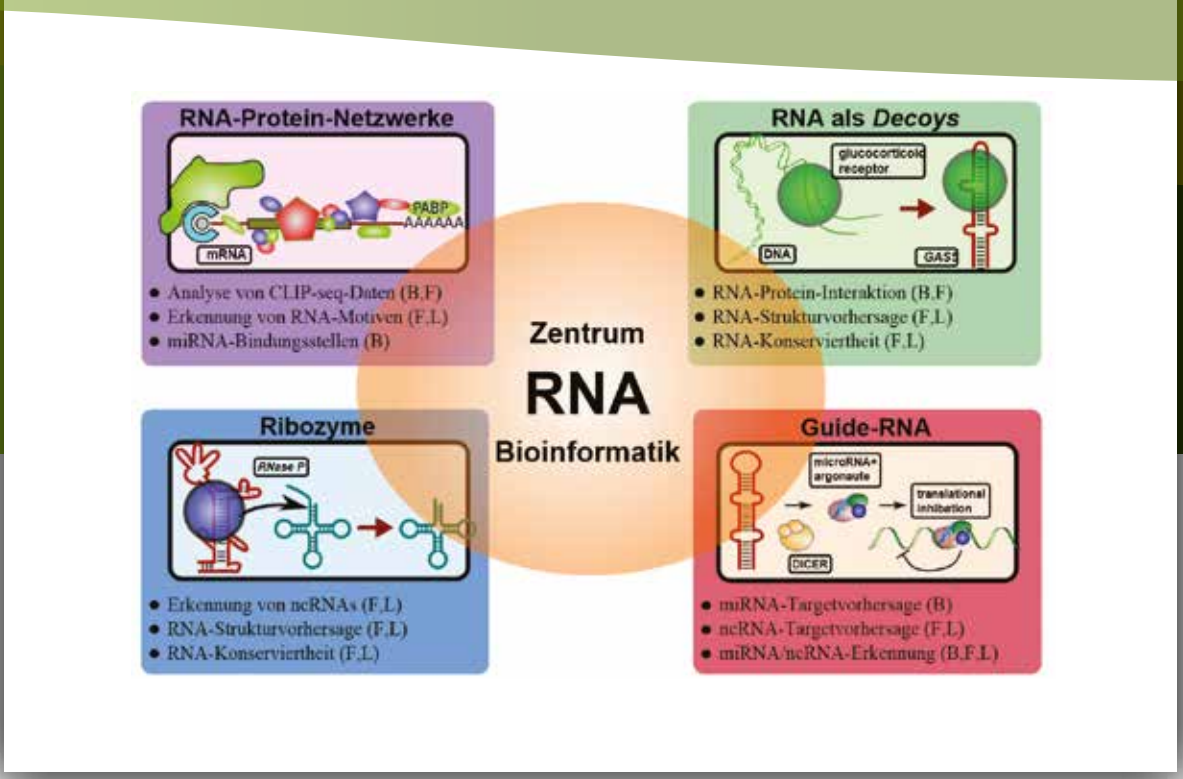


Abbildung 1: Das RNA Bioinformatics Center (RBC), bestehend aus AG Prof. Dr. Rolf Backofen in Freiburg (F), AG Prof. Dr. Peter F. Stadler in Leipzig (L) und AG Prof. Dr. Uwe Ohler und Prof. Dr. Nikolaus Rajewsky (B), ist die zentrale Anlaufstelle für alle RNA-bioinformatischen Fragen (Grafik: Prof. Dr. Rolf Backofen).

(LocARNA, GraphClust, ExpaRNA), zur Vorhersage von ncRNA-Zielstrukturen (ViennaRNA Suite) und zur Definition und Klassifizierung von RNA-Transkripten an.

Außerdem wurden bereits die vom RBC entwickelten und weltweit verbreiteten Tools zur Charakterisierung von ncRNAs in Galaxy integriert. Darunter sind Anwendungen zur Detektion von miRNAs (MirDeep, PipMir, BlockClust, RNAz), von neuen Transkripten (lncRNAs), zum Mapping von HTS-Daten (Segemehl Suite[3]) und zur Vorhersage von RNA-RNA und Protein-RNA-Interaktionen (IntaRNA, RNAup, CopraRNA, PARalyzer[4], Dorina[5], GraphProt[6]). Viele dieser Werkzeuge stehen bereits als individuelle Web-Services zur Verfügung[a].

Galaxy – eine webbasierte Open-Source-Plattform für datenintensive biomedizinische Forschung

Es gibt zahlreiche kommerzielle und frei erhältliche RNA-Analyse-Tools. Was bisher allerdings fehlt, ist ein einfacher Zugang zu einem vereinheitlichten, integrativen System, das eine standardisierte Untersuchung von RNA-basierter Genregulation bietet. Hinzu kommt, dass die meisten Software-Tools von Lebenswissenschaftlern ohne informatische Kenntnisse kaum benutzbar sind, da die Installation zuweilen nicht trivial ist und die Zugänglichkeit oft Programmierkenntnisse voraussetzt. Hierfür bietet Galaxy eine einzigartige Lösung an, weil frei verfügbare und selbst entwickelte Tools sowie Visualisierungen und Datenbanken in Galaxy integriert werden können, so dass diese Komponenten transparent für jeden Lebenswissenschaftler einfach zugänglich und nutzbar werden. In Galaxy verfügbar sind derzeit mehr als 2000 verschiedene Tools, von einfacher Textmanipulation bis hin zur HTS-Analyse (z. B. Mapping, differentielle Genexpressionsanalyse). Mit Hilfe dieser Plattform ist es für jeden Anwender möglich, alle Tools miteinander in Workflows frei zu kombinieren und damit transparent und reproduzierbar einzusetzen. Der Galaxy-Server in Freiburg ist eine der größten Galaxy-Instanzen weltweit und das RBC leistet in großem Umfang Beiträge zur Weiterentwicklung der Galaxy-Plattform. Viele renommierte Universitäten nehmen den RBC-Server bereits als Vorlage zur Etablierung lokaler Galaxy-Instanzen und benutzen die Docker-basierte Virtualisierungslösung des RBC. Ein zentraler Aspekt des RBC-Leistungszentrums ist neben dem Angebot der Datenanalyse auch das Training der Anwender in Hinblick auf die Benutzung der bereitgestellten Tools und Infrastruktur. Alle Partner des RBC bieten mehrmals jährlich verschiedene de.NBI-Workshops und Hands-on Tutorien für die Benutzung von Galaxy und die Auswertung von RNA-Daten an. Die Workshops behandeln unterschiedliche Themen, z. B. Genomannotation oder *High-throughput Sequencing* (HTS)- Datenanalyse, und haben eine Dauer von 1-5 Tagen. Zur Vermittlung RNA-bioinformatischer Kenntnisse beteiligt sich das RBC auch an den vom de.NBI-Netzwerk organisierten *Summer Schools*. Darüber hinaus organisiert das RBC halbjährliche *Hackathons*, bei denen sich Entwickler aus dem In- und Ausland treffen, um neue Tools in die RNA-Workbench zu integrieren und zu testen. Das RBC

[a] http://rna.informatik.uni-freiburg.de/
http://www.bioinf.uni-leipzig.de/webServices.html
https://ohlerlab.mdc-berlin.de/software/

entwickelt zusätzliche ständig aktualisierte Analyse-Workflows für eine standardisierte HTS-Datenauswertung und stellt diese der Wissenschaftsgemeinschaft frei zur Verfügung.

Zugang zur RNA-Workbench

Die RNA-Workbench, die auf Basis von Galaxy entwickelt wurde, kann ohne vorherige Installation auf dem eigenen Rechner ganz einfach über einen Webbrowser benutzt werden. Das bietet für den Anwender den Vorteil, dass er keine großen Datensätze herunterladen muss und die Analyse keinen Speicherplatz auf dem eigenen Computer benötigt. Das RBC stellt einen Galaxy-Server mit u. a. den Tools der RNA-Workbench zum Testen zur Verfügung. Aufgrund der hohen Anforderungen der Tools, was Rechenleistung betrifft, des großen Bedarfs an Speicherkapazitäten für HTS-Daten und individuelle Anforderungen wie Sicherheit und Datenschutz können nicht alle Analysen zentral auf einem Server durchgeführt werden. Hierfür hat das RBC eine Alternative entwickelt, die es jedem Benutzer ermöglicht, die RNA-Workbench unabhängig von einem der öffentlich zugänglichen Server zu benutzen. Diese Alternative nennt sich Galaxy Docker. Galaxy Docker ist eine vorkonfigurierte Galaxy-Plattform mit einer Tool-Sammlung, die individuell zusammengestellt werden kann und auf jedem System benutzbar ist. Aufgrund der abgekapselten Architektur des Galaxy Docker-Projekts ist es möglich, diese Analyse-Umgebung auch in geschlossenen Netzwerken zu betreiben und ist damit optimal für die Auswertung hochsensibler Daten geeignet. So wird das von uns entwickelte Galaxy Docker-Projekt zum Beispiel erfolgreich an der Universität in Oslo eingesetzt, um Dienstleistungen für sensitive Daten anzubieten.

Zusammenarbeit mit den de.NBI-Zentren und anderen Projekten

Das RBC arbeitet in unterschiedlichen Projekten eng mit den verschiedenen Leistungszentren des de.NBI zusammen. Um beispielsweise eine gemeinsame Beschreibungssprache für Kommandozeilen-Tools zu etablieren, welche es ermöglicht diese Tools automatisch in Galaxy einzubinden, arbeitet das RBC eng mit dem Zentrum für integrative Bioinformatik (CiBi) zusammen. Die Nutzung von Docker, welches auch als eine mögliche Lösung zur Virtualisierung aller de.NBI-Tools getestet wird, wird in Kollaboration mit dem Heidelberger Zentrum für Human-Bioinformatik (HD-HuB) und der de.NBI-Datenbank-Abteilung koordiniert. Mit dieser wird außerdem an der Integration der Datenbank SILVA und BacDive, welche rRNA Daten bereitstellt, in Galaxy gearbeitet.

Darüber hinaus bestehen Kooperationen mit dem ELIXIR-Projekt, dessen Aufgabe es ist, Europas führende Organisationen der Lebenswissenschaften zu vereinen, um gemeinsame Strategien zum Management und zur Sicherung der enormen Datenmengen zu entwickeln, die täglich durch öffentlich geförderte Projekte erhoben werden. Das RBC ist in diesen Prozess involviert und treibt die Kooperation durch Teilnahmen an Veranstaltungen des ELIXIR-Projekts und durch Einladungen von ELIXIR-Partnern zu RBC-Veranstaltungen voran. Kooperationen mit der BioJS-Community und dem Jupyter-Projekt ermöglichen deren Nutzung direkt aus Galaxy heraus und fördern den explorativen Charakter der RNA-Workbench.

Das langfristige wissenschaftliche Ziel des RBC ist es eine umfassende Betrachtung der Genregulation aus Sicht der RNA zu erlangen. Das Wissen wird dazu beitragen den Einfluss der RNA-basierten Regulation auf die nachfolgenden Regulationsprozesse der Genregulation, z. B. DNA-Methylierung (Epigenetik), Transkription und Translation, besser zu verstehen. Die Erkenntnisse zur Rolle der Struktur von nicht-kodierenden RNAs sind derzeit auch durch die eingeschränkte Verfügbarkeit bioinformatischer Tools noch stark begrenzt. Vor allem bei der Struktur von mRNAs, die von einem Zusammenspiel verschiedener Interaktionen mit RNA-Bindeproteinen beeinflusst wird, gibt es noch viel Forschungsbedarf. Das RBC leistet mit seinem Service-Angebot und der Entwicklung der RNA-Workbench einen wichtigen Beitrag zur umfassenden, integrativen Analyse der RNA-basierten Genregulation.

Das RBC bietet 2016 folgende Workshops an:

Freiburg:

RBC Hackathon Januar 2016
Galaxy HTS-DatenanalyseWorkshop I – Februar 2016
Galaxy HTS-Datenanalyse Workshop II – September 2016
Genom-Annotation Workshop – Juni 2016

Berlin:

Computational Genomics Workshop – Februar 2016
Galaxy Workshop – Mai 2016
Computational Genomics in Precision Medicine Workshop September 2016

Leipzig:

Summer School mit Jan Gorodkin (ELIXIR)

Abbildung 2: Das RBC Team (von links nach rechts): Dilmurat Yusuf, Sebastian Will, Björn Grüning, Anika Erxleben, Torsten Houwaart, Rolf Backofen, Peter F. Stadler, Uwe Ohler, Altuna Akalin, Bora Uya (Foto: Prof. Dr. Rolf Backofen)

Steckbrief Forschungsprojekt:

Projektname:

Das RBC Bioinformatics Center:
Prof. Dr. Rolf Backofen (RBC Koordinator)
Prof. Dr. Peter F. Stadler
Prof. Dr. Uwe Ohler
Prof. Dr. Nikolaus Rajewsky

Mitglieder des RNA-Bioinformatik-Zentrums:

Dr. Björn Grüning, Dr. Anika Erxleben, Dr. Torsten Houwaart, Dr. Altuna Akalin, Bora Uya, Dr. Dilmurat Yusuf, Dr. Sebastian Will, Prof. Dr. Nikolaus Rajewsky, Prof. Dr. Uwe Ohler, Prof. Dr. Peter F. Stadler, Prof. Dr. Rolf Backofen

Referenzen:

[1] S. Gerstberger, M. Hafner, and T. Tuschl. A census of human RNA-binding proteins. Nature Reviews Genetics 2014; 15: 829–845. doi:10.1038/nrg3813

[2] Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25;11(8):R86. doi:10.1186/gb-2010-11-8-r86

[3] Christian Otto, Peter F. Stadler and Steve Hoffmann: Lacking alignments? The next-generation sequencing mapper segemehl revisited. Bioinformatics 2014 March 13; 30 doi: 10.1093/bioinformatics/btu146

[4] Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U.: PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol. 2011, 12:R79 doi:10.1186/gb-2011-12-8-r79

[5] Blin K, Dieterich C, Wurmus R, Rajewsky N, Landthaler M, Akalin: DoRiNA 2.0-upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. Nucleic Acids Res. 2015 Jan; 43(Database issue): D160-7. doi: 10.1093/nar/gku1180. Epub 2014 Nov 21

[6] Daniel Maticzka, Sita J Lange, Fabrizio Costa and Rolf Backofen: GraphProt: modeling binding preferences of RNA-binding proteins. Genome Biology 2014, 15:R17 doi:10.1186/gb-2014-15-1-r17

Kontakt:

Prof. Dr. Rolf Backofen
RBC Koordinator
Lehrstuhl für Bioinformatik
Institut für Informatik
Albert-Ludwigs-Universität Freiburg
Georges-Köhler-Allee 106, 79110 Freiburg
backofen@informatik.uni-freiburg.de
www.bioinf.uni-freiburg.de

Dr. Björn Grüning
Leiter des Freiburg Galaxy Teams
Lehrstuhl für Bioinformatik
Institut für Informatik
Albert-Ludwigs-Universität Freiburg
Georges-Köhler-Allee 106, 79110 Freiburg
gruening@informatik.uni-freiburg.de
galaxy@informatik.uni-freiburg.de
http://galaxy.uni-freiburg.de/

Foto Herr Backofen: R. Backofen; Foto Herr Grüning: B. Grüning

BRENDA

Von der Datenbank zum Kompetenzzentrum

von Dietmar Schomburg und Ida Schomburg

Enzyme – essentiell für die allermeisten Prozesse des Lebens, unverzichtbar für die industrielle Biotechnologie oder die medizinische Diagnostik. 30 - 40 % der Gene codieren Enzyme, beschleunigen chemische Reaktionen um bis zu 16 Größenordnungen, ermöglichen die fein aufeinander abgestimmten Stoffwechselwege in Zellen und sind unentbehrlich z. B. für die Abwehr von Pathogenen. Manche von ihnen sind hochspezifisch, andere haben eine breite Spezifität. Die Funktionen von Enzymen hängen von vielen Eigenschaften ab, wie z. B. ihrer Sequenz, ihrer dreidimensionalen Struktur, ihrer Stabilität und ihren Interaktionen mit anderen Molekülen.

Um die Rolle eines Enzyms in Krankheitsprozessen oder seine Eignung für eine biotechnologische Anwendung zu verstehen, müssen viele und teils sehr heterogene Daten miteinander kombiniert werden.

Nur eine flexible Datenbank, ausgestattet mit modernen bioinformatischen Werkzeugen kann dies leisten (Chang *et al.,* 2015).

Von der reinen Datenbank hat sich BRENDA zunehmend zu einem Kompetenzzentrum für Enzym-Information entwickelt, bei der neben der Datenbankpflege Bioinformatik-Tools auf breiter Ebene entwickelt werden (Text- und Data-Mining, Genomannotation, Datenintegration, funktionelle Statistik, User-Interface Optimierung, Daten-Visualisierung, Maschinelles Lernen, usw.).

27 Jahre Finanzierung einer biochemischen Infrastruktur

Die Enzymdatenbank BRENDA wurde vor nun 27 Jahren an der damaligen Gesellschaft für Biotechnologische Forschung (GBF), heute Helmholtz-Zentrum für Infektionsforschung in Braunschweig als wissenschaftliche Infrastruktur ins Leben gerufen. Durch stetige Weiterentwicklung, Anpassung an sich ändernde wissenschaftliche Anforderungen, breite Datenintegration und kontinuierliche Aktualisierung hat sie sich zur weltweit größten Ressource für Enzyminformationen entwickelt. Monatlich nutzen mehr als 80.000 Wissenschaftler diese Datenbank. Das Herz der Datenbank umfasst etwa drei Millionen aus der Primärliteratur handannotierte Einzelwerte zu Enzymfunktionen, ihrer Spezifität, Stabilität, Struktur und ihrem Vorkommen. Jedes der

Abbildung 1: Datenkategorien in BRENDA

Nomenclature	Reaction & Specificity	Functional Parameters
Enzyme Names EC-Number	Pathway Catalysed Reaction Reaction Type Natural Substrates & Products Substrates & Products Inhibitors Cofactors Metals/Ions Activating Compounds Ligands Biochemicals Reactions	K_M Value k_{cat} /K_M Value K_I Value IC_{50} Value pI Value Turnover Number Specific Activity pH Optimum pH Range Temperature Optimum Temperature Range **K**inetic **EN**zyme **DA**ta
Organism-related information		
Organism Source Tissue Localization		
Isolation& Preparation		
Purification Cloned Expression Renatured Crystallization		
Stability	**Enzyme Structure**	**Disease, Engineering & Application**
pH Stability Temperature Stability General Stability Organic Solvent Stability Oxidation Stability Storage Stability	Sequence 3D-Structure Molecular Weight Subunits Posttranslational Modification Protein-Specific Search	Disease/ Diagnostics Engineering Application
		References
		References

Grafik: Prof. Dr. Dietmar Schomburg

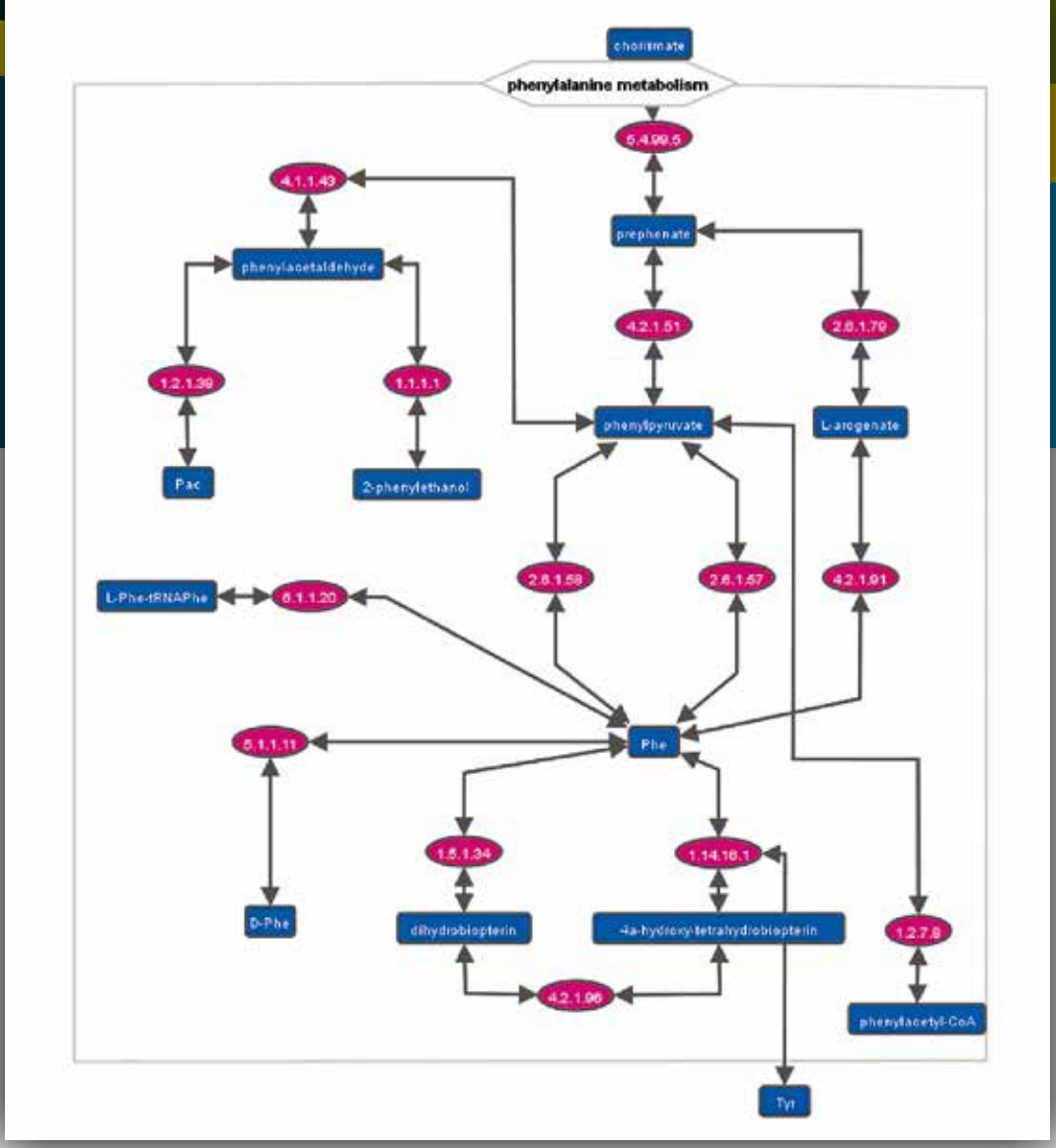


Abbildung 2: BRENDA Pathway des Phenylalaninstoffwechsels (Grafik: Prof. Dr. Dietmar Schomburg).

derzeit enthaltenen Enzyme ist anhand vieler Parameter umfassend charakterisiert (siehe Abbildung 1).

Mehrmals stand BRENDA vor dem finanziellen Kollaps. Eine vollständige Finanzierung war nur selten möglich und die Notwendigkeit der teilweisen Eigenfinanzierung war stets gegeben. So wurde BRENDA von 1987 bis 1996 von der GBF finanziert und in Buchform publiziert. Nach dem Wechsel von Professor Dietmar Schomburg an die Universität zu Köln wurde die Weiterentwicklung der Datenbank ca. ein Jahr lang ausschließlich aus diesen Buchhonoraren finanziert. In den Jahren 2000 bis 2012 gab es verschiedene Finanzierungen meist dank EU-Mitteln. Dies waren häufig Kooperationsprojekte mit dem EMBL-EBI. Im Jahr 2012 drohte wegen der Entscheidung der EU, keine wissenschaftliche Infrastruktur mehr zu finanzieren, das Aus. Glücklicherweise übernahmen das Bundesministerium für Bildung und Forschung und das Land Niedersachsen die weitere Finanzierung, und seit 2015 ist BRENDA ein Bestandteil des Deutschen Netzwerkes für Bioinformatik-Infrastruktur. Etwa 20 – 25 % der notwendigen Finanzmittel werden seit 2001 über einen Verkauf von In-Haus Lizenzen an die Industrie aufgebracht.

Daten aus vielen Quellen

Grundlage der BRENDA-Datenstruktur ist das Enzymklassifizierungssystem der *International Union of Biochemistry and Molecular Biology*. Es teilt alle Enzyme in sechs Hauptklassen ein. Unterklassen definieren weitere Details, wie Cofaktoren, an der Reaktion beteiligte Substrate und die Art der chemischen Bindung, die durch die Reaktion modifiziert wird. Derzeit gibt es etwa 6.500 Enzymklassen.

Enzyminformationen werden nicht in zentralen Repositorien hinterlegt, sondern sind in unstrukturierter Form in Millionen Publikationen in der Primärliteratur vergraben. Um diesen Schatz zu heben, müssen Literaturstellen gesichtet und ihre Bedeutung für die Charakterisierung eines Enzyms beurteilt werden. Dann müssen die im Artikel enthaltenen Daten manuell extrahiert und sorgfältig durch Computerprogramme und Wissenschaftler geprüft werden, bevor sie schließlich in die Datenbank integriert werden und mit anderen Daten, wie z. B. Sequenzen, Proteinstrukturen oder dem NCBI-Taxonomie-Baum verlinkt werden können. Das BRENDA-Team legt dabei sehr großen Wert auf einen hohen Qualitätsstandard.

BRENDA im Deutschen Netzwerk für Bioinformatik-Infrastruktur

BRENDA deckt im Deutschen Netzwerk für Bioinformatik-Infrastruktur (de.NBI) den großen Bereich der Enzyme ab und wird unter der Leitung von Prof. Dr. Dietmar Schomburg an der TU Braunschweig entwickelt. Über die Verknüpfung mit anderen de.NBI-Datenbanken konnte ein Ressourcen-Netzwerk geschaffen werden, das den Nutzern leichten Zugang zu verschiedensten mit-einander verbundenen Daten bietet.

BRENDA-Mitarbeiter arbeiten in zwei *„Special Interest Groups“* von de.NBI mit. Sie führen Kurse für die Wissenschaftsgemeinschaft und für Mitarbeiter der Zentren durch und stellen die umfangreichen und qualitativ hochwertigen BRENDA-Daten für vielfältige Anwendungen der Wissenschaft zur Verfügung.

Eine vollständige Auswertung der ca. 2,5 Millionen relevanten biomedizinischen Fachartikel ist aber nicht möglich. Um BRENDA aktuell zu halten, müssen pro Jahr mindestens 5.000 – 10.000 neue Publikationen manuell ausgewertet werden.

Die handannotierten Daten werden um Informationen ergänzt, die mittels Text-Mining-Methoden generiert werden. Dazu gehören medizinische Aspekte, kinetische Daten, das Vorkommen der Enzyme in den Organismen oder ihre Aktivität in Organen (Schomburg *et al.,* 2013). Weitere Daten werden rechnergestützt vorhergesagt, wie das folgende Kapitel skizziert. Auf diese Weise entsteht eine komplexe Datenbank mit mehr als 150 Millionen Einzeldaten mit unterschiedlichen Datenstrukturen (Texte, numerische Daten, Ontologien, graphische Objekte, Netzwerke).

Rechnergestützte Vorhersagen

Eine Reihe von Daten wird mit bioinformatischen Methoden berechnet und in BRENDA integriert. Hierzu gehören die Analyse des Zusammenhangs von Enzymfunktion oder -Fehlfunktion und Krankheit oder die Verwendung von Enzymen in Diagnostik und Behandlung von Krankheiten.

Hierfür wird ein initialer Text-Mining Schritt der Fachartikel-Abstracts kombiniert mit einer Einteilung des Papers in eine von drei Gruppen (*Causal Interaction, Treatment* oder *Diagnosis*) mittels Methoden des maschinellen Lernens (Söhngen *et al.*, 2011).

Für die sequenzbasierte Enzymfunktionsvorhersage wurde eine Methode entwickelt, die Vorhersagen der wichtigsten Genom-Annotations-Webseiten mit eigenen Berechnungen integriert und so die Vorhersagegenauigkeit erhöht.

Ein dritter Bereich betrifft die Lokalisierung der Enzyme in der Zelle. So wird z. B. die Membranständigkeit und die Zahl der Transmembranhelices für Enzyme sequenzbasiert vorhergesagt.

Heterogene Daten erfordern vielfältige Suchmöglichkeiten

Um den Nutzern einen flexiblen Zugang zu den Daten zu ermöglichen, wurden viele Recherchemöglichkeiten entwickelt. Die einfache Eingabe eines Stichwortes wie z. B. der Name eines Enzyms oder Metaboliten genügt für einen ersten Überblick. Enzyme, die bestimmte Substrate umsetzen oder bei definierten Temperaturen oder pH-Werten stabil sind, können gezielt auch in Kombination gesucht werden. Identifikatoren wie die EC-Nummer (*Enzyme Classification*) oder Referenznummern von Proteinstrukturen können ebenfalls genutzt werden.

BRENDA enthält aber nicht nur Text und numerische Daten, sondern auch Strukturen. Um diese durchsuchbar zu machen, wurden spezielle *„Query-Engines“* entwickelt. Metabolite und andere kleine Moleküle können mit einem einfachen Struktur-Editor gezeichnet und über ihre Struktur gesucht werden.

Neben Liganden haben auch metabolische Pfade eine Struktur. In BRENDA kann sich der Nutzer diese Pfade anzeigen lassen und Informationen aufrufen (siehe Abbildung 2).

Baumstrukturen finden sich z. B. beim Organismus-Taxonomie-Baum oder bei Ontologien. Die Nutzer können sich entlang des Baumes bewegen und bekommen für jede Stufe relevante Informationen angezeigt.

Eine besondere Funktion stellt die BRENDA Tissue Ontology (BTO) dar, die speziell für BRENDA entwickelt werden musste (Gremse *et al.*, 2011). Sie repräsentiert eine hierarchische Einordnung von Organen, Körperteilen und Gewebestrukturen, jeweils versehen mit einer Definition und Literaturreferenzen. Die BTO umfasst Begriffe aus dem Tier- und Pflanzenreich.

Datenvisualisierung – essentiell bei komplexen Datenstrukturen

Rechercheergebnisse werden tabellarisch dargestellt. Die primäre Informationseinheit ist dabei die EC-Nummer mit der dazu gehörigen chemischen Reaktion, dem Ursprungsorganismus und der Literaturreferenz. Numerische Ergebnistabellen lassen sich statistisch auswerten.

Innerhalb der Ergebnistabellen führen Links zu Molekül- und Proteinstrukturen, zu Sequenzen, Ontologien sowie zu Werkzeugen für weitere Analysen. Enzymstrukturen können unter Kennzeichnung ihrer funktionell wichtigen Bereiche dreidimensional und interaktiv frei drehbar dargestellt werden. Strukturdiagramme zeigen die katalysierte Reaktion oder die Strukturen von Inhibitoren und Aktivatoren.

Die ungeheure Datenmenge, die in BRENDA gespeichert ist, erschwert es, einen intuitiven Überblick über Eigenschaften und Bedeutung eines Enzyms zu bekommen. Als Hilfestellung wurden *„Word Maps“* entwickelt, die zeigen, mit welchen Begriffen bestimmte Enzyme in der Literatur assoziiert sind. Abbildung 3 veranschaulicht das für die HIV-Protease bzw. für das typisch biotechnologische Enzym Glucose Oxidase.

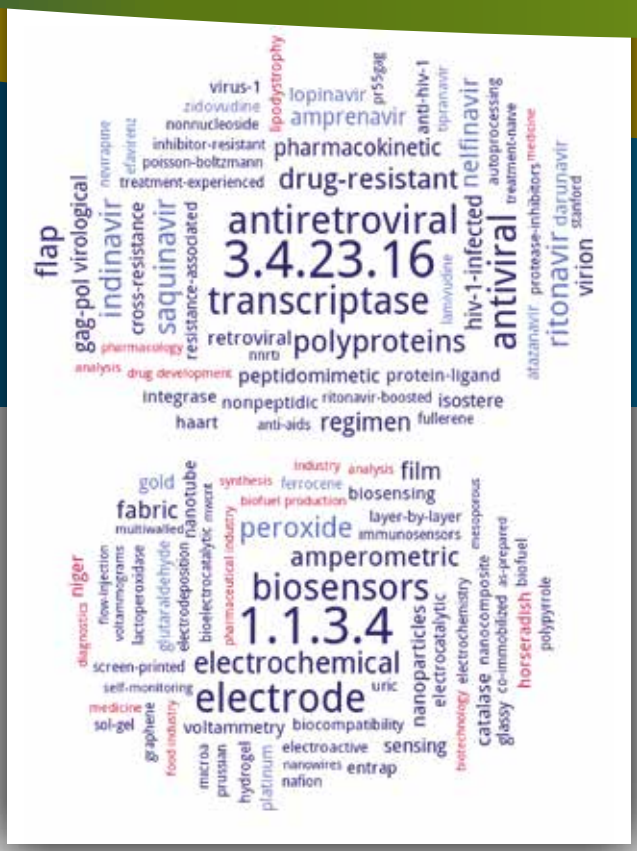


Abbildung 3: Word Maps für Glucose Oxidase und HIV Protease
(Grafik: Prof. Dr. Dietmar Schomburg)

Referenzen:

Chang A., Schomburg I., Placzek S., Jeske L., Ulbrich M., Xiao M., Sensen C.W., and Schomburg D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. Nucleic Acids Res. 43, D439-46

Gremse M., Chang A., Schomburg I., Grote A., Scheer M., Ebeling C., and Schomburg D. (2011) The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. Nucleic Acids Res., 39, D507-513

Schomburg I., Chang A., Placzek S., Söhngen C., Rother M., Lang M., Munaretto C., Ulas S., Stelzer M., Grote A. Scheer M., and Schomburg D. (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. 41, D764-772

Söhngen C., Chang A., and Schomburg D. (2011) Development of a classification scheme for disease-related enzyme information. BMC Bioinformatics, 12, 329

Kontakt:

Prof. Dr. Dietmar Schomburg
Abteilung Bioinformatik und Biochemie
Institut für Biochemie, Biotechnologie und Bioinformatik
Technische Universität Braunschweig
d.schomburg@tu-bs.de

Dr. Ida Schomburg
Enzymeta GmbH
Erftstadt
I.Schomburg@enzymeta.de

www.brenda-enzymes.org

Foto Herr D. Schomburg: D. Schomburg; Foto Frau I. Schomburg: I. Schomburg

NBI-SysBio – der de.NBI datenmanagement knoten

Vom spezialisierten Daten-Silo zu vernetzten Daten

von Wolfgang Müller

Der de.NBI Daten-Management Knoten hat das Ziel, den Systembiologie-Zyklus an einem wichtigen Punkt zu unterstützen: Dem Datenmanagement. Die beiden beteiligen Gruppen, Wolfgang Müller von der HITSgGmbH und Dagmar Waltemath von der Universität Rostock beschäftigen sich seit Jahren mit der Speicherung und Wiederverwendung von systembiologischen Daten, Modellen und Simulationsexperimenten. Hierzu wird das Daten-Management-System SEEK (entwickelt in einer langfristigen Kooperation mit der Gruppe von Carole Goble, Universität Manchester) sowie die Datenbank für reaktionskinetische Daten, SABIORK, eingesetzt und um Modellmanagementlösungen erweitert. Ein weiterer wichtiger Aspekt von NBI-SysBio ist die Entwicklung von Standards für die reproduzierbare Speicherung von Modellen und Simulationsexperimenten.

Im deutschen Bioinformatik-Netzwerk de.NBI bietet der Daten-Management-Knoten allen Nutzern der unterschiedlichen Datenservices eine effiziente und zuverlässige Datenverwaltung an. In SEEK können die erhobenen Daten und Resultate gemeinschaftlich mit Projektpartnern innerhalb und außerhalb de.NBIs organisiert, aufbereitet und miteinander in Beziehung gesetzt werden.

Systembiologisches Datenmanagement als Herausforderung

Die Systembiologie ist ein arbeitsteiliger Prozess, in dem WissenschaftlerInnen verschiedener Fachrichtungen zusammenarbeiten, um Informationen über biologische Systeme zu sammeln, diese mit weiteren Daten zu integrieren und als Basis zur Erstellung bzw. Verfeinerung von Modellen zu verwenden. Aus diesen Modellen können wiederum Hypothesen erzeugt werden, die experimentell getestet werden können.

Man kann diesen Prozess als Anreicherung von *„Big Data“* zu Wissen sehen. Aus großen experimentellen Datensätzen werden Computermodelle, welche dann in einer Publikation beschrieben und diskutiert werden: Konzentriertes Wissen über die Vorgänge im biologischen System.

Der Begriff *Big Data* bezieht sich aber nicht mehr nur auf große einzelne Datensätze. Er bezieht sich auch auf Gebiete, in denen *viele unterschiedliche* Datensätze und Datenformate verarbeitet werden müssen. Genau dies ist in der Systembiologie der Fall. Die immense Vielfalt an experimentellen Methoden führt zu einer ebensolchen Vielfalt an Datensätzen, Datenformaten, Guide-lines und Vokabularen zur Speicherung von Daten und SOPs (*Standard Operating Procedures*). **BioSharing.org** , eine Ko-Entwicklung von den MIBBI- und ISA-Teams listet beispielsweise **621** relevante Standards für die Biologie und Medizin.

Die Diversität biologischer Fragestellungen und das Wachstum der Menge verfügbarer Daten führten in den vergangenen Jahren zu einer Verteilung der Daten auf große Mengen spezialisierter Datenbanken (NAR listet mehr als 1.000 [Galperin, MY and Cochrane, GR, 2009]). In der Regel liegen also die in einem Projekt gewonnenen Daten, Modelle und SOPs verstreut vor. Die Vielzahl an Datenformaten, Standards und Speichersystemen bildet ein Dickicht, das für den Nicht-Datenmanager zunehmend schwieriger zu durchschauen ist.

Das Angebot des Datenmanagement–Knotens im Rahmen von de.NBI

Der Datenmanagement Knoten will dieses Dickicht im Sinne der Nutzer entwirren und nutzbar machen. Hierzu entwickeln und warten die Wissenschaftler Software, stellen bestimmte Datenmanagementdienste bereit (*Software as a Service*, SaaS), partizipieren in verschiedenen Standardisierungsgremien, und beraten die Nutzer schließlich direkt. Ferner führen sie Schulungen durch oder nehmen als Dozenten an Schulungen teil.

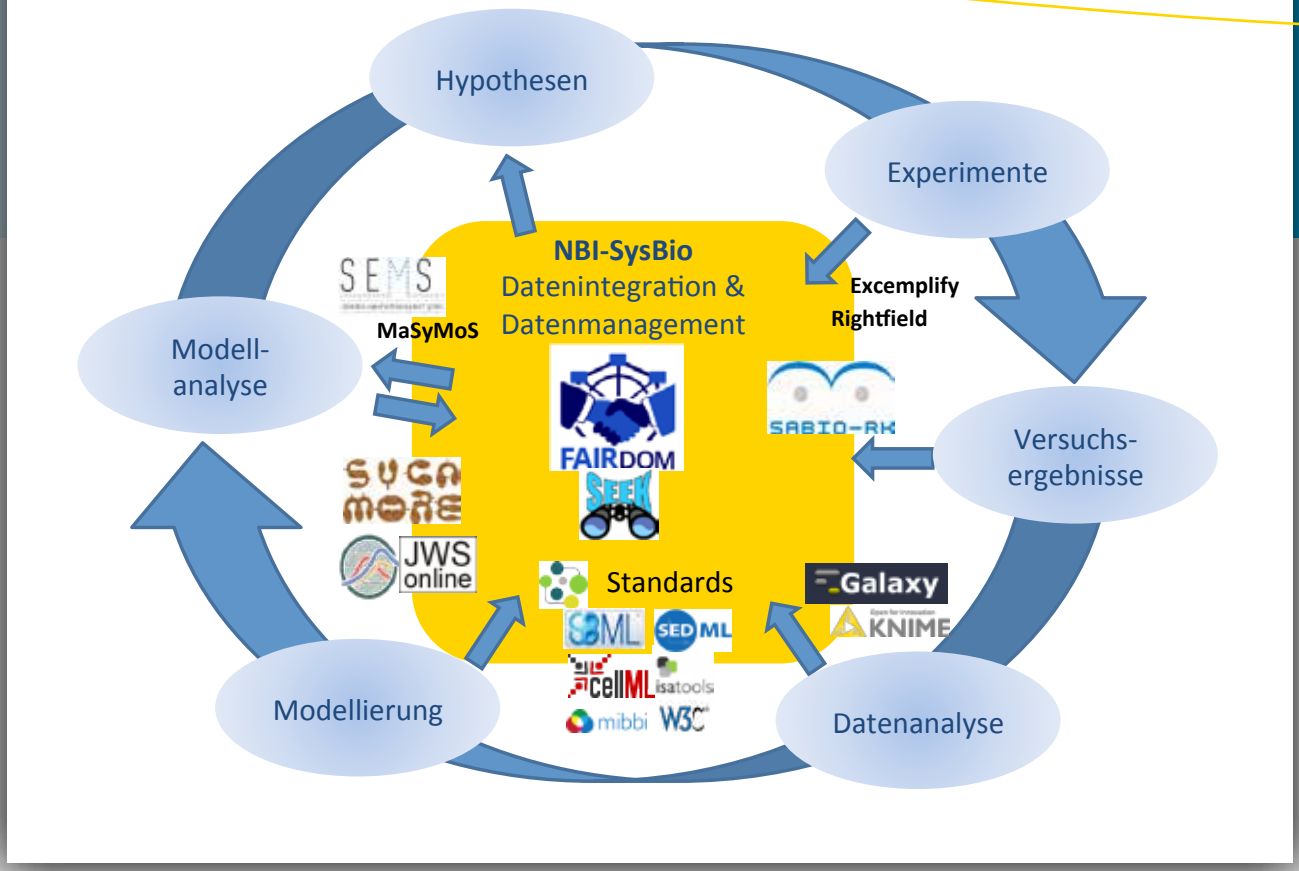


Abbildung 1:
NBI-SysBio unterstützt durch seine Datenmanagement-Dienste die Zusammenarbeit verschiedener Projektpartner im gesamten Zyklus der Systembiologie. Experimentelle Bedingungen (z. B. *Standard Operating Protocols*), Daten oder Modelle können standardisiert in openSEEK (beispielsweise bei http://fairdomhub.org) katalogisiert und miteinander projekt- und datenquellenübergreifend verknüpft werden.

Zusätzliche Werkzeuge wie Excemplify, Rightfield und SABIO-RK helfen bei der Strukturierung der Daten zur späteren Nutzung.

Die Verwendung von Standards wie SBML und SED-ML ermöglicht die Wiederverwendung von in openSEEK gespeicherten Modellen. Darüberhinaus unterstützen weitere Plattformen und Bibliotheken, z. B. JWS Online, Sycamore oder MaSyMos, die Nutzer beim Auffinden, Simulieren und Reproduzieren der in openSEEK gespeicherten Simulationsstudien (Grafik: NBI-SysBio Team).

Die **openSEEK Software** und seine größte öffentlich zugängliche Instanz, **FAIRDOM Hub**, adressieren den Misstand, dass strukturierte *Repositories* meist nach Methoden oder Datentypen, aber nicht nach Projekten organisiert sind. In OpenSEEK können Nutzer ihre Daten, Modelle und SOPs zentral speichern und miteinander in Beziehung setzen. Nutzer können also ohne spezielles Wissen über die darunterliegenden Datenquellen erkennen, wie die Daten, Modelle und SOPs der Projektpartner miteinander zusammenhängen. Die Verwendung standardisierter *Templates* erhöht weiterhin die Interoperabilität der gespeicherten Daten, und macht ganze Studien wiederverwendbar. Die dadurch an Bedeutung gewinnende Suche nach bestimmten Modellen, sowie die Versionierung und Nachvollziehbarkeit der Daten und Modelle in openSEEK sollen durch die Arbeiten des Daten-Management-Knotens im de.NBI Projekt weiter verbessert werden.

Die standardisierte Speicherung von Simulationsexperimenten innerhalb von openSEEK (im **SED-ML** Format) wird die Reproduzierbarkeit von Modellen verbessern. Wir entwickeln hier ein bestehendes Speicherkonzept (MASYMOS) weiter, welches die Simulationsmodelle und dazugehörigen Daten mit dazugehörigen Simulationsexperimenten und Modellparametrisierungen verlinkt, und passende Kombinationen aus Modellen und Experimenten bereitstellt. Dies klingt einfach, Nachlässigkeiten auf diesem Gebiet werden jedoch schnell zu einem Stolperstein für die Reproduzierbarkeit von Modellen.

SABIORK ist eine von Experten kuratierte Datenbank für reaktionskinetische Daten. Diese Daten sind insbesondere für quantitative Modelle der Systembiologie wichtig. Nutzer können hier entweder mit Freitext oder über eine Suchmaske für Fortgeschrittene nach Reaktionen suchen. Sie erhalten eine Liste von „Datenblättern" über Reaktionen. Die so aufbereiteten Informationen lassen sich viel schneller erfassen als deren Beschreibung in einer Publikation. Datenblätter enthalten Informationen über die kinetischen Daten, wie sie beispielsweise für quantitative Simulationen wichtig sind, und Verweise auf weitere Information, z. B. wird jeder ProteinIdentifier auf UniProt verlinkt. Die Daten können aus dem SABIO „Datenblatt" als in SBML kodierte Modellfragmente exportiert werden. Diese Fragmente können erweitert, kombiniert, parametrisiert und schließlich an Simulationsexperimente geknüpft werden.

Services & Trainings

Im Rahmen des de.NBI Netzwerkes werden unterschiedliche *Services* und *Trainings* angeboten:

1. **Nutzerberatung vor Ort oder per Telefon:** Die an dem Datenmanagement beteiligten Spezialisten besuchen Sie gern vor Ort, stellen viele Fragen über das Projekt und erarbeiten darauf basierend einen Datenmanagement-Plan: Was ist der beste Fluss für die Daten durch das Projekt? Wo sind Unklarheiten oder Bedarfe für interne Standardisierung bzw. neue Features? Kann der Einsatz weiterer Werkzeuge und Services aus dem de.NBI-Portfolio oder der internationalen Community die Arbeit erleichtern? Welche Bedarfe müssen zuerst adressiert werden?
2. **Hilfe bei der Entwicklung:** Hierzu gehört neben der Umsetzung von Wünschen die Unterstützung bei der eigenen Umsetzung.
3. **Kuratierung:** Auf sabio.hits.org können Kuratierungsbedarfe formuliert werden: Welche Veröffentlichungen, Pfadwege oder experimentellen Daten sollten kuratiert in SABIORK eingepflegt werden?
4. **Training:** Es werden derzeit verschiedene Workshops und Tutorials für de.NBI-Partner angeboten. Diese reichen von der Organisation eines *ICSB Tutorials in 2015* über die Durchführung einer Weiterbildung zum Thema Modellierung und Modellmanagement bei der *de.NBI Late Summer School* und der CoOrganisation eines ERASysAPP Workshops über *Reproduzierbare und Zitierbare Daten und Modelle*. Weiterhin führen wir offene Webinare zu den Themen Daten und Modellmanagement SABIORK, und Standardisierung durch. Am 30./31.5.2016 wollen wir in einem Workshop mit Training zehn Jahre SABIORK feiern.

Sie können über die de.NBI Homepage oder über nbi-sysbio@denbi.de Kontakt aufnehmen, falls Sie Beratung oder konkrete Hilfe wünschen.

Referenz:

Galperin, MY and Cochrane, GR (2009). Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009. Nucleic Acids Res. 2009 Jan;37(Database issue):D1-4. doi: 10.1093/nar/gkn942.

Wolstencroft K, Owen S, du Preez F, Krebs O, Mueller W, Goble CA, Snoep JL (2011). The SEEK: A Platform for Sharing Data and Models in Systems Biology, Methods in Enzymology, Volume 500: 629-655. doi: 10.1016/B978-0-12-385118-5.00029-3.

Kontakt:

Foto: HITS

Priv.-Doz. Dr. Wolfgang Müller
HITS gGmbH
Heidelberg Institute for Theoretical Studies
Heidelberg
wolfgang.mueller@h-its.org

Für Fragen zu unseren Services und benachbarten de.NBI-Diensten erreichen Sie uns unter: nbi-sysbio@denbi.de

FAIRDOM Hub: Verwaltung von systembiologischen Daten als Teil von de.NBI und dem transnationalen FAIRDOM-Projekt http://fairdomhub.org

SABIO-RK: Kuratierte Datenbank für systembiologische Daten http://sabiork.h-its.org

entwicklung von software-lösungen für mikrobielle bioinformatik

Institutsportrait

Bioinformatik Zentrum der Justus-Liebig-Universität Gießen

von Alexander Goesmann

Das Institut für Systembiologie der Justus-Liebig-Universität in Gießen unter der Leitung von Professor Alexander Goesmann ist Teil des de.NBI-Leistungszentrums Mikrobielle Bioinformatik (BiGi), an dem neben der Universität Gießen auch die Universität Bielefeld beteiligt ist. Im Fokus der Arbeiten des Leistungszentrums steht die Bereitstellung und Weiterentwicklung von Software-Lösungen zur Analyse des Erbguts von medizinisch und biotechnologisch relevanten Mikroorganismen. Darüber hinaus stellt das Leistungszentrum eine umfangreiche Hardware-Infrastruktur zur Verfügung, die im Rahmen der de.NBI-Förderung stufenweise erweitert wird.

Mit der Berufung von Alexander Goesmann im Jahr 2013 wurde das neue Institut an der Justus-Liebig-Universität (JLU) bereits mit einer exzellenten IT-Infrastruktur ausgestattet, die vom Land Hessen und von der JLU finanziert wurde. Diese Infrastruktur wird nun mit Hilfe der Fördermittel von de.NBI seit dem Start des Projektes im März 2015 weiter ausgebaut und an den Bedarf angepasst. Neben einem Rechencluster mit circa 1.000 Rechenkernen steht für Kooperationspartner auch Spezialhardware zur Verfügung, wie beispielsweise ein SMP-Server mit 160 Rechenkernen und zwei Terabyte Arbeitsspeicher sowie ein TimeLogic DeCypher-System zur Beschleunigung von Homologiesuchen. Mit diesem energieeffizienten System lassen sich pro Tag mehrere hunderttausend Sequenzen mit großen Datenbanken vergleichen. Eine Auflistung der im Rahmen von de.NBI geplanten Investitionen findet sich in Tabelle 1.

Das Institut für Bioinformatik und Systembiologie verfügt damit über Speicher- und Rechnerkapazitäten, die optimal auf die besonderen Anforderungen der Bioinformatik ausgelegt sind. In den kommenden Monaten sollen Teile der vorhandenen Ressourcen auch über eine Cloud-Computing-Infrastruktur zur Verfügung gestellt werden, sodass Nutzern eine noch größere Flexibilität zur Entwicklung eigener Werkzeuge geboten werden kann.

Tabelle 1: Stufenweiser Ausbau der Speicher- und Rechenkapazitäten

Im Rahmen der de.NBI-Förderung erfolgt ein stufenweiser Ausbau der Speicher- und Rechenkapazitäten an der JLU Gießen. In der Tabelle sind die zu beschaffenden Systeme mit ihrem jeweiligen Einsatzzweck aufgeführt.

2015	• TimeLogic DeCypher-System J1 • Datenbank-Server	Hochdurchsatz-Sequenzvergleiche Bereitstellung von Datenbank-Anwendungen
2016	• Erweiterung Speicherkapazität (360 TB) • Erweiterung Rechenkapazität (480 Cores)	Speicherung von Projektdaten Allgemeine Datenauswertung & Cloud Computing
2017	• TimeLogic DeCypher-System J1 • Web-Server	Hochdurchsatz-Sequenzvergleiche Bereitstellung von Web-Anwendungen
2018	• Erweiterung Speicherkapazität (300 TB) • Erweiterung Rechenkapazität (320 Cores)	Speicherung von Projektdaten Allgemeine Datenauswertung & Cloud Computing
2019	• Erneuerung SMP-Server (2 TB RAM)	Speicherintensive Anwendungen wie zum Beispiel Genomassemblierungen

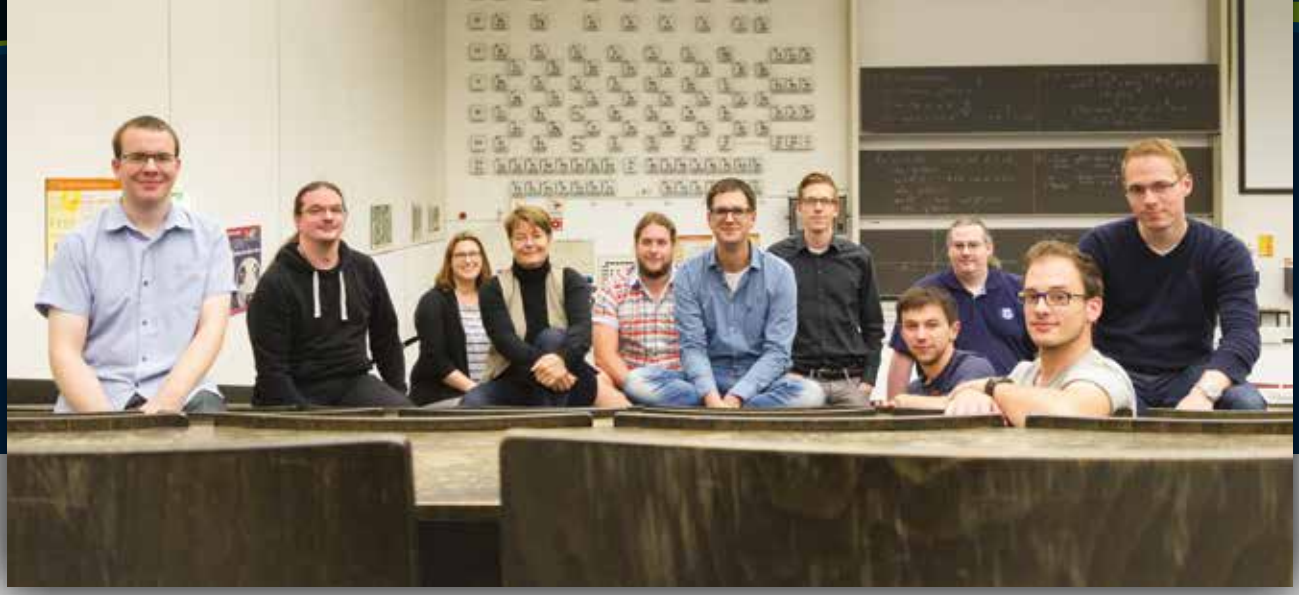


Abbildung 1: Arbeitsgruppe Prof. Goesmann
Die Mitarbeiterinnen und Mitarbeiter der Arbeitsgruppe von Prof. Goesmann entwickeln Software-Lösungen zur Analyse des Erbguts von medizinisch und biotechnologisch relevanten Mikroorganismen (Foto: Lukas Jelonek).

Im Deutschen Netzwerk für Bioinformatik-Infrastruktur stellt BiGi interessierten Nutzern aber auch zahlreiche Software-Pakete sowie regelmäßig aktualisierte öffentliche Datensammlungen zur Verfügung, die zur bioinformatischen Analyse von Genom- und Postgenomdaten benötigt werden. Dabei können Sequenzdaten einzelner Organismen aber auch Metagenomdaten analysiert werden. Zusätzlich bietet BiGi Anwendern Unterstützung im Umgang mit dieser Software und vermittelt den Umgang mit diesen Werkzeugen in speziell ausgerichteten Trainingskursen.

Zu den bekanntesten Software-Entwicklungen der Arbeitsgruppe gehört das Genomannotationssystem GenDB (Meyer *et al.*, 2003). GenDB ist eine webbasierte Plattform zur automatischen und manuellen Annotation, die Forscherteams weltweit in mehr als 400 Projekten zur gemeinschaftlichen Validierung, Verbesserung und langjährigen Verwaltung der Annotationsdaten prokaryotischer Genome nutzen.

Im Bereich der komparativen Genomik steht die Plattform EDGAR (Blom *et al.*, 2009) zur Verfügung, mit der sich die genetische Ausstattung verschiedener Bakterien anwenderfreundlich vergleichen lässt. Mit der Software lassen sich auf einfache Weise verschiedene Informationen über Gruppen von Genomen extrahieren, deren Gemeinsamkeiten und Unterschiede identifizieren und verwandtschaftliche Beziehungen phylogenetisch auswerten. Mit über 15.000 Genomanalysen in 500 Projekten stellt EDGAR eines der populärsten Werkzeuge im Feld der komparativen Genomik dar. Für noch unveröffentlichte Datensätze können wie bei allen hier dargestellten Software-Anwendungen auf Anfrage passwortgeschützte Projekte erstellt werden. Diese Möglichkeit haben seit Beginn der de.NBI-Förderung bereits mehr als 45 Kooperationspartner genutzt.

Liegen Genom- oder Transkriptomdaten aus einer Sequenzierung als nicht assemblierte Sequenzfragmente (Reads) vor, die zunächst an einem Referenzgenom angeordnet werden müssen, bietet sich die frei verfügbare Software ReadXplorer (Hilker *et al.*, 2014) zur Analyse und Visualisierung der Daten an. Zur Unterstützung einer möglichst benutzerfreundlichen Datenauswertung kombiniert die Software eine ganze Reihe von Analysemethoden und entsprechende Visualisierungsverfahren. Zu den wichtigsten automatischen Funktionen gehört die detaillierte Klassifizierung aller Reads, die Detektion genomischer Varianten, die Identifizierung und Validierung von Operons oder Transkriptionsstarts, sowie Analysen zur differentiellen Genexpression.

Der einfache Zugang zu Sequenzdaten aus Hochdurchsatztechnologien hat in den letzten Jahren mit der Metagenomik ein weiteres Themenfeld eröffnet, bei dem der Fokus auf der Erfassung und Analyse hochkomplexer mikrobieller Gemeinschaften liegt. Mit der MGX-Software stellt die Arbeitsgruppe von Goesmann ein benutzerfreundliches und hochflexibles Framework zur Analyse solcher Daten bereit. Aufbauend auf standardisierten *Conveyor-Workflows* (Linke *et al.*, 2011) ermöglicht MGX die effiziente Prozessierung metagenomischer Datensätze z. B. aus Umweltproben oder klinischen Isolaten. Umfangreiche Visualisierungen und statistische Methoden erleichtern die Auswertung von taxonomischen und funktionalen Analysen und erlauben eine einfache Interpretation der erzielten Ergebnisse.

Zu den herausragenden Ereignissen im letzten Jahr gehörte die erste de.NBI-Sommerschule an der JLU Gießen. 20 internationale Teilnehmer lernten an fünf Tagen verschiedenste Teilaspekte der Analyse genomischer Daten kennen. Ausgehend von der grundlegenden Qualitätskontrolle über die Assemblierung von Sequenzierdaten bis hin zur Genomannotation und komparativen Auswertung konnten die zunächst theoretisch vermittelten

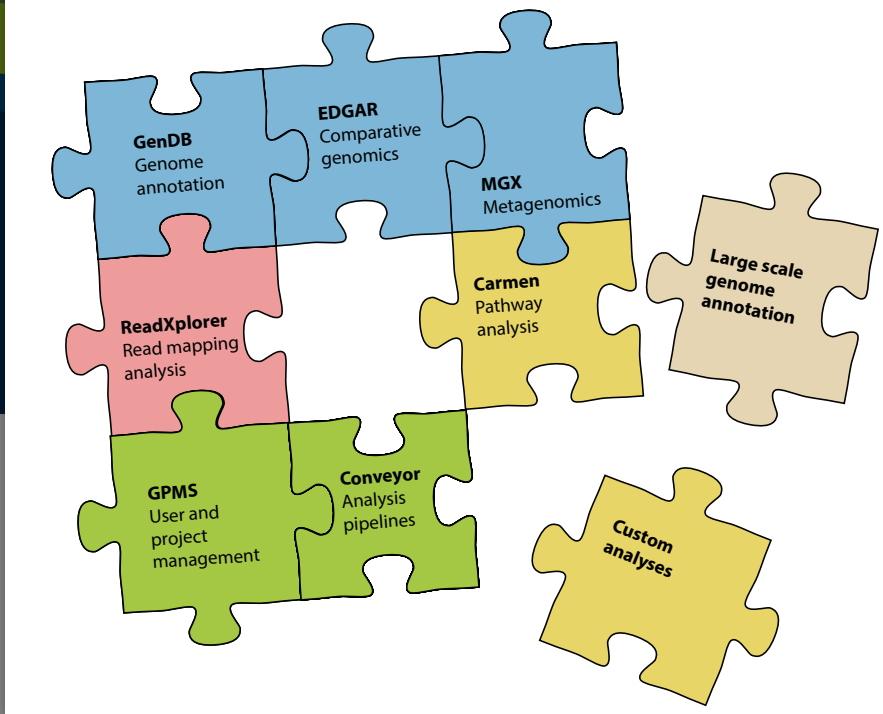


Abbildung 2: Die Abbildung zeigt verschiedene Software-Entwicklungen und deren Einsatzbereiche im Rahmen des de.NBI-Netzwerks. (Abbildung: Lukas Jelonek).

Fachkenntnisse schließlich auch an eigenen Datensätzen in der Praxis erprobt werden. Da mit Gene Myers, Tatiana Tatusova, Paul Kersey, Jan Gorodkin und Ursula Kummer ausgesprochen hochkarätige externe Vortragende gewonnen werden konnten, war auch das internationale Interesse an der de.NBI-Sommerschule entsprechend groß.

Im Fokus der aktuellen Arbeiten steht neben zahlreichen Verbesserungen an den oben dargestellten Software-Plattformen auch die weitergehende Automatisierung der Analyse-Workflows, um zukünftig auch tausende von Genomen in kürzester Zeit möglichst zielgerichtet auswerten zu können. Ein ehrgeiziges Ziel dabei ist es, 1.000 Genome pro Tag analysieren zu können. Dazu muss einerseits für den Benutzer eine möglichst einfache Spezifikation aller Daten und der zugehörigen Arbeitsschritte ermöglicht werden. Andererseits muss ein modularer Baukasten von Analysewerkzeugen und bedarfsorientierten Visualisierungsmethoden entwickelt werden, um maßgeschneiderte Ergebnisberichte generieren zu können.

Interessierte Nutzer finden weitere Informationen und Zugang zu der zur Verfügung stehenden Software sowie zu den angebotenen Trainingskursen unter http://bigi.computational.bio.

Eine erste Kontaktaufnahme kann über die Email-Adresse bigi@computational.bio erfolgen.

Autoren

Dr. Jochen Blom, Dr. Karina Brinkrolf, Dr. Rolf Hilker, Sebastian Jaenicke, Lukas Jelonek, Dr. Burkhard Linke, Oliver Rupp, Oliver Schwengers, Prof. Dr. Alexander Goesmann

Referenzen:

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O., Giegerich, R. and Pühler, A. (2003) GenDB - an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. 31(8):2187-95.

Blom, J., Albaum, S., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M. and Goesmann, A. (2009) EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinformatics 10(1): 154, 2009.

Hilker, R., Stadermann, K.B., Doppmeier, D., Kalinowski, J., Stoye, J., Straube, J., Winnebald, J. and Goesmann, A. (2014) ReadXplorer - Visualization and Analysis of Mapped Sequences. Bioinformatics, 30, 2247–2254.

Linke, B., Giegerich, R. and Goesmann, A. (2011) Conveyor: a workflow engine for bioinformatic analyses. Bioinformatics 27(7), 903–911.

Kontakt:

Prof. Dr. Alexander Goesmann
Systembiologie mit dem Schwerpunkt Genomik, Proteomik und Transkriptomik
Justus-Liebig-Universität Gießen
Gießen
Alexander.Goesmann@Computational.Bio.Uni-Giessen.DE

http://computational.bio

Foto: Björn Fischer

ELIXIR – damit der datenfluss nicht ins stocken kommt

Interview mit Niklas Blomberg

Seit Mai 2013 ist Niklas Blomberg Direktor von ELIXIR, einer neuen Forschungsinfrastruktur auf der ESFRI-Roadmap (European Strategy Forum for Research Infrastructures). ELIXIR hat sich zum Ziel gesetzt, europaweit ein Netzwerk aufzubauen, das die einzelnen nationalen Bioinformatik-Strukturen untereinander verknüpft. Anfang 2016 waren bereits 16 europäische Länder Mitglied im ELIXIR-Konsortium.

Systembiologie.de: *Herr Dr. Blomberg, können Sie uns in wenigen Sätzen beschreiben was ELIXIR ist?*

Dr. Niklas Blomberg: ELIXIR ist eine EU-weite Forschungsinfrastruktur, welche nationale Bioinformatik-Kapazitäten – von Forschungslaboren bis zu Datenzentren – in Europa miteinander vernetzt. Für die gesamte Forschungsgemeinschaft werden verschiedenste Angebote bereitgestellt. Dazu gehören beispielsweise analytische Werkzeuge für Bioinformatik und die Forschung mit lebenswissenschaftlichen Daten und Datenbanken sowie die Bereitstellung von Methoden für sicheren Datenaustausch. All diese Instrumente sollen dabei helfen, die einzelnen Datensammlungen, ihre Qualitätssicherung als auch das Aufbewahren der großen Datenmengen aus dem lebenswissenschaftlichen Bereich zu koordinieren. Das ist sehr wichtig, denn ohne gutes Netzwerk geht es nicht: Der Datenfluss kommt ins Stocken!

Wer beteiligt sich an ELIXIR?

Der ELIXIR Hub, also die zentrale Schalt- und Koordinierungsstelle von ELIXIR, befindet sich im Vereinigten Königreich und ist neben dem EMBL-EBI – dem European Molecular Biology Laboratory´s European Bioinformatics Institute – in Hinxton, Cambridge angesiedelt. Die ELIXIR Nodes – die nationalen Knotenpunkte – verteilen sich über die einzelnen Mitgliedsstaaten und sind dort jeweils an führenden biowissenschaftlichen Instituten angesiedelt.

Wie finanziert sich ELIXIR?

ELIXIR wird überwiegend aus Beiträgen seiner Mitglieder finanziert, also der Länder, die das sogenannte ELIXIR Consortium Agreement unterschrieben haben. Das Budget von 2015, welches auf der Basis des vom ELIXIR Board beschlossenen Programms berechnet wird, betrug beispielsweise rund zwei Millionen Euro. Über geförderte Forschungsprojekte kommen von der Europäischen Kommission dann noch andere Beträge dazu. Die Höhe der Mitgliedsbeiträge orientiert sich an der Wirtschaftsleistung des jeweiligen Landes, also am Nettonationaleinkommen (NNE). Im Falle von Deutschland entspräche dies beispielsweise für das Jahr 2016 einem Mitgliedsbeitrag von ca. 900.000 EUR. Die Gelder werden fast ausschließlich für die Unterstützung von relevanten Vorhaben der Mitglieder verwendet.

Inwieweit können auch Nicht-Mitglieder die ELIXIR-Angebote nutzen? Oder anders gefragt: Welche Vorteile und zusätzlichen Möglichkeiten bietet eine ELIXIR-Mitgliedschaft?

Viele der Dienstleistungen, die der Forschungsgemeinschaft im Rahmen von ELIXIR angeboten werden, sind kostenfrei und unabhängig von einer ELIXIR-Mitgliedschaft. Sie können von allen Wissenschaftlern in Europa genutzt werden, egal ob ihr Land Mitglied bei ELIXIR ist oder nicht.

Eine der großen heutigen Herausforderungen in der dateninstensiven Forschung im lebenswissenschaftlichen Bereich ist allerdings die Bereitstellung von sicheren und effektiven Netzwerken für Datenaustausch und Analyse: Dies ist nicht immer kostenfrei. Nur ELIXIR-Mitglieder haben die Möglichkeit, an der Gestaltung der Infrastruktur teilzuhaben und dazu beizutragen, wo entsprechende Schwerpunkte für die zukünftige Entwicklung gesetzt werden. Zudem stehen Mitgliedern zum Beispiel vielfältige Weiterbildungen, die von ELIXIR angeboten werden, zur Verfügung

Welche Rolle spielt das EMBL-EBI als Gastgeber für den ELIXIR Hub?

ELIXIR nutzt die Rechtspersönlichkeit des EMBL als eine internationale Organisation, und der ELIXIR Hub ist auf dem Wellcome

ELIXIR-Direktor Niklas Blomberg vernetzt nationale Bioinformatik-Kapazitäten in Europa miteinander (Foto: Niklas Blomberg/ELIXIR).

Genome Campus in Hinxton neben dem EMBL-EBI angesiedelt. Das EMBL-EBI ist wiederum selbst ein ELIXIR Node, einer der Knotenpunkte von Bioinformatik-Ressourcen. In diesem Sinne ist das Verhältnis von ELIXIR zum EMBL-EBI ähnlich wie das zu einem der ELIXIR Mitgliedsstaaten. Natürlich hält das EMBL-EBI signifikante Datenressourcen, deren Unterhaltung von den EMBL-Mitgliedsstaaten finanziert wird. Der zusätzliche Gewinn im Rahmen von ELIXIR für das EMBL-EBI, wie für alle Nodes, ist die internationale Vernetzung dieser Ressourcen mit anderen nationalen Ressourcen, die in den jeweiligen ELIXIR Mitgliedsstaaten unterhalten werden, und die Verbundeffekte, die daraus entstehen.

Was sind die Vorteile dieser Europäischen Infrastruktur gegenüber z. B. der US-amerikanischen Infrastruktur NCBI?

Das NCBI – das National Center for Biotechnology Information – ist kein verflochtenes Netzwerk, sondern ein Monolith, ähnlich wie das EMBL-EBI. Man kann es daher nicht mit einer vernetzten, internationalen Infrastruktur wie ELIXIR vergleichen, welche nationale Ressourcen auf internationaler Ebene zusammenbringt und koordiniert. Bei ELIXIR geht es darum, eine Verbindung bestehender Strukturen zu schaffen. Natürlich brauchen wir das EMBL-EBI, ebenso wie das NCBI. Wir brauchen aber auch die Verbindung zwischen den vielen weiteren Instituten in den ELIXIR Mitgliedsstaaten mit dem EMBL-EBI, ebenso wie die Verbindung zwischen ELIXIR und NCBI.

Sehen Sie Synergien zur anderen europäischen Infrastruktur-Initiativen, wie z. B. EUDAT2020, das europaweit Forschungsdaten verfügbar machen möchte, oder zur europäischen Supercomputing Community PRACE?

Ja, auf jeden Fall! Von den europäischen Strukturen ist der GÉANT-Verbund vielleicht die bedeutendste Initiative für uns, denn gute Datenflüsse sind sehr wichtig. Lassen Sie mich das am Beispiel Roaming erläutern: Wir arbeiten bei ELIXIR intensiv daran, uns mit anderen Netzwerken wie beispielsweise education roaming (eduroam) zu verbinden. Das erfordert sehr wenig Technik, zudem sind viele technische Möglichkeiten hier schon weit fortgeschritten. Dennoch sind sehr viele Abstimmungsprozesse erforderlich. Die größte Aufgabe besteht hier also in der Governance.

Zum Abschluss noch eine persönliche Frage: Was reizt Sie an Ihrem Job ganz besonders?

Bevor ich zu ELIXIR gekommen bin, habe ich eine Arbeitsgruppe bei AstraZeneca geleitet. Das war eine interessante und spannende Zeit. Hier bei ELIXIR mache ich jetzt eine Arbeit, die ich für wirklich wichtig halte - wenn wir erfolgreich sind, dann wird das eine richtig gute Sache, die einen weitreichenden, positiven Einfluss auf die Forschung mit Daten aus der Lebenswissenschaft haben kann.

Das Gespräch führten Marcus Garzón und Vera Grimm vom Projektträger Jülich.

Kontakt:

Dr. Niklas Blomberg
ELIXIR Director
ELIXIR Hub
Wellcome Genome Campus
Hinxton, Cambridgeshire, CB10 1SD, UK
niklas.blomberg@elixir-europe.org

www.elixir-europe.org

omics-infrastrukturen für forschung und lehre

Ein Konzept der Leopoldina zum Umbruch in den Lebenswissenschaften

Von Alfred Pühler

Entstehung des Zukunftsreports

Die Nationale Akademie der Wissenschaften Leopoldina sieht es als eine ihrer Hauptaufgaben an, sowohl Politik als auch Öffentlichkeit in wissenschaftlichen Fragen zu beraten. Eine solche Beratung erfolgt meistens durch die Herausgabe unabhängiger Stellungnahmen, die sich entweder einzelnen Fachdisziplinen widmen oder das gesamte Wissenschaftssystem in Deutschland im Blick haben. Auf dem Gebiet der Lebenswissenschaften hat die Leopoldina dazu eine Publikationsreihe mit dem Titel „Zukunftsreport Wissenschaft“ ins Leben gerufen. Der erste Zukunftsreport Wissenschaft nimmt Bezug auf den Paradigmenwechsel in den Lebenswissenschaften, der sich vor Kurzem aufgrund der Omics-Technologien vollzogen hat. Er analysiert die Herausforderungen für Forschung und Lehre und entwickelt ein Konzept, wie dieser Umbruch bewältigt werden kann.

Zur Erstellung des ersten Zukunftsreports wurde eine Arbeitsgruppe aus Omics-Experten gebildet, die in mehreren Arbeitssitzungen Vertreter aus Universitäten, außeruniversitären Einrichtungen und Ministerien zu Fachgesprächen einluden. Diese Arbeitsgruppe formulierte anschließend den ersten Zukunftsreport Wissenschaft mit dem Thema „Lebenswissenschaften im Umbruch - Herausforderungen der Omics-Technologien für Deutschlands Infrastrukturen in Forschung und Lehre“. Der erste Zukunftsreport Wissenschaft wurde im Mai 2014 nach einer ausführlichen Begutachtung durch acht Experten aus dem In- und Ausland vom Präsidium der Leopoldina verabschiedet. Im September 2014 wurde er in einer Pressekonferenz und anschließend in einer öffentlichen Veranstaltung Vertretern aus Politik und Öffentlichkeit vorgestellt (Abb. 1).

Paradigmenwechsel in den Lebenswissenschaften durch Omics-Technologien

Das molekulare Verständnis von zellulären Vorgängen wurde insbesondere nach der Entdeckung der DNA als Erbsubstanz enorm beschleunigt. So konnten einzelne Gene, Transkripte, Proteine und Metabolite intensiv molekularbiologisch analysiert werden. Diese isolierte Betrachtungsweise von zellulären Einzelbausteinen wurde erst nach der Etablierung der Omics-Technologien von einer ganzheitlichen Sicht abgelöst. Insbesondere konnten auf dem Gebiet der DNA-Sequenzierung Technologieentwicklungen eingeschlagen werden, die es heute erlauben, alle Gene beliebiger Organismen zu erfassen und damit einer ganzheitlichen Betrachtung zugänglich zu machen. In jüngster Zeit wurde darüber hinaus die RNA-Sequenzierung perfektioniert, so dass die Gesamtheit aller transkribierten Gene in die Analyse einbezogen werden kann. Neben Genomik und Transkriptomik wurden auch die Omics-Technologien Proteomik und Metabolomik optimiert, so dass nahezu alle Proteine und viele Metabolite einer Zelle hochparallel erfasst werden können. Die Entwicklung all dieser Omics-Technologien wäre allerdings ohne die entscheidenden Beiträge einer anwendungsorientierten Bioinformatik nicht denkbar gewesen, denn Omics-Technologien produzieren große Datenmengen, deren Analyse nur mittels spezifizierter bioinformatischer Werkzeuge erfolgen kann. Damit ist der Einsatz der angewandten Bioinformatik aber nicht erschöpft. Sie kommt speziell bei der sich heute entwickelnden Systembiologie zellulärer Vorgänge nochmals prominent ins Spiel und hilft, den Zusammenhang zwischen identifizierten zellulären Bestandteilen systembiologisch zu erfassen.

Damit ist der Paradigmenwechsel durch Omics-Technologien bereits detailliert beschrieben. Er beruht auf dem Wechsel der Analyse von einzelnen zellulären Bestandteilen hin zu einer ganzheitlichen Betrachtung des zellulären Geschehens. Omics-

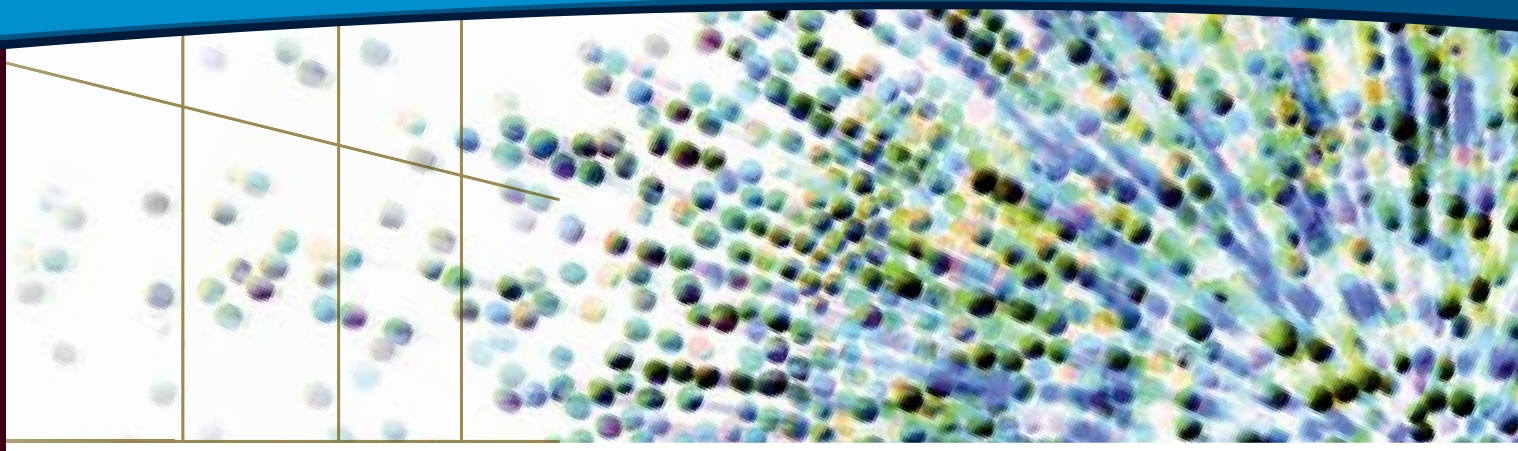
Abbildung 1: Vorstellung des Zukunftsreports anlässlich einer öffentlichen Veranstaltung im September 2014.
Abgebildete Personen von links nach rechts: D. Scheel (Halle), A. Pühler (Bielefeld), R. Kahmann (Marburg), M. Hecker (Greifswald) und R. Eils (Heidelberg) (Foto: David Ausserhofer/Leopoldina).

Technologien haben offensichtlich die Grundlagenforschung in den Lebenswissenschaften revolutioniert und in der Zwischenzeit auch die angewandten Lebenswissenschaften erreicht. Speziell in der Medizin wird eine maßgeschneiderte Diagnostik und Therapie angeboten, die auf Biomarkern aus Omics-Verfahren beruht. Diese Entwicklung zielt auf die Etablierung einer personalisierten oder individualisierten Medizin. Auch in der Biotechnologie werden Omics-Verfahren zur Entwicklung von maßgeschneiderten bakteriellen Produktionsstämmen für den industriellen Einsatz verwendet. Dabei wird das in früheren Jahren so erfolgreich praktizierte Mutations- und Selektionsverfahren durch eine rationale Stammentwicklung ersetzt, die hauptsächlich von einer genombasierten Systembiologie geprägt ist. Aber auch in der Pflanzenforschung werden Omics-Verfahren für die Entwicklung von robusten und ertragreichen Pflanzen eingesetzt. Dabei steht die Metabolomik im Vordergrund, mit der Metabolite identifiziert werden können, die als metabolische Marker in Züchtungsprogrammen genutzt werden können.

Herausforderungen der Omics-Technologien für Forschung und Lehre an deutschen Universitäten

Die rasante Entwicklung auf dem Gebiet der Omics- und Bioinformatik-Technologien haben die Lebenswissenschaften zu einer Leitdisziplin des 21. Jahrhunderts werden lassen. Nun gilt es, die Potenziale dieser Technologien auszubauen und diese vor allem in Forschung und Lehre zu integrieren. An erster Stelle sind hier die Universitäten gefordert, denn sie sind es, die die Ausbildung der Studierenden in Bachelor-, Master- und Promotionsstudiengängen sicherstellen müssen. Sollen diese Studierenden eine zeitgemäße Ausbildung erhalten, dann haben die deutschen Universitäten eine riesige Infrastrukturmaßnahme zu bewältigen. Sie müssen die Omics-Technologien etablieren, um diesen modernen Zweig der Lebenswissenschaften angemessen in den Lehr- und Forschungsbetrieb einzubauen. Eine solche Infrastrukturmaßnahme lässt sich zurzeit an den meisten deutschen Universitäten aber nur schwer verwirklichen, da deren Finanzierung von den Ländern getragen werden müsste und diese hierfür keinen genügenden Finanzspielraum besitzen. Dabei handelt es sich nicht um eine Anschubfinanzierung, sondern um eine nachhaltige Finanzierung der Omics-Infrastrukturen, die langfristig Investitions-, Verbrauchs- und Personalmittel benötigt. Die nachhaltige Finanzierung der Omics-Infrastrukturen ist also das Kardinalproblem an deutschen Universitäten. Offensichtlich erhalten universitäre Absolventen auf dem Omics-Sektor zurzeit keine zeitgemäße Ausbildung und sind somit nicht in der Lage, in den großen außeruniversitären und industriellen Forschungseinrichtungen an vorderster Front anstehende Forschungsfragen zeitgemäß zu bearbeiten.

Die von der Leopoldina eingerichtete Arbeitsgruppe hat insbesondere den Stand der heutigen universitären Ausbildung auf dem Gebiet der Omics-Technologien abgefragt. Dabei hat sich ergeben, dass Omics-Technologien in der Lehre meist nur theoretisch angesprochen, aber nicht experimentell vermittelt werden. Besonders gravierend ist, dass in einem überfrachteten Medizinstudium Omics-Technologien als Lehrinhalt unterrepräsentiert sind. Ein besonders großes Defizit wurde für die Bioinformatik ausgemacht. Diese Fachrichtung bildet, gemessen an



(Bildrechte: Sisters of Design/Leopoldina)

dem zukünftigen Bedarf, besonders wenige Absolventen aus. Aus den geführten Fachgesprächen ergab sich auch, dass Studierende, aber auch Lehrende, eine zu geringe interdisziplinäre Kompetenz besitzen: Bioinformatiker besitzen zu wenig Kompetenz auf dem Sektor der Omics-Technologien und umgekehrt haben Lebenswissenschaftler zu wenig Einblick in die Leistungsfähigkeit von Bioinformatik-Programmen. Bei der Etablierung von Omics-Technologien an deutschen Universitäten sollte besonders darauf geachtet werden, dass Mitarbeiterstellen mit attraktiven Karrierewegen geschaffen werden. Hier ist auch darüber nachzudenken, wie sich neben zeitlich befristeten Postdoc-Stellen unbefristete Stellen für Expertenpersonal generieren lassen.

Szenarien zur Einrichtung einer Omics-Infrastruktur in Deutschland

Die Arbeitsgruppe der Leopoldina hat sich nach Analyse der Ausgangslage zum Stand der Omics-Technologien in Forschung und Lehre an deutschen Universitäten eindeutig dafür ausgesprochen, dass ein deutlicher Ausbau sowie eine Stärkung der Omics-basierten Forschung und Lehre in den Lebenswissenschaften angestrebt werden sollte. Es wird vorgeschlagen, dass dazu eine Omics- und IT-Infrastruktur aufgebaut wird, die aus einem Netzwerk an bundesweit verteilten Zentren besteht, welche die Omics-Spitzentechnologien vorhalten und sicherstellten, dass diese Technologien auch anderen Universitäten in Forschung und Lehre zur Verfügung gestellt werden. Diese Zentren sollen sich besonders dadurch auszeichnen, dass dort Arbeitsgruppen angesiedelt sind, die die Omics-Technologien nicht nur nutzen, sondern auch aktiv an deren Fortentwicklung mitwirken.

Der erste Zukunftsreport Wissenschaft der Leopoldina entwickelt zusätzlich Modelle, wie ein solches Netzwerk einschließlich einer zentralen Koordinationsstelle verwirklicht werden könnte. Es werden zwei Szenarien genannt, nämlich das DFG-Szenario und

Steckbrief zum Zukunftsreport Wissenschaft der Leopoldina

Im Jahr 2011 beschloss die Leopoldina die Entwicklung des Wissenschaftssystems in Deutschland am Beispiel der Omics-Technologien zu beleuchten. Dazu wurde eine Arbeitsgruppe installiert, die in den Jahren 2012 und 2013 zunächst Fachgespräche mit einschlägigen Experten aus dem In- und Ausland veranstaltete. Parallel dazu wurde 2013 eine bundesweite Umfrage an lebenswissenschaftlichen und medizinischen Fakultäten durchgeführt. Bis Dezember 2013 erarbeitete die Arbeitsgruppe dann den Zukunftsreport, der nach internationaler Begutachtung im Mai 2014 vom Präsidium der Leopoldina verabschiedet wurde.

Der Zukunftsreport widmet sich zunächst dem technologisch bedingten Paradigmenwechsel in den Lebenswissenschaften, beschreibt anschließend Omics-Technologien und deren bisherige Förderung. Er beleuchtet strukturelle Herausforderungen in der Organisation der Lebenswissenschaften, bevor Szenarien für die Etablierung einer nationalen Omics- und IT-Infrastruktur für Forschung und Lehre entwickelt werden. Der Zukunftsreport ist im Internet unter http://www.leopoldina.org/de/politikberatung/wissenschaftliche-kommissionen/zukunftsreport-wissenschaft/ abrufbar.

Mitglieder der Leopoldina Arbeitsgruppe:

Prof. Dr. Rudolf Amann, Max-Planck-Institut für Marine Mikrobiologie, Bremen
Prof. Dr. Roland Eils, Deutsches Krebsforschungszentrum, Heidelberg
Prof. Dr. Michael Hecker, Zentrum für Funktionelle Genomforschung, Greifswald
Prof. Dr. Regine Kahmann, Max-Planck-Institut für Terrestrische Mikrobiologie, Marburg
Prof. Dr. Alfred Pühler, Centrum für Biotechnologie, Universität Bielefeld
Prof. Dr. Dierk Scheel, Leibniz-Institut für Pflanzenbiochemie, Halle (Saale)

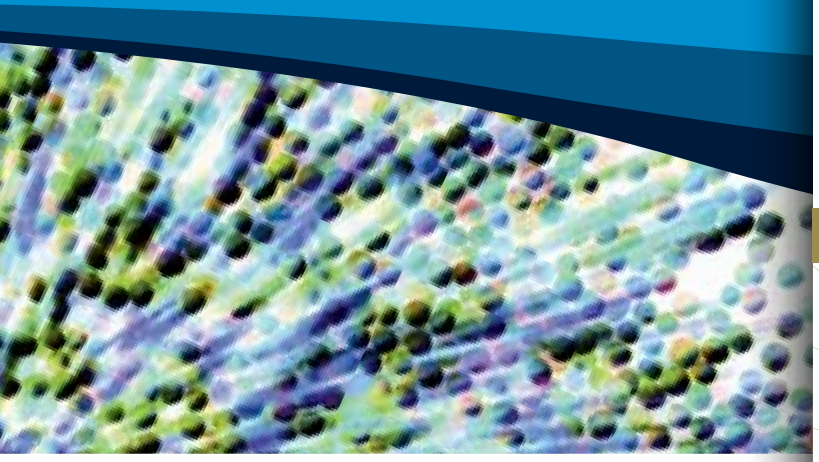


Abbildung 2: Titelseite des Zukunftsreports
(Bildrechte: Sisters of Design/Leopoldina)

das Schweizer Szenario. Das DFG-Szenario sieht vor, dass die DFG den Aufbau einer nationalen Omics-Infrastruktur durch die Gründung einer DFG-Senatskommission unterstützt. Es könnte zusätzlich ein DFG-Panel für Omics-Technologien ins Leben gerufen werden, das die Finanzierung von Omics-Infrastrukturen zum Inhalt hat. Das Schweizer Szenario hingegen orientiert sich an der föderativen Organisationsstruktur des Schweizer Instituts für Bioinformatik (SIB). Hier wurde eine rechtlich und finanziell eigenständige Organisation gegründet, die seit 1998 die bioinformatische Infrastruktur und Datenanalyse in der Schweiz betreibt. Dieses Szenario könnte auch als Vorbild für Deutschland dienen und den Aufbau einer Omics- und IT-Infrastruktur über eine eigenständige Organisation steuern.

Der erste Zukunftsreport Wissenschaft der Leopoldina fand seit seiner Veröffentlichung im September 2014 eine starke Beachtung in der Tages- und Wochenpresse und erzeugte auch Resonanz auf politischer Ebene. So wird seit seiner Publikation sowohl im BMBF als auch in der DFG darüber diskutiert, mit welchen Ansätzen die geforderte Omics- und IT-Infrastruktur in Deutschland etabliert werden kann. Gemessen an der Dringlichkeit dieser Aufgabe kann man mit dem langen Diskussionsprozess natürlich nicht zufrieden sein. Andererseits wurde parallel zu diesem Diskussionsprozess mit der Etablierung des Deutschen Netzwerks für Bioinformatik-Infrastruktur (de.NBI) ein Teil der Leopoldinaforderung bereits verwirklicht. Wie in diesem Sonderheft separat dargestellt, konnte in verhältnismäßig kurzer Zeit – durch das BMBF finanziert – das de.NBI-Netzwerk eingerichtet werden. Mit dieser Etablierung von de.NBI wurde ein gangbarer Weg aufgezeigt, wie insgesamt eine Omics-Infrastruktur bestehend aus einem Netz von Omics-Zentren und einer Koordinationseinheit angelegt sein könnte. Man kann sogar noch einen Schritt weiter gehen und in Erwägung ziehen, das vorhandene de.NBI-Netzwerk als Kern zu nutzen und daran eine Omics-Infrastruktur anzugliedern. Diese Überlegung resultiert aus der Beobachtung, dass die existierenden de.NBI-Servicezentren fast alle an Universitäten angesiedelt sind, die in den vergangenen Jahren in Omics-Technologien investiert haben und daher gut vorbereitet sind, eine Rolle in der zu etablierenden Omics-Infrastruktur zu spielen. Der hier skizzierte Vorschlag ist sicherlich attraktiv und sollte zeitnah von Leopoldina, BMBF und DFG diskutiert werden.

Literatur:

http://www.leopoldina.org/de/politikberatung/wissenschaftliche-kommissionen/zukunftsreport-wissenschaft/

http://www.leopoldina.org/nc/de/publikationen/detailansicht/?publication[publication]=604&cHash=83666e6055d29d2e79bb62b724aa4d7b

Kontakt:

Prof. Dr. A. Pühler
Centrum für Biotechnologie
Universität Bielefeld
Bielefeld
puehler@cebitec.uni-bielefeld.de

Foto: Centrum für Biotechnologie, Bielefeld

eine globale initiative zur erforschung von krebserkrankungen

Das Pan-Cancer Analysis of Whole Genomes (PCAWG) Projekt

von Jan O. Korbel[@,1], Sergei Yakneen[1], Sebastian M. Waszak[1], Matthias Schlesner[2], Roland Eils[2] und Fruzsina Molnár-Gábor[3]

Das *Pan-Cancer Analysis of Whole Genomes* (PCAWG) Projekt, ein innerhalb des Internationalen Krebsgenomkonsortiums (ICGC) organsiertes Forschungsprojekt, ist Wegbereiter des breiten Einsatzes von Analysenmethoden mittels Hochleistungsrechensystemen und Cloud Computing. Neu entwickelte analytische Werkzeuge ermöglichen die standardisierte Untersuchung von Krebsgenomen und assoziierten Datensätzen (z. B. Transkriptom, DNA-Methylierung und klinischen Daten) im Petabyte Bereich und erlauben so die Beantwortung systembiologischer Fragestellungen in ungekannter Dimension. Dieser Artikel stellt das PCAWG Projekt vor, hebt Bedeutung und Potential für Grundlagen- und translationale Forschung heraus, und betont wissenschaftliche, technische und normative Herausforderungen.

„Big Data“: Chancen für die Erforschung des Genoms und von Erkrankungen

Durch Technologieinnovationen im Bereich der massiv-parallelen DNA Sequenzierung ist die Anzahl sequenzierter humaner Genome in den letzten Jahren sprunghaft angestiegen. So wuchs die Menge der in öffentlichen Datenbanken gespeicherten genomischen Daten kürzlich auf mehrere Petabyte an (1 Petabyte entspricht 10^{15} Bytes oder dem Speicherumfang von rund 20.000 modernen Smartphones) – eine Dimension, wie sie im akademischen Bereich vorher vor allem in der Physik bekannt war. Innerhalb des Internationalen Krebsgenomkonsortiums (*International Cancer Genome Consortium*, ICGC; www.icgc.org) haben wissenschaftliche Forschungsverbünde aus 17 Ländern in den vergangenen fünf Jahren eine Datenmenge angehäuft, die dem Speicherplatz von 500.000 DVDs entspricht (Stein *et al.*, 2015).

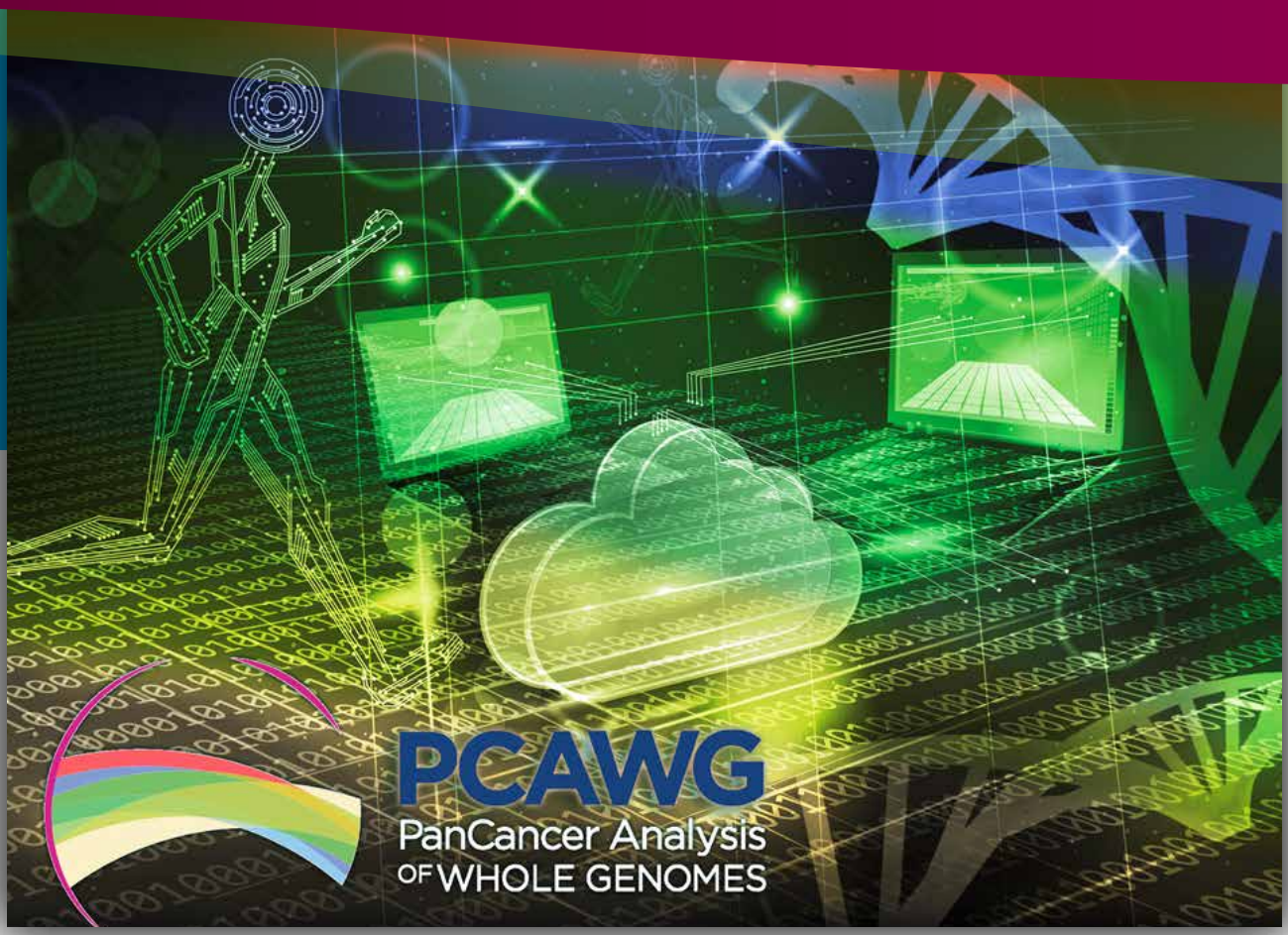
Das Hauptaugenmerk von ICGC Studien liegt hierbei auf der Analyse kompletter Genome von Krebspatienten. Auf diesem Wege konnten schon zahlreiche Fortschritte in der Erforschung molekularer Ursachen bestimmter Krebsformen erzielt werden. Ein Beispiel hierfür ist Prostatakrebs bei Männern zwischen 40 und 50 Jahren, welcher vermehrt im Zusammenhang mit genetischen Umlagerungen auftritt, die durch das männliche Sexualhormon Androgen ausgelöst werden können (Weischenfeldt *et al.*, 2013). Im Medulloblastom, einem Hirntumor, hingegen zeigte sich, dass genregulatorische Elemente (z. B. *enhancer*) an der Krebsentstehung beteiligt sind und durch gezielte Reorganisation des genetischen Materials (d. h. durch genomische Strukturvarianten wie Translokationen) zweckentfremdet werden können (Northcott *et al.*, 2014).

Eine Besonderheit des Krebsgenoms ist die große Heterogenität mit der krebsspezifische Mutationen im Genom auftauchen. In der Tat gleicht kein Krebsgenom dem anderen – klinisch relevante Genmutationen treten vielfach mit einer Häufigkeit von unter 5 % (gelegentlich sogar unter 1 %) innerhalb einer Patientengruppe auf (Lawrence *et al.*, 2014). Um in der Zukunft personalisierte Krebstherapien zu ermöglichen, muss deshalb eine große Anzahl von Tumoren zur Klärung klinisch relevanter Zusammenhänge untersucht werden. Hier setzt das *Pan-Cancer Analysis of Whole Genomes* (PCAWG) Projekt an. Diese Initiative ermöglicht es, integrative Studien, sogenannte „Pan-Krebsgenom Analysen“, also vergleichende Analysen komplett sequenzierter Patientengenome zwischen verschiedenen (oder innerhalb von) Tumortypen durchzuführen. Die große Anzahl von insgesamt 2.800 Krebsgenomen* (Abbildung 1), assoziierten molekularen

[1] European Molecular Biology Laboratory (EMBL), Heidelberg
[2] German Cancer Research Center (DKFZ), Heidelberg
[3] Ruprecht-Karls-Universität Heidelberg, Heidelberg

@Anfragen an: jan.korbel@embl-heidelberg.de

* Die Mehrzahl dieser Genome wird durch das ICGC beigesteuert (68 %). Die restlichen 32 % trägt das US-Amerikanische Cancer Genome Atlas (TCGA) Projekt bei.



Das *Pan-Cancer Analysis of Whole Genomes* (PCAWG) Projekt, eine internationale Initiative zur Erforschung von Krebs-Mutationsmustern, ist Wegbereiter des breiten Einsatzes von Analysenmethoden mittels Hochleistungsrechensystemen und Cloud Computing. Innerhalb von PCAWG vergleichen weltweit über 700 Wissenschaftler mehr als 2.800 Krebsgenomdaten verschiedener Tumortypen miteinander. Fokus dieser Untersuchungen sind die Ursachen und Konsequenzen somatischer und Keimbahnvariationen in kodierenden aber auch nicht-kodierenden Bereichen des Genoms (Copyright©EMBL; Design by P. Riedinger).

Informationen (z. B. Genexpression und DNA-Methylierung) sowie klinischen Daten lässt innerhalb von PCAWG die Beantwortung einer Reihe neuer Fragestellungen zu: So wissen wir immer noch wenig über regulatorische krebsspezifische Mutationen, welche jenseits der 1 - 2 % des Protein-kodierenden Genoms in nichtkodierenden DNA-Regionen auftreten, obwohl es deutliche Hinweise daraufhin gibt, dass Mutationen oder strukturelle Variationen in diesen Regionen für die Krebsentstehung hochrelevant sein können (Horn *et al.*, 2013; Northcott *et al.*, 2014). Krebsgenomuntersuchungen im größeren Maßstab könnten zudem dabei helfen, neue Einblicke in krebsauslösende Viren oder Bakterien zu gewinnen. Weiterhin können integrative Analysen von Erbgut, Krebsmutationen und klinischen Daten dabei helfen, den Einfluss von erblichen Faktoren bei der Krebsentstehung besser zu verstehen.

Herausforderungen für Forschung und IT

Krebsgenomanalysen finden in der biomedizinischen Forschung bereits breite Verwendung – wo genau setzt die PCAWG Initiative also an? Ein wesentliches Problem ist die schiere Menge an genomischen Daten. Mit einer Internetanbindung wie sie typischerweise an deutschen Universitäten zu finden ist benötigen Wissenschaftler viele Monate, um publizierte Sequenzdaten auf ihr lokales Netzwerk zu übertragen bevor Analysen überhaupt beginnen können. In der Realität führt dies dazu, dass wichtige Studien nicht durchgeführt werden können. Ein zusätzliches Problem sind die Hardware-Kosten für die Datenspeicherung und -prozessierung: Aufgrund der durch das ICGC bereits erzeugten Datenmengen würden allein für die Datenspeicherung innerhalb eines professionellen Netzwerks Beträge von über einer Million Euro pro Jahr anfallen. Wissenschaftler können auch nicht einfach auf bereits vorprozessierte Mutationsdaten zurückgreifen (was weniger Speicherplatz notwendig machen würde), denn es fehlt derzeit an vereinheitlichten Prozessabläufen. Die Verwendung einer Vielzahl unterschiedlicher Analyseprotokolle für Genomanalysen führt dazu, dass Ergebnisse verschiedener Forschungsstudien nur sehr begrenzt oder überhaupt nicht miteinander verglichen werden können. Solche Schwierigkeiten beim Zugriff auf große Datensätze und deren Vergleich schränken biomedizinisch relevante Analysen, welche von der rapide ansteigenden Zahl weltweit verfügbarer Genomdatensätze erst ermöglicht würden, erheblich ein.

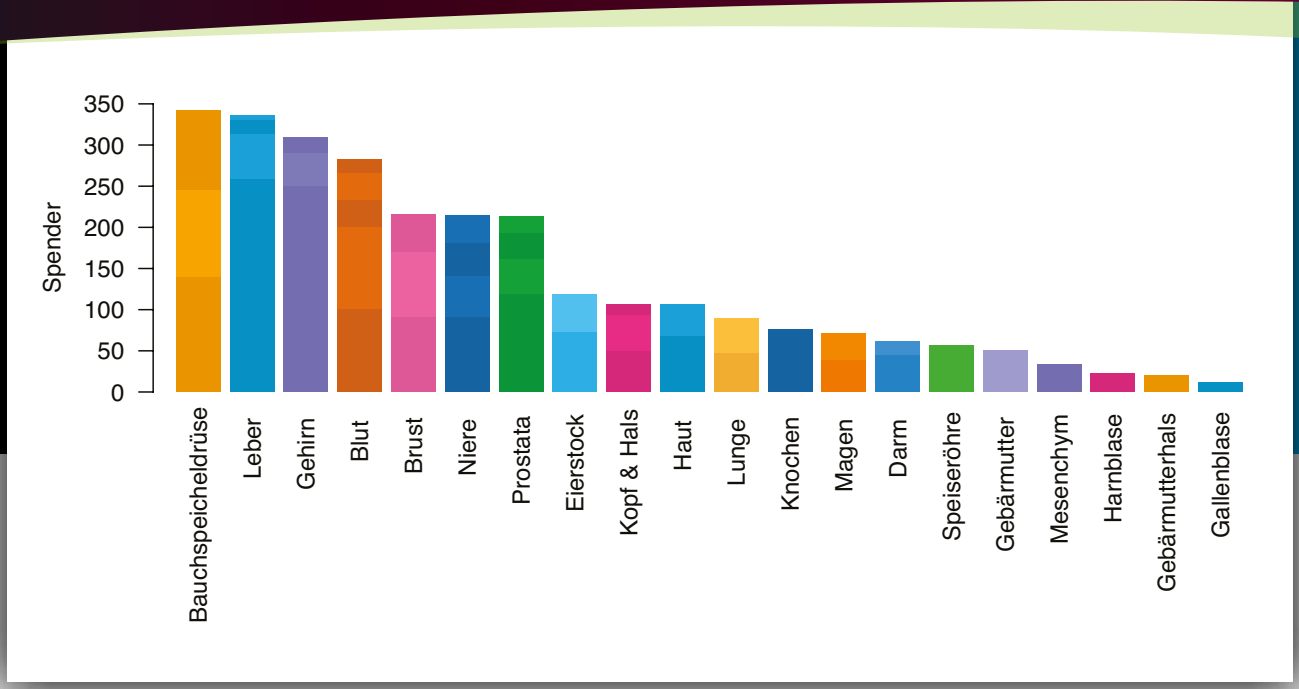


Abbildung 1: Aufteilung von PCAWG Krebspatienten nach Tumortyp sowie ICGC Projekten (helle/dunkle Schattierungen).
(Quelle: https://dcc.icgc.org/pcawg/)

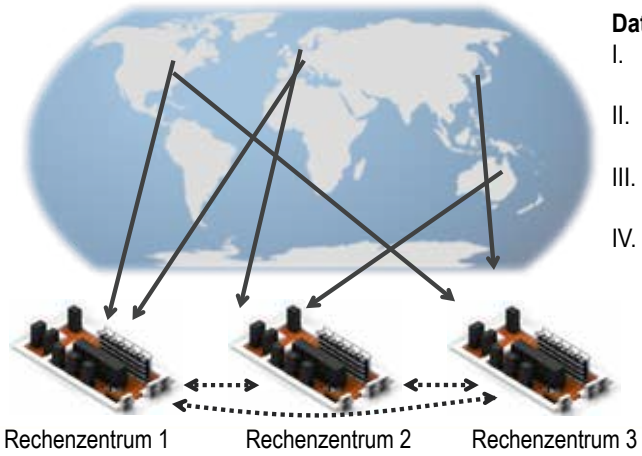
In diesem Zusammenhang gilt das PCAWG Projekt (https://dcc.icgc.org/pcawg) als Modellprojekt zur gemeinsamen Datennutzung von Krebsgenomdaten und internationalen Standardisierung von Analyseabläufen. Um Proben miteinander vergleichen zu können werden innerhalb von PCAWG Sequenzdaten von ca. 1 Petabyte durch Einsatz standardisierter Protokolle zum Alignment von DNA-Sequenzen sowie zur Mutationserkennung vereinheitlicht analysiert. Prozessierte Daten werden allen Projektpartnern unmittelbar zur Verfügung gestellt, sodass alle an dem vereinheitlichten Datensatz rechnen können.

Die einheitliche Verarbeitung einer derartigen Datenmenge stellt eine enorme Herausforderung dar. Insgesamt sieben projekteigene IT-Zentren (in Heidelberg, London, Barcelona, sowie in Nordamerika und Asien) stellen dem Projekt Rechenkapazitäten mit insgesamt 8.824 Computerkernen, 24 TB RAM und 6 PB an Festplattenspeicherplatz zur Verfügung. Zusätzlich wurde ein kommerzieller Cloud-Anbieter, Amazon Web Services (AWS), zur Analyse von ca. 20 % der Proben eingesetzt (Stein *et al.*, 2015). Weitere industrielle Partner (u. a. Annai Systems, Fujitsu, Intel, SAP und Seven Bridges Genomics) unterstützen einzelne Analysen innerhalb des Projektes. Vor allem Cloud-Dienste offerieren ihren Nutzern beträchtliche Speicher- und Rechenkapazitäten, die nach einem *„pay-as-you-go“* Prinzip angeboten werden. Da Cloud-Dienste über das Internet verfügbar sind und viele Nutzer somit auf dieselbe Hardware zugreifen können, müssen hohe Datensicherheitsstandards gelten.

Bei der Verarbeitung personenbezogener Daten in der Cloud dürfen die Persönlichkeitsrechte der betroffenen Patienten nicht verletzt werden. Fundament von Datenschutz und Datenschutzrecht in vielen europäischen Ländern ist das Recht auf informationelle Selbstbestimmung der Betroffenen. Erhöhte Kontrolle kommerzieller Cloud-Anbieter bei der Datenanalyse (und somit verminderte Kontrolle durch den Benutzer) führen zu Fragen bezüglich Datensicherheit und Beachtung von Datenschutzbestimmungen insbesondere bei international verfügbaren Cloud-Diensten. Vor allem deshalb wurden Krebsgenomdaten zahlreicher europäischer ICGC-Projekte bisher nicht innerhalb der Amazon-Cloud, sondern stattdessen in den zahlreich beteiligten akademischen Zentren analysiert. Tatsächlich fehlen derzeit rechtliche Bestimmungen für den Umgang mit solchen neuartigen internationalen Herausforderungen im Datenschutz – sowohl auf internationaler Ebene als auch in vielen europäischen Ländern. Zukünftige Bestimmungen sollten die Definition von personenbezogenen Daten und die Zuständigkeiten für die Verarbeitung und Verteilung derartiger Daten in Clouds überdenken, und sollten zusätzlich die länderübergreifende gemeinsame Datennutzung regeln, um weitere Fortschritte in Krebsforschung, Technologieentwicklung und personalisierter Medizin zu ermöglichen.

PCAWG: Standardisierte Analysen von Krebsgenomdaten im globalen Kontext

Wie aber funktioniert die standardisierte Analyse von Daten innerhalb von PCAWG? Eine zentrale Entwicklung von PCAWG sind Prozessabläufe, die es ermöglichen Genomdaten dezentralisiert (d. h. auf verschiedene Rechenzentren verteilt) zu bearbeiten, wobei gleichzeitig standardisierte Genomanalysen gewährleistet werden (Abbildung 2). Diese Abläufe beinhalten Methoden der Qualitätskontrolle und Genomanalyse, wie Sequenz-Abgleich, Fehler-Korrektur, Identifizierung von Duplikaten und Identifi-



Datenanalyse-Phasen:

I. Datentransfer zu verschiedenen IT-Rechenzentren (bzw.Clouds)
II. Standardisierte Datenprozessierung (für alle Proben die gleichen Prozessabläufe)
III. Teilen prozessierterter Daten (genetische Varianten, RNA-Seq-Daten)
IV. Datamining, Hypothesengenerierung und Hypothesentestung

Abbildung 2: Ansatz der geteilten Verarbeitung innerhalb verschiedener Computerzentren und Clouds innerhalb von PCAWG. Die Datenverarbeitung geschieht innerhalb verschiedener IT-Zentren und Clouds – wie der EBI Embassy Cloud, welche als akademische Gemeinschaft-Cloud am EMBL-EBI genutzt wird, sowie dem DKFZ Hochleistungs-Rechenzentrum. (Grafik: Jan Korbel und PCAWG Projekt).

zierung genetischer Varianten in Keimbahn und Tumor. Derzeit werden dabei drei standardisierte Prozessabläufe eingesetzt (unter amerikanischer, britischer und deutscher Leitung) um sicherzustellen, dass die Genomanalyse effizient und mit ausreichender Aussagekraft durchgeführt wird. Dass diese Abläufe größtenteils Probe für Probe unabhängig durchgeführt werden können, erleichtert die in Abbildung 2 beschriebene parallelisierte Analyse wesentlich. Nach Ablauf dieser Analyseschritte werden prozessierte Daten zwischen akademischen IT-Zentren synchronisiert und so zu einer standardisierten globalen Datenressource für die Forschergemeinschaft zusammengefügt.

Eine besondere Neuerung innerhalb von PCAWG ist die Nutzung virtueller Maschinen bzw. „virtueller Container“ wie der Open Source Software „Docker“, die es dem Projekt ermöglicht standardisierte IT-Abläufe unabhängig von der jeweiligen IT-Infrastruktur durchzuführen. Um weitere Entwicklungen im Feld zu fördern veröffentlicht PCAWG neben einer Datenressource von 2.800 Krebsgenomen auch die im Projekt entwickelten standardisierten Prozessabläufe. Ziel ist dabei nicht nur, Krebsgenomanalysen wie in PCAWG reproduzierbarer zu machen, sondern auch in Zukunft Voraussetzungen für die weitere Vergrößerung der Zahl gemeinsam analysierbarer Patientengenome zu schaffen.

Ausblick

PCAWG ist ein Modellprojekt zur verteilten Nutzung von IT-Ressourcen und Prozessabläufen in der Krebsgenomforschung. Wenngleich die Zahl der derzeit analysierten Krebsgenome schon immens erscheint: Neuartige Systeme zur DNA-Sequenzierung (insbesondere die Illumina HiSeq X Ten Plattform) werden voraussichtlich schon innerhalb der nächsten 5 - 10 Jahre die Sequenzierung von über einer Million Patientengenomen weltweit ermöglichen, die mit klinischen Verlaufsdaten (wie z. B. Tumormarker, Medikamentendosierungen usw.) in Zusammenhang gebracht werden könnten. Um eine breite Anwendung dieser Daten in Forschung und personalisierter Medizin Wirklichkeit werden zu lassen, werden standardisierte Prozessabläufe mit entsprechender Skalierbarkeit benötigt. Dezentral nutzbar gemachte Hochleistungs-IT-Lösungen, insbesondere das *„Cloud-Computing“*, werden dabei eine wichtige Rolle spielen. Dies unterstreicht die Bedeutung normativer Ansätze sowie der Entwicklung spezifischer Regelungen zum *„Cloud-Computing“*. Ein kürzlich für einen solchen Kontext entwickeltes Zukunftsmodell einer *„Genome Cloud“* für Forschung und Translation ist online unter www.genome-cloud.de öffentlich verfügbar.

Danksagungen:

Wir danken den an PCAWG Beteiligten Wissenschaftlern, insbesondere der *„Technical Working Group“*, die international die Analyse einer vorher unerreichten Menge an DNA Sequenzierdaten in verschiedenen IT-Zentren koordiniert. Weiterhin danken wir Nina Habermann für wertvolle Hilfe bei der Übersetzung und Fertigstellung dieses Manuskriptes. J.O.K und F.M.G. danken der Heidelberger Akademie der Wissenschaften für die Förderung ihres normativen Forschungsprojektes.

Steckbrief Forschungsprojekt PCAWG

Im Rahmen des *Pan-Cancer Analysis of Whole Genomes* (PCAWG) Projektes, einer internationalen Initiative zur Erforschung von Krebs-Mutationsmustern, werden mehr als 2.800 Krebsgenomdaten des Internationalen Krebsgenomkonsortiums (ICGC) untersucht. Fokus von PCAWG sind die Ursachen und Konsequenzen somatischer und Keimbahnvariationen in kodierenden aber auch nicht-kodierenden Bereichen des Genoms, im Gegensatz zu vorangegangenen Untersuchungen innerhalb des ICGC sowie des Cancer Genome Atlas Research Projekts, die vor allem auf kodierende Regionen ausgerichtet waren. Innerhalb von PCAWG vergleichen über 700 Wissenschaftler verschiedene Tumortypen miteinander und untersuchen dabei das Genom der Tumore und die dazugehörenden Keimbahngenome. Alle Daten durchlaufen die gleichen standardisierten IT-Prozessabläufe in lokalen Rechenzentren sowie in „Clouds" – so wird eine Standardisierung der Genomanalysen sichergestellt, aufgrund dessen Analysen mit ausreichender Aussagekraft durchgeführt werden können.

Projektname:
Pan-Cancer Analysis of Whole Genomes (PCAWG) Projekt
Beteiligte Partner (Deutschland):
EMBL (Jan Korbel, Wolfgang Huber);
DKFZ (Roland Eils, Peter Lichter, Benedikt Brors, Christoph Plass);
UKE Hamburg (Guido Sauter, Thorsten Schlomm);
Universität Kiel (Reiner Siebert)
Koordinator (Mitglied im Leitungsgremium und verantwortlich für Deutschland):
Jan Korbel (EMBL)

Referenzen:

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., *et al.* (2013). TERT promoter mutations in familial and sporadic melanoma. Science 339, 959-961.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495-501.

Northcott, P.A., Lee, C., Zichner, T., Stutz, A.M., Erkek, S., Kawauchi, D., Shih, D.J.H., Hovestadt, V., Zapatka, M., Sturm, D., *et al.* (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature 511, 428-+.

Stein, L.D., Knoppers, B.M., Campbell, P., Getz, G., and Korbel, J.O. (2015). Data analysis: Create a cloud commons. Nature 523, 149-151.

Weischenfeldt, J., Simon, R., Feuerbach, L., Schlangen, K., Weichenhan, D., Minner, S., Wuttig, D., Warnatz, H.J., Stehr, H., Rausch, T., et al. (2013). Integrative Genomic Analyses Reveal an Androgen-Driven Somatic Alteration Landscape in Early-Onset Prostate Cancer. Cancer Cell 23, 159-170.

Kontakt:

Foto: EMBL

Dr. Jan Korbel
European Molecular Biology Laboratory (EMBL)
Genome Biology Unit
Heidelberg
korbel@embl.de

https://dcc.icgc.org/pcawg
und
www.genome-cloud.de

ICSB 2016
International Conference on Systems Biology

BARCELONA

16 - 20 September 2016

17th INTERNATIONAL CONFERENCE ON SYSTEMS BIOLOGY

DISCOVER THE LATEST TRENDS AND NEW EXITING RESULT IN SYSTEMS BIOLOGY:

Basic Systems Biology

- Modelling Networks and Circuits
- Signalling Pathways
- Cellular decision making
- Large-Scale Networks
- SB of multicellular tissues and organs
- Image-driven Systems Biology
- Variability and Noise: from single-cells up to populations
- Evolutionary Systems Biology
- Cellular populations and ecological interactions

Applications

- Systems Biology of Cancer and Multifactorial Diseases
- Systems Immunology
- Systems Biology of Stem Cells
- Systems and Personalised Medicine
- Synthetic Biology and Biotechnology

Highlighted Topics

- Chemotaxis
- Systems Neuroscience

KEY DATES:

- **1 June 2016 :** Abstracts submission deadline
- **4 July 2016 :** Abstracts submitters notification
- **18 July 2016 :** Early bird registration deadline

Organised by:

ISSB The International Society for Systems Biology

CRG Centre for Genomic Regulation

upf. Universitat Pompeu Fabra Barcelona

@ICSB2016 www.icsb2016barcelona.org #ICSB2016

sensible genomdaten

EURAT beantwortet ethische und rechtliche Fragen in der Genomforschung

von Sebastian Schuol und Eva C. Winkler

Die Genomsequenzierung bringt neben vielen Chancen auch neue ethische und rechtliche Fragen mit sich. Zu den wichtigsten Themen zählen: der Umgang mit Zusatzbefunden, die Gestaltung der Aufklärung und Einwilligung von Patienten bzw. Probanden, der Schutz personenbezogener Daten und die Verantwortung der Forscher. Um diesen kontextnah und lokal zu begegnen, arbeiten in Heidelberg Wissenschaftler aus Natur- und Normwissenschaften im EURAT-Projekt zusammen. Mit einer Stellungnahme reagieren sie auf die aufgeworfenen Fragen und plädieren angesichts von gesetzlichen Regulierungslücken für die Selbstregulierung der Wissenschaft.

Die Genomsequenzierung wirft Fragen auf

Angetrieben von den enormen Fortschritten in der Sequenzierungstechnologie, der bioinformatischen Auswertung von Daten und den Fortschritten in der digitalen Datenspeicherung entwickelt sich die Genomforschung seit der Jahrtausendwende zu einem beachtlichen Forschungszweig mit vielfachem Erkenntnisgewinn. Die zunehmende Wirtschaftlichkeit und Zeiteffizienz der neuen Sequenziertechnologien des *next generation sequencing* haben entscheidend zu einer Erweiterung des Blicks von isolierten Genen hin zu ganzen Genomen beigetragen. Dabei spielt die Genomsequenzierung mittlerweile nicht nur in der Grundlagenforschung, sondern auch in der translationalen Forschung, eine bedeutende Rolle. Im besonderen Maße gilt dies für die Krebsforschung (Winkler *et al.*, 2013), in der beispielsweise genetische Veränderungen im Tumorgewebe im Vergleich zur Keimbahn ermittelt werden um für eine präzisere Therapie Mutationen zu identifizieren, welche das Tumorwachstum bzw. die Therapiewirksamkeit beeinflussen. Gemessen an einem solchen Nutzen darf bereits jetzt spekuliert werden, dass die Verwendung genomweiter Analysen zukünftig Eingang in die klinische Routinediagnostik finden wird.

Das EURAT-Projekt entwickelt Lösungen

Um diesen ethischen und rechtlichen Fragen möglichst kontextnah und lokal zu begegnen hat sich 2011 ein Konsortium von

Der enorme Wissensfortschritt der Genomforschung eröffnet Handlungsbereiche, in deren Lichte sich neuartige ethische und rechtliche Fragen stellen. Dabei werden im Wesentlichen die folgenden Themenschwerpunkte kontrovers diskutiert (Winkler & Schickhardt 2014):

1. Die wiederholte Datenanalyse (Folgeforschung) erhöht die Wahrscheinlichkeit, auf nicht erwartete aber medizinisch relevante Befunde, sog. Zusatzbefunde, zu stoßen. Wie soll mit ihnen umgegangen werden?
2. Die erhobenen genetischen Daten führen möglicherweise zu Einsichten des Patienten, die seine Zukunft verändern. Wie soll dies im Aufklärungs- und Einwilligungsprozess des Patienten berücksichtigt werden?
3. Die genomischen Daten sind identifizierend und können sensible Informationen über eine Person enthalten. Wie lassen sich der Schutz der Privatsphäre und die Forschung mit diesen Daten vereinbaren?
4. Der Prozess der Genomsequenzierung stellt ein stark arbeitsteiliges Forschungsgebiet dar, sodass gerade hier Verantwortungsdiffusion droht. Welche Verantwortung tragen die jeweiligen Forscher darin?

Mikrofluidiksystem eines Sequenziergeräts (Quelle: Marsilius-Kolleg).

Medizinern, Naturwissenschaftlern, Bioinformatikern, Juristen und Ethikern aus der Ruprecht-Karls-Universität Heidelberg, dem Max-Planck-Institut (MPI) für ausländisches öffentliches Recht und Völkerrecht, dem Deutschen Krebsforschungszentrum (DKFZ) und dem Europäischen Laboratorium für Molekularbiologie (EMBL) mit einer anerkannten Expertise in den mit der Genomsequenzierung assoziierten Forschungsbereichen zusammengefunden, um normative Fragen zur Genomsequenzierung zu untersuchen und gemeinsam praxisnahe Vorschläge zu erarbeiten. EURAT (Ethische und Rechtliche Aspekte der Totalsequenzierung des menschlichen Genoms) identifiziert dabei Probleme aus der Praxis, diskutiert diese unter Einbeziehung der Normwissenschaften – Ethik und Recht – und gibt Lösungsvorschläge zurück. Dadurch wird abseits der langsameren Regulierung durch den Gesetzgeber den vor Ort aufgeworfenen ethischen und rechtlichen Problemen praxisnah und selbstregulativ begegnet. In den beteiligten Einrichtungen werden dadurch Kommunikationsprozesse über die normativen Herausforderungen initiiert und gemeinsam Dokumente erarbeitet. Auf letztere wird nun genauer eingegangen:

Im Rahmen der erstmals 2013 veröffentlichten EURAT-Stellungnahme „Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung“ wurden ein Kodex für Forscher sowie zwei Mustertexte für Patienteninformationen und Einwilligungserklärungen entwickelt. Der Forscherkodex begegnet der Herausforderung, dass mit der Genomsequenzierung ein in ethischer und rechtlicher Hinsicht unterregulierter Handlungsbereich betreten wird, da z. B. nichtärztliche Forscher nicht in gleicher Weise über ein verankertes Berufsethos wie Mediziner verfügen, das zur Orientierung beitragen könnte und das Gendiagnostikgesetz nur für die klinische Versorgung gilt, die Forschung dagegen aber explizit ausklammert (GenDG § 2, Abs. 2); diese gesetzliche Regulierung hängt den rasanten Entwicklungen insgesamt hinterher (Bartnek 2009).

Der Kodex begründet neue Rechte und Pflichten der Forscher im Umgang mit genetischem Wissen über Patienten (und deren Familien). Aufgrund der Arbeit mit Patientengenomen und seines humangenetischen Wissens kommt dem Einzelforscher eine berufsspezifische Handlungsverantwortung und den Leitungsgremien eine Organisationsverantwortung zu. In einzelnen Richtlinien wird u. a. auf eine Risikominimierung im Hinblick auf das Forschungsvorhaben, die Notwendigkeit einer vorliegenden Einwilligungserklärung des Patienten und eines Ethik-Votums sowie die Einwilligung des Patienten zur Aufbewahrung der Proben eingegangen. Von zentraler Bedeutung ist der Komplex Daten und Forschungsbefunde: Um bestmöglichen Patientenschutz zu gewährleisten, dürfen die ermittelten Daten nur verschlüsselt und nur für vom Patienten genehmigte Forschungsbereiche verwendet werden. Anders als bei der Anonymisierung, bei der die Verbindung zwischen Daten und Patient gekappt wird, erlaubt die Pseudonymisierung im Falle von im Forschungsverlauf auftretenden Befunden mit therapeutischem Nutzen, diese dem Patienten zugutekommen zu lassen. Die Forscher sind verpflichtet, Erkenntnisse, welche voraussichtlich zum Wohl des Patienten beitragen, dem behandelnden Arzt zu melden, der dann über das weitere Vorgehen entscheidet. Sie haben aber keine Pflicht, aktiv danach zu suchen. Der Forscherkodex schafft somit einen institutionalisierten Rahmen, der zwar Forschern einige Pflich-

Chemische Lösungen für den Betrieb der Sequenzierinstrumente (Foto: J. Ritzerfeld).

ten auferlegt, ihnen aber innerhalb dieser Grenzen zugleich Handlungssicherheit gibt und sie damit schützt.

Die Wirkungsweite der Genomsequenzierung betrifft allerdings nicht nur die Forscherseite, sondern insbesondere jene Patienten, deren Genome sequenziert werden. Die EURAT-Gruppe betont die Notwendigkeit einer informierten Patienteneinwilligung (*informed consent*). Um eine Balance zwischen möglichst umfassender Information über wichtige Implikationen und möglichst hoher Verständlichkeit für Patienten zu erreichen, hat die EURAT-Gruppe Mustertexte für spezifische Anwendungsbereiche entwickelt.

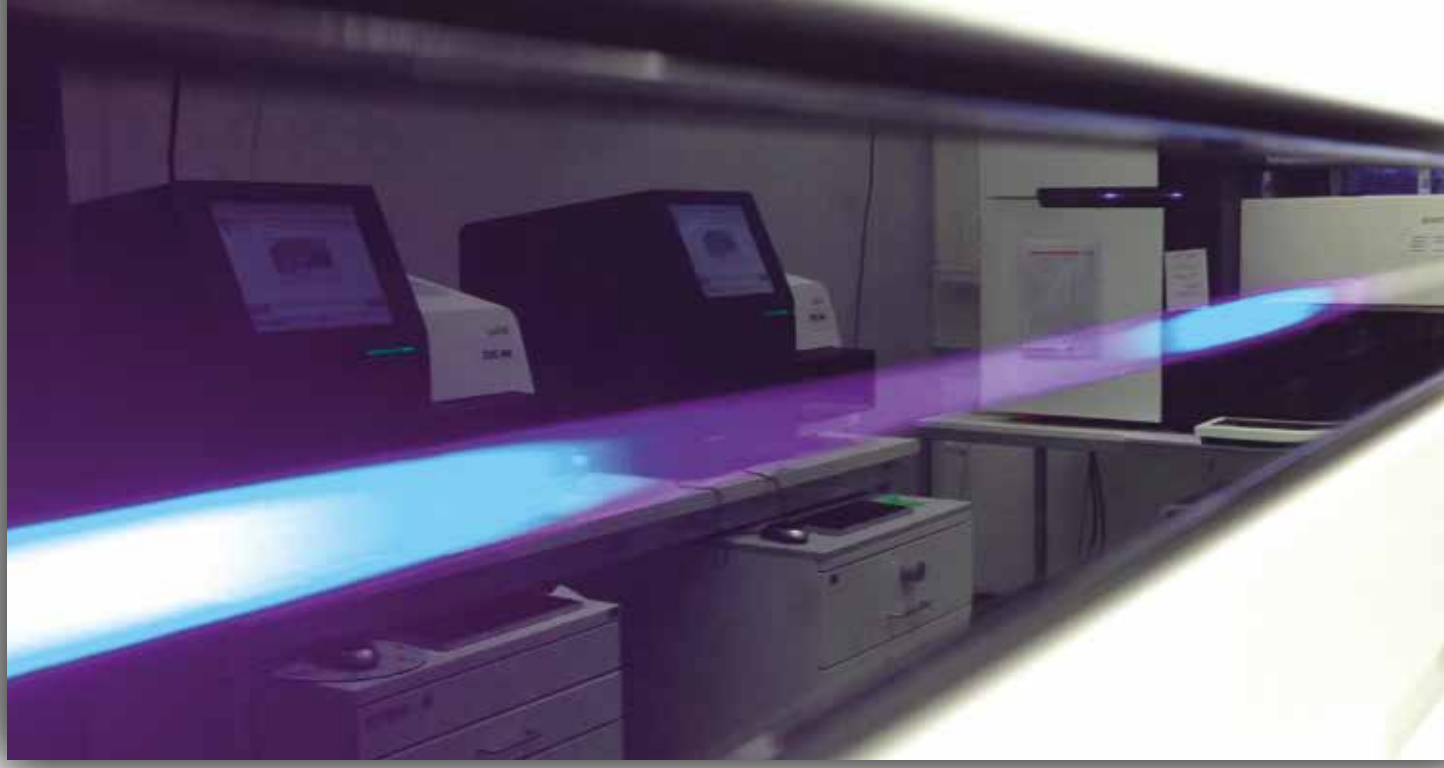
Im Hinblick auf Zusatzbefunde ist ein Spagat zu leisten: Einerseits soll der Patient entscheiden, ob und wenn ja, welche Zusatzbefunde er erfahren will. Andererseits ist eine vollumfängliche Aufklärung über Zusatzbefunde angesichts von über 5.000 monogen verursachten Krankheiten nicht möglich (OMIM).

Daher sollten Zusatzbefunde nach Kategorien (behandelbar, vorbeugbar, nicht behandelbar) klassifiziert und an Beispielen erklärt werden. Zudem steht das Thema Daten im Zentrum der Aufklärung: Da genomischen Daten das Potential innewohnt, Personen zu identifizieren und sensible Informationen über sie zu erhalten, werden unter anderem die Methode der Pseudonymisierung, die Weitergabe an inter- und nationale Forschungsinstitute, die Möglichkeiten und Grenzen der Datenlöschung beim Widerruf der Einwilligung und die Datenzugänglichkeit zu Überprüfungszwecken erklärt.

Wie die EURAT-Lösungen wirksam werden

Zu den Erfolgen der EURAT-Stellungnahme auf einer institutionellen Ebene zählt, dass der Kodex für Forscher bald nach seiner Veröffentlichung sowohl von der Universität Heidelberg mittels offiziellem Senatsbeschluss als auch dem DKFZ angenommen wurde und seitdem hier für alle Forscher auf dem Gebiet der Genomsequenzierung als verbindlich gilt. Verstöße gegen den Verhaltenskodex können nun auch sanktionierende Konsequenzen haben, etwa arbeits- oder haftungsrechtlicher Art. Das Ziel des Kodex ist es jedoch nicht zu sanktionieren, sondern in einem ansonsten unregulierten Bereich Orientierung zu stiften. Dass dies gelingt, zeigt seine positive Resonanz, sowohl in der Fachwelt als auch in den öffentlichen Medien. Doch entwickelt sich der Bereich der Genomforschung im rasanten Tempo weiter. Daher wurde im November 2015 die EURAT-Stellungnahme in einer zweiten aktualisierten Fassung publiziert (EURAT 2015). Diese enthält nun unter anderem konkrete Lösungen zum Datenschutz, wie z. B. eine Erklärung des inzwischen am DKFZ in Zusammenarbeit mit EURAT entwickelten Rahmendatenschutzkonzeptes zum Umgang mit personenbezogenen Daten für Forschungsprojekte mit genomischen Daten.

Der Themenkomplex Daten kristallisiert sich zu einem der wichtigsten ethischen Themenfelder im Rahmen der Genomforschung heraus und wird die Zukunft von EURAT weiterhin prägen. Wegen der zunehmenden Wichtigkeit der Weitergabe genomischer Daten in internationalen Forschungskonsortien und aus Kostengründen gewinnt das Thema *Cloud-Computing* in der Genomforschung gegenwärtig an Bedeutung. Da z. B. beim internationalen *Datasharing* verschiedene Rechtsordnungen betroffen sind, aber die Gerichtsbarkeit vom Sitz des Cloud-Anbieters abhängen und der Schutz personenbezogener Daten variieren kann, stellen sich neue ethische und rechtliche Fragen zur Weitergabe von Daten in Cloud-Lösungen, für die bald praktikable Lösungen entwickelt werden müssen.



Ein Sequenzierinstrument in Betrieb (Quelle: Marsilius-Kolleg).

Steckbrief Forschungsprojekt:

EURAT (Ethische und Rechtliche Aspekte der Totalsequenzierung des menschlichen Genoms) ist ein Projekt zu normativen Fragen der Genomsequenzierung am Forschungsstandort Heidelberg, das Wissenschaftler der Universität Heidelberg, des Universitätsklinikums, des Deutschen Krebsforschungszentrums (DKFZ), des Europäischen Laboratoriums für Molekularbiologie (EMBL) sowie des Max-Plank-Instituts (MPI) für ausländisches öffentliches Recht und Völkerrecht vereint. Das EURAT-Projekt wurde 2011 im Rahmen der Exzellenzinitiative durch das Marsilius-Kolleg initiiert.

Mitglieder von EURAT sind:

Prof. Dr. Claus R. Bartram, Humangenetik | Prof. Dr. Roland Eils, Bioinformatik | Prof. Dr. Hanno Glimm, Onkologie | Prof. Dr. Christof von Kalle, Onkologie | Prof. Dr. Dr. h.c. Paul Kirchhof, Verfassungsrecht | Dr. Jan Korbel, Bioinformatik/Genomsequenzierung | Prof. Dr. Andreas E. Kulozik, Onkologie | Prof. Dr. Peter Lichter, Tumorgenetik/Genomsequenzierung | Prof. Dr. Peter Schirmacher, Pathologie/Biobanking | Prof. Dr. Klaus Tanner, Ethik/Theologie (Projektsprecher von 2011-2013) | Prof. Dr. Stefan Wiemann, Genomsequenzierung | Prof. Dr. Dr. Eva Winkler, Onkologie/Medizinethik (Projektsprecherin) | Prof. Dr. Dr. h.c. Rüdiger Wolfrum, Verfassungsrecht/Völkerrecht.

www.uni-heidelberg.de/totalsequenzierung

Referenzen:

Winkler, E.C., Ose, D., Glimm, H., Tanner, K., von Kalle, C. (2013): Personalisierte Medizin und Informed Consent: Klinische und ethische Erwägungen im Rahmen der Entwicklung einer Best Practice Leitlinie für die biobankbasierte Ganzgenomforschung in der Onkologie. Ethik in der Medizin, 25(3), 195-203.

Winkler, E.C. und Schickhardt, C. (2014): Ethische Herausforderungen der Genomsequenzierung in der translationalen Forschung und Antworten aus dem EURAT-Projekt. J Lab Med, 38(4), 211–220.

EURAT (2015): Stellungnahme: Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung (Heidelberg: Nino Druck GmbH).

OMIM – Online Mendelian Inheritance in Man. www.omim.org

Bartnek, T. (2009): Das Gendiagnostikgesetz: Ein lückenhafter Schutz. Gen-ethischer Informationsdienst 194, 50-54.

Kontakt:

Prof. Dr. med. Dr. phil. Eva Winkler
Sprecherin des EURAT-Projekts und Leiterin des Schwerpunktes „Ethik und Patientenorientierung in der Onkologie“
Nationales Centrum für Tumorerkrankungen Heidelberg
eva.winkler@med.uni-heidelberg.de

Sebastian Schuol, M.A.
Koordinator des EURAT-Projekts
Nationales Centrum für Tumorerkrankungen Heidelberg
Sebastian.Schuol@med.uni-heidelberg.de

Foto Schuol: S. Schuol; Foto Winkler: Medienabteilung Uniklinik Heidelberg

big data: perspektiven in der krebstherapie

Eindrücke aus der Industrieperspektive

von Ajay Kumar

Big-Data-Technologien werden in der Industrie und der medizinischen Forschung immer mehr eingesetzt. Entscheidungsträger in der Industrie können dadurch ihre bestehenden Geschäftsmodelle optimieren und einen höheren Mehrwert für ihre Kunden schaffen. In der Krebsforschung könnten Big-Data-Technologien einen Paradigmenwechsel ermöglichen! Während eines Forschungsaufenthalts im Sommer am Deutschen Krebsforschungszentrum (DKFZ) in Heidelberg konnte ich mir einen Einblick verschaffen, wie Big-Data-Methoden im spezifischen Anwendungsfall („*Use Case*“) der Krebsforschung eingesetzt werden. Durch die offene und kooperative Atmosphäre in dieser Forschungseinrichtung konnte ich meine Erfahrungen teilen und gleichzeitig selbst durch das einzigartige interdisziplinäre Umfeld Ideen sammeln, die ich an meinen Arbeitsplatz in der Industrie übertragen konnte.

Von der Industrie zur Krebsforschung

Beim Zeitunglesen an einem Sonntagmorgen stieß ich auf einen interessanten Artikel über die Verwendung von Big Data in der Krebsforschung. Der Artikel war ein Interview mit zwei Forschern am Deutschen Krebsforschungszentrum (DKFZ) in Heidelberg, Prof. Roland Eils und Prof. Christof von Kalle. In dem Artikel wurde erläutert, dass Big-Data-Technologien immer mehr zu einem unabdingbaren Hilfsmittel in der Krebsforschung werden, um die Erkrankung auf einer tieferen Ebene zu verstehen. Als Mitarbeiter in einem Automobilunternehmen war ich mit dem Einsatz von Big-Data-Technologien in der Konsumgüter- und Maschinenbauindustrie vertraut. Die Liste der Anwendungsfälle ist lang und reicht von der Optimierung von Supermarktregalen bis zur Effizienzsteigerung von Maschinen und Prozessen. Ein Forschungsaufenthalt am DKFZ bot die einzigartige Gelegenheit zu erfahren, wie technische Fortschritte bei Hochdurchsatz-Technologien sich direkt auf die Krebsforschung und Krebstherapie auswirken.

Die vier Phasen der industriellen Revolution

Mein Interesse an Big Data entstand durch einen starken Trend im Produktionsumfeld. Industrie 4.0 oder die vierte industrielle Revolution ist vielleicht einer der in der deutschen Industrie heutzutage am meisten diskutierten Begriffe (Kagermann *et al.*, 2013). Die ersten drei industriellen Revolutionen waren das Ergebnis von Mechanisierung, Elektrifizierung und Informationstechnologie (Abbildung 1). Vor der ersten industriellen Revolution im späten 18. Jahrhundert wurde die Produktionsarbeit manuell oder mithilfe von natürlichen Ressourcen wie Wasser oder Wind ausgeführt. Die Erfindung der Dampfmaschine und ihr Einsatz in der Produktion haben die industrielle Landschaft verändert. Textilfabriken in Großbritannien haben diese Veränderung zuerst erfahren. Zu Beginn des 20. Jahrhunderts haben die Elektrifizierung und Arbeitsteilung zur zweiten industriellen Revolution geführt. Die Produktion des T-Modells von Henry Ford verdeutlicht diese Veränderung am besten. Sie machte das Automobil erschwinglich, solange alle das gleiche kauften. Die Massenproduktion standardisierter Güter zu niedrigen Kosten wurde Wirklichkeit.

In den frühen 1970er-Jahren hat der Einfluss von Elektronik und Informationstechnologie auf die Produktion die dritte industrielle Revolution ausgelöst. Komplexe Lieferketten für Produkte konnten mit softwarebasierten Systemen geplant werden (Kagermann *et al.*, 2013). Der Automatisierungsgrad wurde mit Robotern erhöht, die die manuelle Arbeit an Fließbändern ersetzten, und mithilfe von numerischen Steuerungsmaschinen konnten komplexe Produktionsmuster mit hohen technischen Anforderungen realisiert werden. Damit begann die moderne Automobilherstellung in Europa, bei der viele verschiedene Automodelle mit hohen Qualitätsstandards produziert werden können.

Die vierte industrielle Revolution basiert auf der Idee des Internets der Dinge („*Internet of things – IoT*“) und Dienste (Abbildung 1). In Fabriken sind Maschinen, Materialien, Dienstleistungen, Produkte, Mitarbeiter, Lieferanten und Kunden miteinander vernetzt. In dieser vernetzten Umgebung interagieren Maschinen

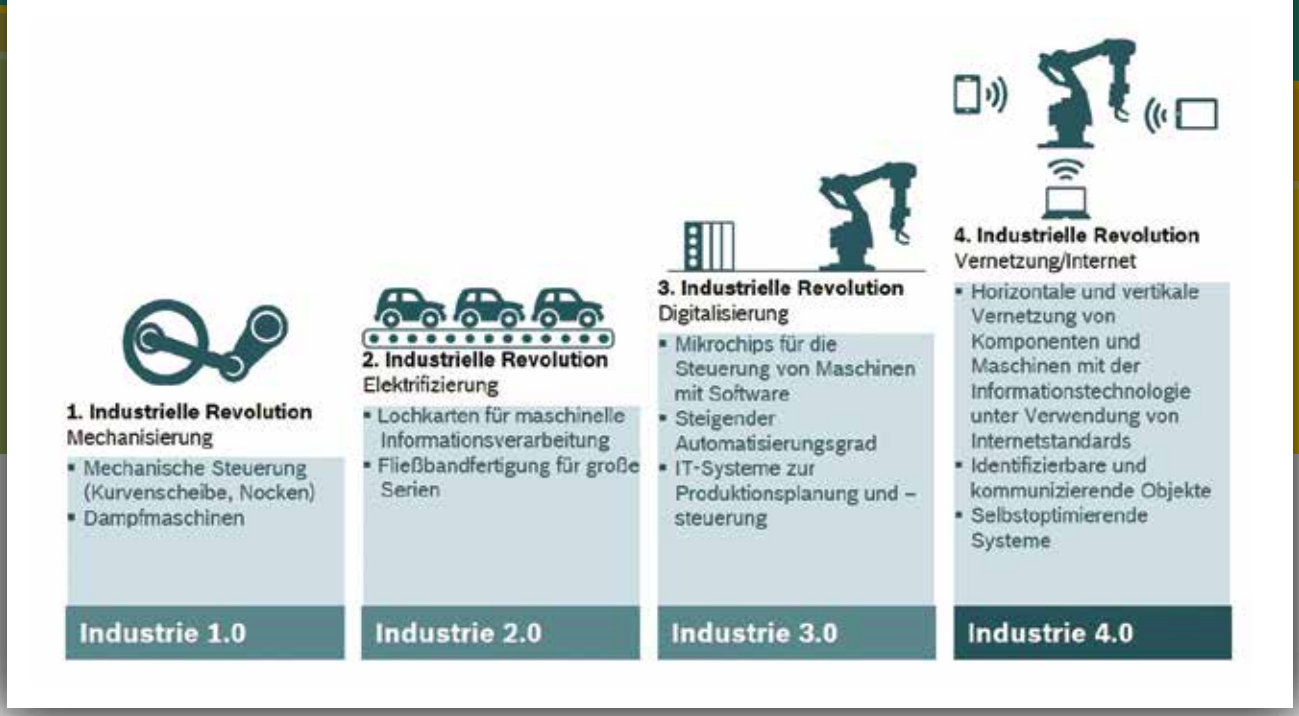


Abbildung 1: Die vier Phasen der industriellen Revolution (Quelle: Bosch Rexroth).

miteinander, Roboter arbeiten mit Mitarbeitern in sogenannten cyber-physischen Systemen zusammen (Kagermann *et al.*, 2013) und Produktionssysteme sind in der Lage zu lernen und sich selbst zu optimieren. Außerhalb der Fabrik sind Kunden und Produkte mit dem Internet verbunden und eröffnen neue Geschäftsmöglichkeiten, während sie gleichzeitig die bestehenden Geschäftsmodelle ins Wanken bringen.

Alle industriellen Revolutionen haben die Komplexität des Produktionsumfelds erhöht und eine Steigerung des industriellen Outputs ermöglicht. Darüber hinaus haben sie sich entscheidend auf unsere Gesellschaft und die Wettbewerbsfähigkeit der Industrie ausgewirkt.

Big-Data-Technologien verbessern die Krebsbehandlung

Die Abteilung *Theoretische Bioinformatik („eilslabs")* am DKFZ befindet sich auf dem Universitätscampus im Neuenheimer Feld in Heidelberg, auf dem viele verschiedene Forschungseinrichtungen und Kliniken angesiedelt sind.

Ich begann meinen Forschungsaufenthalt in der Arbeitsgruppe *Data Management and Genomics IT* in den *eilslabs*. Die Arbeitsgruppe besteht aus Bioinformatikern, Computeringenieuren und Softwareentwicklern. Ich wusste nur sehr wenig über die molekularen Mechanismen, die Krebserkrankungen zugrunde liegen, und erfuhr schnell, dass Krebs aufgrund von Veränderungen in unserer DNA entsteht. Die Identifizierung dieser auslösenden Mutationen wäre ein großer Durchbruch und stellt eines der Hauptziele der modernen Krebsforschung dar. Eine Art Mutationswörterbuch zum Thema Krebs.

Das Wissen über diese genetischen Veränderungen in Verbindung mit klinischen Patientendaten ermöglicht es Ärzten, individuelle Behandlungspläne für Krebspatienten auszuarbeiten. Dieser gemeinschaftliche, interdisziplinäre Ansatz führt zu Synergien im Therapieprozess, senkt dadurch die Behandlungskosten und – was noch wichtiger ist – verbessert das Wohlbefinden und die Überlebenschancen der Patienten.

Enabler für Big Data: Sequenzierungskosten und das Internet

Ich habe mich gefragt, warum dieses Verfahren noch nicht zum Standard in der Krebstherapie gehört. Vor zehn Jahren stellten hohe Sequenzierungskosten und Engpässe in der IT-Infrastruktur die größten Hürden für die Übertragung der Genomforschung auf die klinische Praxis dar. 2001 betrugen die Kosten für die Sequenzierung eines einzelnen menschlichen Genoms etwa 100 Millionen USD (Abbildung 2). Fortschritte in der Sequenzierungstechnologie haben die Kosten über die Jahre auf 5.000 USD pro Patientengenom im Jahr 2014 gesenkt (Hayden, 2014). Bis 2020 sollte es möglich sein, ein menschliches Genom für weniger als 800 USD zu sequenzieren.

Heutzutage ermöglichen niedrige Sequenzierungskosten und die Verfügbarkeit des Hochleistungsrechnens einen Durchbruch für die Krebsforschung und die personalisierte Medizin. Ärzte können so mithilfe der Genomdaten eines einzelnen Patienten eine genau abgestimmte, individuelle Therapie entwickeln. Die personalisierte Onkologie sollte für einen Großteil der Krebspatienten schon bald Realität sein!

In der Industrie fallen riesige Datenmengen durch die Verfügbarkeit des Internets zu Hause, unterwegs und an fast allen Orten an, das es möglich macht Produkte, Dienstleistungen und Menschen miteinander zu vernetzen. Analysten des Beratungsunternehmens *Machina Research* gehen davon aus, dass sich die Zahl der an das Internet angeschlossenen Einzelgeräte bis 2022 auf 14 Milliarden erhöht, von IP-fähigen Autos über Heizungsanlagen bis hin zu Sicherheitskameras, Sensoren und Produktionsmaschinen (Abbildung 3).

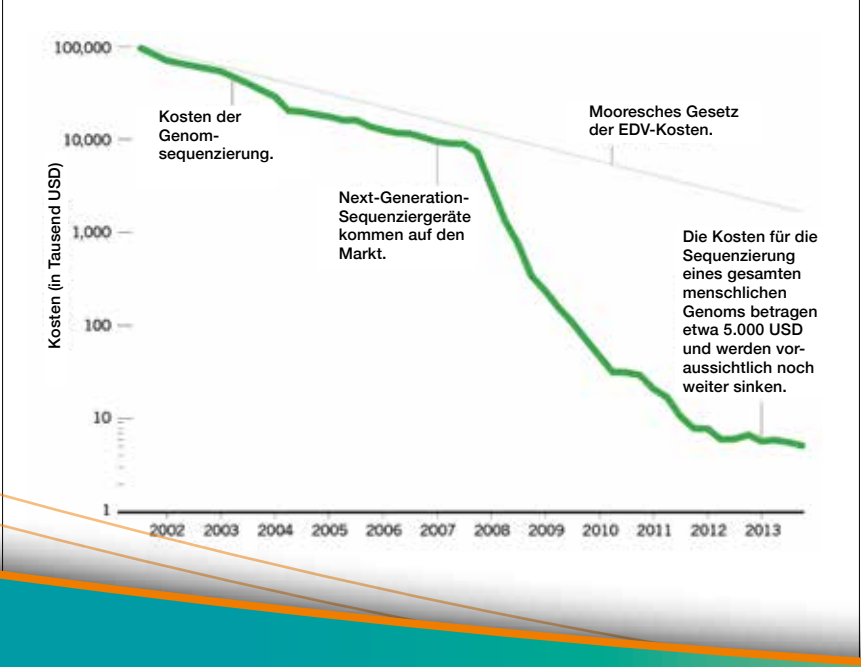
Die Sammlung und Analyse der Daten, die von den Geräten über das Internet gesendet werden, kann die Nutzererfahrung oder die Effizienz in Fabriken erheblich verbessern. Beispielsweise können alle Maschinen eines Fließbands in einem Werk an das Internet oder ein gemeinsames Netzwerk angeschlossen werden. Sensoren und Vorrichtungen in der Maschine überwachen den Maschinenzustand und den Produktionsprozess. Durch die Analyse dieser Daten sind Ingenieure in der Lage, Abweichungen im Produktionsprozess zu korrigieren oder die Wartung der Maschine zu planen, um unvorhergesehene Ausfälle zu vermeiden.

Während meines Forschungsaufenthaltes lernte ich, dass bei der Realisierung des Anwendungsfalls am DKFZ viele spezifische Probleme bewältigt werden mussten. Die drei Hauptprobleme sind hier das Datenmanagement, ethische Fragen und die Automatisierung von Arbeitsschritten bei der Datenanalyse.

Datenmanagement: Speicherung und Sicherheit

Die Handhabung und Verarbeitung von Genomdaten ist eine kritische Aufgabe. Um die Dimension des Anwendungsfalls zu verdeutlichen, kann man davon ausgehen, dass jeden Tag die Genome von 50 bis 80 Patienten sequenziert werden. Die Sequenzierungsdaten eines einzelnen menschlichen Genoms haben eine Größe von etwa 80 GB. Mit 2 Proben, d. h. Tumorprobe und Kontrolle, würden jeden Tag 8 bis 12 Terabytes an Daten erzeugt werden. Dies entspricht nahezu dem täglichen

Abbildung 2: Entwicklung der Sequenzierungskosten



In den ersten Jahren nach dem Abschluss des Humangenomprojekts folgte die Entwicklung der Genomsequenzierungskosten grob dem Mooreschen Gesetz, nach dem ein exponentieller Rückgang der EDV-Kosten zu erwarten ist. Nach 2007 sanken die Sequenzierungskosten rasant. (Quelle: überarbeitet nach Hayden, 2014)

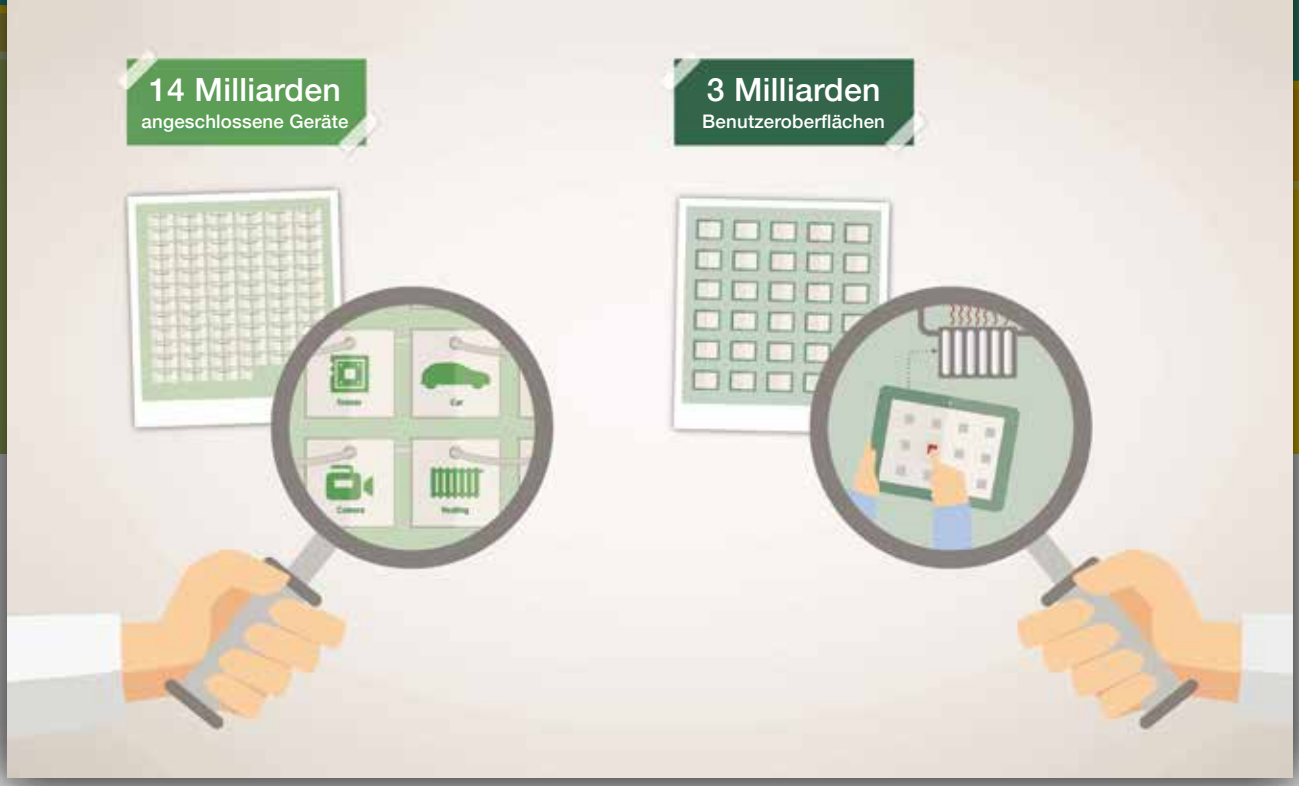


Abbildung 3: Bis 2022 an das Internet angeschlossene Geräte; 14 Milliarden angeschlossene Geräte und 3 Milliarden Benutzeroberflächen (Quelle: überarbeitet nach Machina Research/Bosch Software Innovations GmbH).

Datenvolumen auf Twitter! Für die aktuelle Forschung mit etwa 500 Patienten im Jahr ist bereits ein ausgezeichneter Petabyte-Speicher vorhanden. Um den Anwendungsfall in den nächsten Jahren fortzuführen, wurden verschiedene Optionen beurteilt, wie Cloud, Industriepartnerschaften oder ausgesourcte Einrichtungen. Die Cloud-Speicherung ermöglicht Skalierbarkeit, erfordert jedoch Maßnahmen im Hinblick auf Datensicherheit und Konnektivität. Eine ausschließlich lokale Speicherung würde hohe Investitionen und Betriebskosten mit sich bringen, allerdings auch einen schnellen Zugriff und hohe Sicherheit bieten. Partnerschaften für die Installation von Datenfarmen zur Speicherung von medizinischen und IoT-Daten mit ähnlichen Datenschutzstandards könnten die Kosten von Speicherungslösungen weiter senken. Ethische Fragestellungen in Bezug auf Genomdaten müssen von spezialisierten interdisziplinären Projektgruppen wie EURAT diskutiert werden (siehe Artikel auf Seite 40).

Der Workflow für die Datenanalyse: One Touch Pipeline (OTP)

Der Workflow für die Datenanalyse ist das Kernelement des Anwendungsfalls. Die One-Touch-Pipeline (OTP) ist eine Automatisierungsplattform für das Management von Genomdaten aus Versuchen des Next Generation Sequencing, die den Datentransfer unterstützt und darüber hinaus die Qualitätskontrolle, das Sequenzalignment im Vergleich mit einem Referenzgenom und die Identifizierung von einzelnen Nukleotidvarianten in der Patienten-DNA durchführt. OTP ist wie ein virtuelles Fließband. Ein ausgezeichnetes Fließband verfügt über stabile Prozesse, ist nach Bedarf skalierbar, hat eine geringe Ausfallzeit und arbeitet wirtschaftlich. Das Material durchläuft zügig die Fertigungsschritte zur Herstellung des Endprodukts. Ist das Gegenteil der Fall, führt dies zur Anhäufung von Lagerbeständen, hohen Aussonderungsraten, Qualitätsverlusten und hohen Kosten.

OTP muss wie ein ausgezeichnetes Fließband laufen, um den schnellen Transfer der Genomdaten in die klinische Praxis zu gewährleisten. Das DKFZ muss dafür disziplin- und abteilungsübergreifend arbeiten und mit Softwareunternehmen zusammenarbeiten, um OTP auf zukünftige Anforderungen vorzubereiten.

Das Datenmanagement-Team ist für die Entwicklung und Pflege von OTP zuständig. Das Team erhält viele Benutzeranfragen, muss enge Fristen einhalten und verfügt nur über begrenzte Ressourcen. Um diese Herausforderungen zu bewältigen, nutzt die Datenmanagement-Gruppe bei der Entwicklung von OTP Scrum, eine iterative, agile Produktentwicklungsmethode. Das Arbeitsumfeld ist flexibel, kooperativ und teamorientiert und verfügt über kurze Feedbackzyklen. Das Team analysiert die Bedürfnisse der Benutzer, ordnet diese nach ihrer Priorität und plant einen Zeitraum von etwa 3 Wochen (Sprints), um die Probleme zu lösen (Abbildung 4). Ich war inspiriert, einige Elemente dieser Methode an meinen Arbeitsplatz mitzunehmen. Industrielle Arbeitsumfelder haben eine feste Organisationsstruktur und eine definierte Anzahl an Managementebenen. Ziele und Strategien werden von oben nach unten kommuniziert. Dies sorgt für Ein-

Abbildung 4: Tägliches Stand-Up-Meeting des OTP Scrum-Teams. Dabei geben die Teammitglieder ein kurzes Feedback zur aktuellen Arbeit (Quelle: G. Zipprich).

heitlichkeit, macht die Organisationsstruktur jedoch auch starr und langsam. Industrie 4.0 stellt nicht nur eine technologische Veränderung dar, sondern auch eine Veränderung für unsere Arbeitsplätze. Flexibilität und Agilität sind wichtige Faktoren. Einige Prinzipien von Scrum können diese Bedürfnisse erfüllen und zu einer höheren Zufriedenheit von Mitarbeitern am Arbeitsplatz führen.

Der Blick in die Zukunft: wirksame Krebstherapie und vernetzte Welt

Big-Data-Technologien werden die wichtigsten Enabler für die personalisierte Krebstherapie und Industrie 4.0 sein. In Anwendungsfällen der Industrie werden Big-Data-Technologien zu einer Verbesserung des Arbeitsumfelds und der Wettbewerbsfähigkeit, sowie zu einer höheren Kundenzufriedenheit beitragen. In der Krebsforschung und Krebstherapie werden sie einen starken Impuls für die personalisierte Krebstherapie geben. Zur Umsetzung dieser Anwendungsfälle sind starke Partnerschaften und eine enge Zusammenarbeit zwischen den verschiedenen Projektpartnern erforderlich. Die Kooperationspartner müssen in den frühen Phasen der Umsetzung Risiken auf sich nehmen und sind auf eine starke Unterstützung durch Entscheidungsträger angewiesen.

Referenzen:

Kagermann, H., Wahlster, W., Helbig, J. – Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0 der Forschungsunion Wirtschaft – Wissenschaft und acatech – Deutsche Akademie der Technikwissenschaften e.V. (2013).

Hayden, E.C. (2014) Technology: The 1000$ Genome. Nature 507, 294-295

Kontakt:

Foto: A. Kumar

Ajay Kumar
Direktor im Produktions- und Operationsmanagement
in der Automobilindustrie
Heidelberg
ajaykumar@web.de

fachgruppe bioinformatik – interessen mit einer stimme vertreten

FaBI – der Zusammenschluss von Fachgruppen aus fünf deutschen Fachgesellschaften aus den Lebenswissenschaften und der Informatik

von Matthias Rarey

Bioinformatik hat sich zu einem zentralen Baustein lebenswissenschaftlicher Forschung entwickelt. Aus der Komplexität der Fragestellungen gepaart mit großen Datenmengen resultieren große Herausforderungen für die Informatik. Bioinformatik-Forschung war und ist heute mehr denn je vom Wechselspiel zwischen Anwendung und Methodenentwicklung geprägt. Biologen und Informatiker, Chemiker, Physiker, Mathematiker und Statistiker, Pharmazeuten und Mediziner – in der Bioinformatik findet man aus gutem Grund Wissenschaftler jeglicher Couleur. So interessant und zielführend diese Heterogenität für die Wissenschaft ist, sie birgt auch ihre Schwierigkeiten. In der Wissenschaftsorganisation spielen Fachgesellschaften eine zentrale Rolle, und die Bioinformatik ist naturgemäß in vielen Gesellschaften verankert. Eine Repräsentation des gesamten Fachgebietes ist somit schwierig, und Ansprechpartner sind für Außenstehende meist nur schwer zu identifizieren. 2014 beschlossen daher fünf große Fachgesellschaften mit Bezug zu den Lebenswissenschaften und der Informatik die Gründung der gemeinsamen Fachgruppe Bioinformatik, um der Bioinformatik in Deutschland eine Interessensvertretung, ein Gesicht und eine Stimme zu geben.

Die Bioinformatik kann in Deutschland auf eine lange Tradition zurückblicken. Bereits 1985 fand eine erste Tagung zum Thema „Biotechnologie und Information“ statt, die nun seit 1993 jährlich als **German Conference on Bioinformatics (GCB)** an wechselnden Orten in Deutschland ausgerichtet wird. Heute erlebt die Bioinformatik eine enorme Dynamik, ausgehend sowohl von den Informations- als auch den Lebenswissenschaften. Unabhängig davon, ob man sich primär der Informatik oder den Lebenswissenschaften zugehörig fühlt, ist ein gemeinsames Handeln aller Akteure in diesem interdisziplinären Feld von großer Bedeutung. Dazu dienen Fachgesellschaften. Sie spielen in nahezu allen wissenschaftlichen Disziplinen eine tragende Rolle in der Gestaltung von Fachgebieten. Beispielsweise ging der Förderung der Bioinformatik in Form von Forschungszentren durch das BMBF und die DFG ein wichtiges Positionspapier unter maßgeblicher Beteiligung der Gesellschaft für Chemische Technik und Biotechnologie (DECHEMA) und der Gesellschaft für Informatik (GI) voraus. Aus diesem Grund wurde 2014 die gemeinsame Fachgruppe Bioinformatik (FaBI) von den Fachgesellschaften aus Informatik und Lebenswissenschaften ins Leben gerufen. Die FaBI vertritt heute über 700 Bioinformatiker. Sie tritt für eine international kompetitive Bioinformatik-Forschung und -Lehre in Deutschland ein und ist Ansprechpartnerin für Politik und Gesellschaft.

FaBI
Fachgruppe Bioinformatik

FaBI-Beirat (v.l.n.r.): Bertram Weiß (Bayer Pharma AG), Sven Rahmann (U. Duisburg-Essen, TU Dortmund), Oliver Kohlbacher (U. Tübingen), Matthias Rarey (U. Hamburg), Caroline Friedel (LMU München), Ina Koch (U. Frankfurt). Weitere Beiratsmitglieder (nicht abgebildet): Thomas Engel (LMU München), Heike Pospisil (FH Wildau), Martin Vingron (MPI für Genetik, Berlin) (Quelle: Dechema e. V.).

Die FaBI hat bereits einiges erreicht. Prioritäten lagen auf einer konsensorientierten Meinungsbildung zum Stand der Bioinformatik in Deutschland einerseits und der Steigerung von Sichtbarkeit und Transparenz andererseits. Auf der Basis einer breit angelegten Umfrage wurde im ersten Halbjahr 2015 ein Positionspapier erarbeitet. Seine Kernaussagen betonen die Eigenständigkeit der Disziplin, die Notwendigkeit der Schaffung nachhaltiger Infrastrukturen, die Internationalisierung und Aspekte der Ausbildung und Nachwuchsförderung. Unter http://www.bioinformatik.de ist ein Informationsportal entstanden, welches nicht nur die vielfältigen Aktivitäten der FaBI darstellt. Erstmals ist auch ein Verzeichnis von akademischen Gruppen mit Bioinformatik-Bezug in Deutschland verfügbar. Gerade für ein interdisziplinäres Fach vereinfacht sich damit die Suche nach Experten und Kooperationspartnern deutlich. Zur Orientierung von Schülerinnen und Schülern und Studierenden werden auf der Website alle Studiengänge mit Bioinformatikinhalten tabellarisch dargestellt. Neuigkeiten, Konferenzankündigungen und Stellenangebote runden das Informationsangebot für Bioinformatiker ab.

Wichtige Aktivitäten der Fachgruppe Bioinformatik in 2014 und 2015

German Conference on Bioinformatics (GCB)

Vergabe der Ausrichtung der jährlichen, internationalen Konferenz GCB an eine deutsche Hochschule; Unterstützung bei der Programmgestaltung und Organisation

Internetpräsenz http://www.bioinformatik.de

Betrieb und redaktionelle Gestaltung der Internetpräsenz www.bioinformatik.de. Die Seite enthält neben Informationen zur Fachgruppe, zur GCB, zur Bioinformatik im Allgemeinen auch aktuelle Nachrichten, Stellenangebote und ein Verzeichnis deutscher Bioinformatik-Arbeitsgruppen und Studienangebote.

Positionspapier „Bioinformatik in Deutschland – Perspektive 2015“

Darstellung des Standes der Bioinformatik in Deutschland aus Sicht der Akteure incl. sechs Empfehlungen für die Weiterentwicklung. Das Positionspapier wurde aus einer Umfrage unter Bioinformatikern entwickelt und wird von einer großen Gruppe von Wissenschaftlern getragen (siehe http://bioinformatik.de/Positionspapier)

Fachgesellschaften mit Vertretung in der FaBI, in alphabetischer Reihenfolge

- **DECHEMA:**
  http://www.dechema.de
  Gesellschaft für Chemische Technik und Biotechnologie e. V.
- **GBM:**
  https://www.gbm-online.de
  Gesellschaft für Biochemie und Molekularbiologie e. V.
- **GDCh:**
  https://www.gdch.de
  Gesellschaft Deutscher Chemiker e. V.
- **GI:**
  https://www.gi.de
  Gesellschaft für Informatik e. V.
- **GMDS:**
  http://www.gmds.de
  Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V.

Kontakt:

Prof. Dr. Matthias Rarey
Universität Hamburg
ZBH - Zentrum für Bioinformatik
rarey@zbh.uni-hamburg.de
http://www.zbh.uni-hamburg.de/rarey

Sprecher des Beirats der Fachgruppe Bioinformatik (FaBI)
der DECHEMA e. V., GBM e. V., GDCh e. V., GI e. V. und GMDS e. V.
sprecher@bioinformatik.de
http://www.bioinformatik.de

Foto: Berit Schulz / R.Rarey

Definition der Bioinformatik als Grundlage der FaBI

„Die Bioinformatik ist eine interdisziplinäre Wissenschaft. Unter Bioinformatik verstehen wir die Erforschung, Entwicklung und Anwendung computergestützter Methoden zur Beantwortung molekularbiologischer und biomedizinischer Fragestellungen. Im Fokus stehen Modelle und Algorithmen für Daten auf molekularer und zellbiologischer Ebene, beispielsweise für

- ***Genome und Gene,***
- ***Gen- und Proteinexpression und -regulation,***
- ***metabolische und regulatorische Pfade und Netzwerke,***
- ***Strukturen von Biomakromolekülen, insb. DNA, RNA und Proteine,***
- ***molekulare Interaktionen zwischen Biomakromolekülen untereinander und zwischen Biomakromolekülen und weiteren Substanzen wie beispielsweise Substraten, Transmittern, Botenstoffen und Inhibitoren sowie***
- ***die molekulare Charakterisierung von Ökosystemen.“***

Quelle: Fachgruppe Bioinformatik, http://bioinformatik.de/de/bioinformatik-3/was-ist-bioinformatik

Bundesministerium
für Bildung
und Forschung

Neuigkeiten aus dem BMBF

BMBF erleichtert Flüchtlingen den Zugang zum Studium

Bildung ist für eine gelungene Integration zentral. Mehr als die Hälfte der Flüchtlinge, die derzeit nach Deutschland kommen, ist unter 25 Jahre alt – also in einem Alter, in dem sie eine Ausbildung benötigen. Um denjenigen unter ihnen, die studieren möchten und hierfür ausreichende Qualifikationen mitbringen, den Zugang zum Studium zu ermöglichen, wird das Bundesbildungsministerium die Hochschulen in den kommenden Jahren mit gezielten Maßnahmen unterstützen und so bei der Integration der Flüchtlinge helfen.

Das Bundesbildungsministerium hat hierfür ein umfangreiches Maßnahmenpaket in Höhe von 100 Millionen Euro geschnürt. Es besteht aus drei Bausteinen, die eine Basis für den erfolgreichen Zugang zum Studium bilden: Kompetenzen und Potenziale der Flüchtlinge sollen ermittelt, ihre Studierfähigkeit durch fachliche und sprachliche Vorbereitung sichergestellt sowie die Integration an den Hochschulen unterstützt werden.

„Bildung ist der Schlüssel für die Integration der Flüchtlinge, insbesondere für jene, die dauerhaft eine Perspektive in Deutschland haben. Die Hochschulen sind dabei von großer Bedeutung. Auch deshalb, weil sie jahrelange Erfahrungen mit ausländischen Studierenden gesammelt haben. Wir haben bereits passende Instrumente, um mit sprachlichen Herausforderungen oder unterschiedlichen Qualifikationen umzugehen und ausländische Studierende gut zu beraten. Auf diese Infrastruktur können wir jetzt bei der Integration von Flüchtlingen an den Hochschulen aufbauen“, sagte Bundesbildungsministerin Johanna Wanka bei der Vorstellung des Maßnahmenpakets. „Wer studieren will und kann, dem wollen wir ein Studium ermöglichen – auch zum Nutzen Deutschlands.“

www.bmbf.de/de/fluechtlinge-durch-bildung-integrieren-1944.html

Mehr Geld für Bildung und Forschung

Mit dem Haushalt 2016 steigt der Etat des BMBF erneut, diesmal um 1,1 Milliarden Euro auf rund 16,4 Milliarden Euro. Das zeigt, wie wichtig Bildung und Forschung für die Politik der Bundesregierung sind.

Für die Förderung der beruflichen Bildung stehen insgesamt rund 515 Millionen Euro bereit. Damit wird dieser Schwerpunkt mit einem Plus von 15 Prozent im Vergleich zum Jahr 2015 deutlich ausgebaut.

Für die Finanzierung zusätzlicher Studienplätze sind im Rahmen des Hochschulpaktes von 2007 bis 2015 rund acht Milliarden Euro vom Bund an die Länder geflossen. Im Jahr 2016 werden weitere gut zwei Milliarden Euro hinzukommen. Darüber hinaus werden 50 Millionen Euro in die Verbesserung der Lehrerbildung investiert.

Insgesamt rund 5,5 Milliarden Euro stellt das Ministerium für die institutionelle Forschungsförderung bereit. Ebenfalls im Haushalt 2016 enthalten ist eine Aufstockung der Mittel für die Forschungsförderung im Rahmen der neuen Hightech-Strategie der Bundesregierung.

www.bmbf.de/de/der-haushalt-des-bundesministeriums-fuer-bildung-und-forschung-202.html

Sprachkompetenz wird hier besonders gefördert: Johanna Wanka mit Schülern eines Deutschkurses.

Quelle: BMBF/Hans-Joachim Rickel

Kampagne macht Mut zum Lesen und Schreiben lernen

Rund 7,5 Millionen Menschen in Deutschland sind sogenannte funktionale Analphabeten. Sie können zwar einzelne Sätze lesen und schreiben, nicht jedoch zusammenhängende Texte verstehen. Weitere 2,3 Millionen Menschen gelten als vollständige Analphabeten. Sie können selbst einzelne Sätze weder schreiben noch verstehen. Aus Angst und Scham trauen sich nur wenige, aktiv Hilfe zu suchen.

Darum hat das Bundesbildungsministerium mit weiteren Partnern eine bundesweite Aufklärungskampagne gestartet. Die Kampagne will möglichst viele Menschen sensibilisieren und über Möglichkeiten informieren, im Erwachsenenalter Lesen und Schreiben zu lernen. Dazu werden Erzählungen von Erwachsenen über den Schlüsselmoment in ihrem Leben in den Mittelpunkt gestellt. Es geht um den Zeitpunkt, in dem sie Mut fassten und sich entschieden, besser Lesen und Schreiben zu lernen. TV-Spots, Kino- und Außenwerbung sowie Postkartenaktionen setzen den Augenblick aufmerksamkeitsstark in Szene.

Bund und Länder wollen in den kommenden zehn Jahren die Lese- und Schreibfähigkeiten von Erwachsenen in Deutschland deutlich verbessern. Sie sollen künftig mehr Angebote bekommen, die sie beim Lesen und Schreiben lernen unterstützen. Das Bundesbildungsministerium wird in den nächsten zehn Jahren mit bis zu 180 Millionen Euro Alphabetisierungsprojekte fördern sowie Kurskonzepte und Selbstlernmöglichkeiten schaffen.

„Menschen, die nicht ausreichend lesen und schreiben können, fühlen sich häufig ausgeschlossen, denn in der modernen technik- und dienstleistungsorientierten Arbeitswelt sind diese Fähigkeiten das Fundament für gesellschaftliche Teilhabe und sichere Beschäftigung“, sagt Bundesbildungsministerin Johanna Wanka. „Gemeinsam mit den Ländern und vielen weiteren Partnern wollen wir in den nächsten zehn Jahren erreichen, dass mehr Menschen den Mut finden, auch in späteren Lebensphasen ihre Fähigkeiten im Lesen und Schreiben zu verbessern.“

www.mein-schluessel-zur-welt.de

Nur Mut! Es ist nie zu spät für den nächsten Schritt. Das neue Motiv der Aufklärungskampagne soll betroffene Menschen motivieren, im Erwachsenenalter Lesen und Schreiben zu lernen.

Quelle: BMBF

Neues Wissenschaftsjahr widmet sich der Meeresforschung

Die Meeresforschung ist das zentrale Thema des neuen Wissenschaftsjahres, das Mitte 2016 beginnt. Die Themen reichen vom Lebensraum Meer über die Bedeutung der Ozeane für Wetter und Klima bis hin zur gesellschaftlichen Bedeutung der Meere und Küstenregionen als Kulturräume, Sehnsuchtsorte und Reiseziele. Bundesweit werden mehrere hundert Veranstaltungen, Gesprächsrunden, Ausstellungen und Wettbewerbe stattfinden.

„Meere und Ozeane sind für unser Leben von immenser Bedeutung. Den Großteil der Meere kennen wir noch gar nicht. Trotzdem werden sie vielfach rücksichtslos ausgebeutet und verschmutzt“, sagt Bundesforschungsministerin Johanna Wanka. „Die Meere sind für uns Nahrungsquelle, Wirtschaftsraum und Klimamaschine. Dieses große Thema wollen wir den Menschen nahe bringen. Das bedeutet für die Bürgerinnen und Bürger den größten Lebensraum des Planeten besser kennen und zu schützen lernen.“

Wie wird sich das Ökosystem der Meere und Ozeane verändern? Eine von vielen Fragen im Wissenschaftsjahr 2016*17.

Quelle: flysafe340 - Fotolia

Das derzeit laufende Wissenschaftsjahr 2015 steht im Zeichen der „Zukunftsstadt". Wissenschaftlerinnen und Wissenschaftler suchen gemeinsam mit Kommunen und Wirtschaft sowie Bürgerinnen und Bürgern konkret und vor Ort Lösungen für große gesellschaftliche Herausforderungen zu Themen wie sichere Energie, klimaangepasstes Bauen oder Mobilität.

www.bmbf.de/de/meere-und-ozeane-entdecken-nutzen-schuetzen-2338.html und www.wissenschaftsjahr-zukunftsstadt.de

Fusionsanlage „Wendelstein 7-X" geht an den Start

Die neue Fusionsanlage Wendelstein 7-X hat vor kurzem in Greifswald den Betrieb aufgenommen. Das nationale Forschungsgroßgerät wurde vom Max-Planck-Institut für Plasmaphysik errichtet und soll die Kraftwerkstauglichkeit von Fusionsanlagen des sogenannten „Stellarator"-Typs demonstrieren.

„Wendelstein 7-X steht für Spitzenforschung in Deutschland. Mit Wendelstein beschreiten wir einen neuen Weg in der Kernfusion, der perspektivisch zu grundlastfähigen Kraftwerken führen kann, die stabil große Mengen an Strom bereitstellen", sagt Bundesforschungsministerin Johanna Wanka.

Die Investitionskosten belaufen sich auf 370 Millionen Euro und die Gesamtkosten des für den Bau gegründeten Instituts in Greifswald auf rund 1,1 Milliarden Euro. Die Finanzierung erfolgte gemeinsam durch Bund, EU und Mecklenburg-Vorpommern.

www.bmbf.de/de/jetzt-wird-s-heiss-fusionsanlage-wendelstein-7-x-nimmt-betrieb-auf-2160.html

Spektakuläres Skelett von Tyrannosarus rex in Berlin zu sehen

Der König aller Dinosaurier ist der neue Star einer Sonderausstellung im Naturkundemuseum Berlin. Das Skelett des Tyannosaurus rex „Tristan Otto" ist einer der am besten erhaltenen Funde weltweit. Drei Jahre wird es in Berlin für die Öffentlichkeit zu bestaunen sein.

„Von 300 Knochen, die so ein Tier hat, sind 170 Knochen bei uns im Original ausgestellt", sagt Jochen Vogel, Direktor des Museums. Wissenschaftler wollen das Skelett in den kommenden Jahren noch weiter erforschen. Einiges konnten sie bereits herausfinden, zum Beispiel, dass Tristan Otto einen Knochentumor im Kiefer hatte und ziemliche Zahnschmerzen gehabt haben muss. Wie der Dino gelebt hat, wie alt er wurde und was er gegessen hat, das wollen die Wissenschaftler noch genauer herausfinden. Tristan Otto wurde im Jahr 2012 in Montana, USA, gefunden. Er ist etwa 66 Millionen Jahre alt.

„Das Museum für Naturkunde – als Leibniz-Einrichtung gemeinsam von Bund und Ländern getragen – steht für einmalige Sammlungen, exzellente Forschung und Ausstellungen, die es vermögen, Menschen zu begeistern. Mit der Ausstellung zu Tristan, diesem spektakulären Fund eines Tyrannosaurus rex, stellt dies das Museum ein weiteres Mal unter Beweis", sagte Bundesforschungsministerin Johanna Wanka anlässlich der Ausstellungsöffnung.

www.bmbf.de/de/dino-mit-zahnschmerzen-skelett-des-t-rex-fasziniert-forscher-2230.html

Es ist eine Attraktion: das zwölf Meter lange Skelett des Tyrannosaurus rex „Tristan Otto". Drei Jahre wird es im Naturkundemuseum Berlin zu sehen sein.

Quelle: BMBF/ Hans-Joachim Rickel

BMBF stärkt Innovationskraft des deutschen Mittelstands

Auf den kleinen und mittleren Unternehmen (KMU) fußt das deutsche Wirtschaftsmodell: Sie bieten 16 Millionen Menschen Arbeit und stellen etwa 83 Prozent der betrieblichen Ausbildungsplätze. Zwar geht es vielen Unternehmen gut. Angesichts des globalen Wettbewerbs, der Digitalisierung und neuer Geschäftsmodelle ist es aber keineswegs sicher, dass das so bleibt.

Das Bundesforschungsministerium möchte die KMU dabei unterstützen, neue Ideen zu entwickeln und Forschungsergebnisse für sich zu nutzen. „Unser Ziel ist es, auch Unternehmen zu eigenen Innovationsanstrengungen zu ermuntern, die hier bisher wenig aktiv waren“, sagte Bundesministerin Johanna Wanka bei der Vorstellung des neuen Programms „Vorfahrt für den Mittelstand“, mit dem das Ministerium die Förderung der KMU bis einschließlich 2017 auf rund 320 Millionen Euro pro Jahr erhöht.

Denn es gibt Grund zur Sorge: Die KMU haben ihre Ausgaben für Innovationen in den vergangenen Jahren nicht erhöht – ganz anders als die großen Unternehmen. Dieser Trend muss umgekehrt werden. Dazu soll das neue Programm beitragen. Der Zugang für KMU zu den Fachprogrammen des Bundesforschungsministeriums wird weiter verbessert, gerade wenn es um digitale Wirtschaft, gesundes Leben und nachhaltiges Wirtschaften geht. Zudem möchte das Ministerium die KMU mit starken Partnern zusammenbringen, mit Hochschulen, Forschungseinrichtungen und Großunternehmen. Auch möchte es dabei helfen, dass ihnen die nötigen Fachkräfte zur Verfügung stehen und Nachwuchs exzellent ausgebildet wird.

www.bmbf.de/de/fuer-einen-innovativen-mittelstand-2326.html

G7-Wissenschaftsminister gehen globale Probleme an

Die Bekämpfung armutsbedingter Krankheiten war eines der Schwerpunktthemen der deutschen G7-Präsidentschaft im Jahr 2015. Die G7-Wissenschaftsminister haben auf einem Treffen in Berlin beschlossen, die Forschung der G7 auf den gesamten Kreis der vernachlässigten armutsbedingten Infektionskrankheiten auszuweiten. Diese umfassen mehr als 25 Infektionskrankheiten, die vor allem in ärmeren und tropischen Ländern auftreten. Vereinbart wurde, alle bereits laufenden Maßnahmen zu armutsbedingten Infektionskrankheiten zu erheben und ihre Forschungsförderung zu koordinieren. In einer Konferenz 2016 soll eine gemeinsame Forschungsinitiative beschlossen werden.

„Wissenschaft und Forschung haben in allen G7-Staaten einen hohen Stellenwert. Wir müssen unser Wissen und Können einbringen, um dort zu helfen, wo Menschheitsaufgaben zu lösen sind“, sagte Bundesforschungsministerin Johanna Wanka anlässlich des Treffens.

Wanka kündigte zudem an, die Förderung der Produktentwicklungspartnerschaften, durch die gezielt die Entwicklung bestimmter Medikamente gefördert wird, von deutscher Seite in den nächsten fünf Jahren mit weiteren 50 Millionen Euro zu unterstützen.

Treffen der G7-Wissenschaftsminister in Berlin.

Quelle: BMBF/Hans-Joachim Rickel

www.bmbf.de/de/die-deutsche-g7-praesidentschaft-273.html

Kontakt:

Informationen zu diesen und anderen interessanten Themen zur neuen Hightech-Strategie für Deutschland finden Sie unter: **www.hightech-strategie.de**

SIMULATION LABORATORIES - SUPERCOMPUTER-SUPPORT UND FORSCHUNG FÜR COMMUNITIES

Keine Angst vor großen Rechnern

Seit 2008 gibt es am Jülich Supercomputing Centre (JSC) Simulation Laboratories. Diese Supporteinrichtungen bestehen aus Mitarbeiterteams, die selbst aktiv Forschung betreiben und daher die spezifischen Anforderungen ihrer jeweiligen Forschungscommunity im Hinblick auf den Einsatz von Supercomputern genau kennen.

Supercomputer erreichen ihre hohe Leistung durch die Aufteilung einer Berechnung auf viele Prozessoren, die Daten extrem schnell untereinander austauschen können. Solche Systeme können daher ihre Rechenleistung nur dann ausspielen, wenn die darauf laufende Software gut skaliert, d. h. so programmiert ist, dass die Berechnungsgeschwindigkeit möglichst proportional zur Anzahl der eingesetzten Prozessoren zunimmt. Neben speziell auf Parallelverarbeitung optimiertem Code erfordert dies oft auch die Verwendung völlig anderer Algorithmen als auf dem PC. Gegenwärtig verfügen nur wenige wissenschaftliche Arbeitsgruppen über Software-Entwickler mit einem entsprechenden Hintergrund, so dass es nicht selten zu Berührungsängsten mit den Großrechnern kommt und nur ein kleiner Teil der möglichen Anwendungen auf Supercomputern portiert wird. Genau hier setzt die Idee der Simulation Laboratories (kurz: SimLabs) an: Jedes der SimLabs am JSC bietet ein für die jeweils eigene Forschungscommunity (Klima, Plasmaphysik, etc.) zugeschnittenes Supportangebot rund um die Softwareentwicklung für Supercomputer.

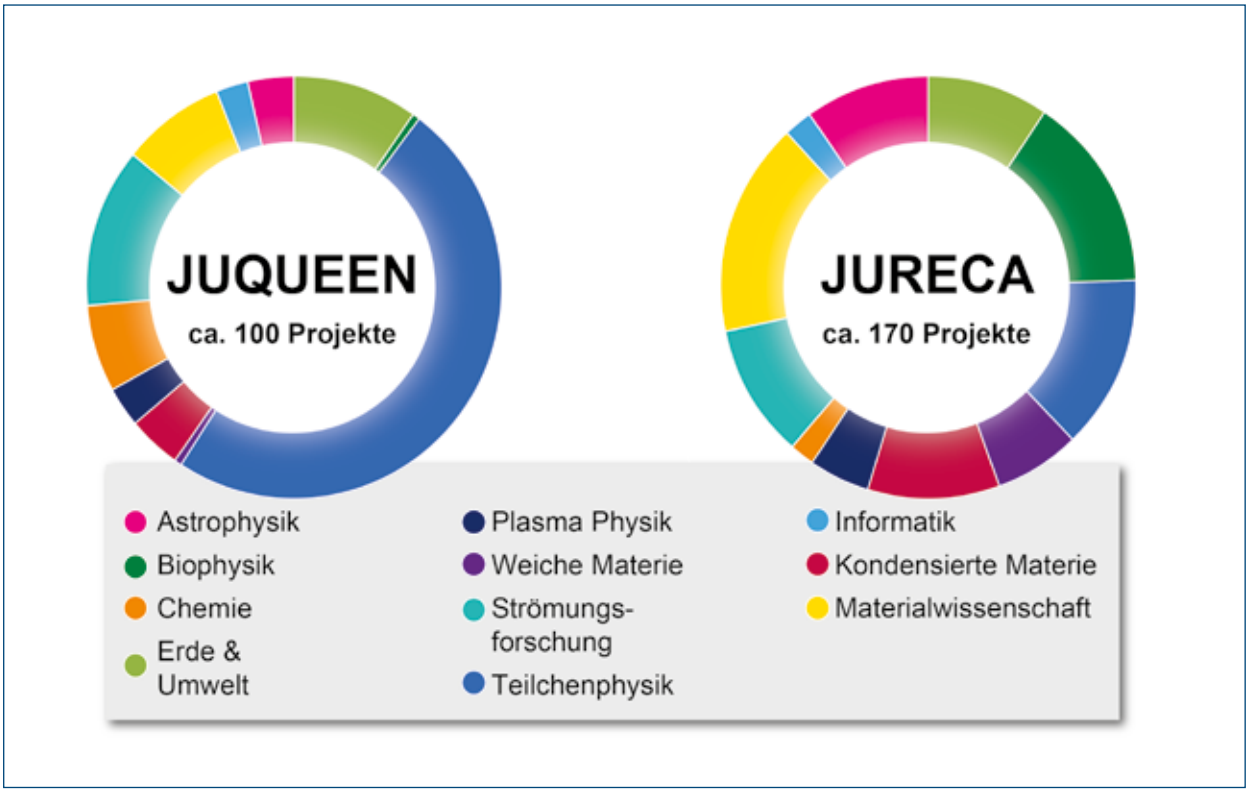


Abbildung 1: Die beiden Jülicher Supercomputer werden in vielen Wissenschaftsgebieten eingesetzt. Während JUQUEEN auf hochparallele Anwendungen spezialisiert ist, laufen auf JURECA eine Vielzahl verschiedener Programme. Der Anteil der biologischen Projekte an der auf JURECA vergebenen Rechenzeit beträgt z. Z. ca. 16 % (Stand: Nov 2015).

Quelle: © Jülich Supercomputing Centre

Abbildung 2: Der 2015 im JSC in Betrieb genommene JURECA Supercomputer erreicht mit seinen 3.768 Prozessoren und 174 Grafikkarten eine Leistung von ca. 2 Billiarden Fließkommaoperationen pro Sekunde.

Quelle: © Forschungszentrum Jülich

DAS SIMULATION LABORATORY BIOLOGY

Für die Supercomputerbelange der Life-Science-Community ist im JSC das Simulation Laboratory Biology zuständig. Es unterstützt Wissenschaftler bei der Portierung, Optimierung und Skalierung ihrer Programme. Zum Angebot gehören aber auch Kurse über parallele Programmierung in C++ und Python, sowie zur Visualisierung, Datenanalyse und zum maschinellen Lernen in Python. Das SimLab Biology dient außerdem als Ansprechpartner für Supercomputernutzer sowie für Wissenschaftler aus der Biologie, Biomedizin und Biotechnologie, die einen Antrag auf Rechenzeit planen.

Neben den Supportaufgaben unterhält jedes SimLab ein eigenes Forschungsportfolio. Das SimLab Biology hat seinen Forschungsschwerpunkt in den Bereichen Proteinstrukturvorhersage und molekulare Simulation. Während es bei der Proteinstrukturvorhersage darum geht, verschiedene bioinformatische Methoden wie Sekundärstrukturvorhersage, Faltungserkennung, Energieberechnung und Modellbewertung vorteilhaft in komplexen Workflows zu kombinieren, verwenden die Jülicher Forscher bei der molekularen Simulation den Supercomputer als eine Art Supermikroskop, um molekulare Details biologischer Prozesse besser zu verstehen. Eine Spezialität der Gruppe ist die Entwicklung von Methoden, die es erlauben, auch große Konformationsänderungen von Proteinsystemen wie sie z. B. bei der Proteinfaltung und der Aggregation von Peptiden auftreten, in atomarer Auflösung zu simulieren. Solche Prozesse, die Hunderte von Millisekunden oder noch länger dauern, liegen jenseits der Reichweite von Standardverfahren wie der Molekulardynamik. Denn obwohl sich die Berechnung der Bewegung der einzelnen Atome gut parallelisieren lässt, müssen bei der Molekulardynamik sehr viele winzige Zeitschritte nacheinander berechnet werden. Mit der am SimLab Biology entwickelten Software ProFASi, die auf einem anderen Verfahren, der Markov-Ketten Monte-Carlo-Simulation beruht, konnte 2013 erstmals die Faltung eines Proteins mit 92 Aminosäuren und einer experimentell gemessenen Faltungs-

DAS JSC :

Das Jülich Supercomputing Centre (JSC) gehört weltweit zu den führenden Höchstleistungsrechenzentren und betreibt mit JUQUEEN und JURECA zwei der größten Supercomputer Europas. Neben dem Betrieb der Rechner entwickeln die mehr als 200 Mitarbeiter des JSC neue Konzepte für energieeefiziente Rechnerarchitekturen, Werkzeuge für die Softwareentwicklung auf Parallelrechnern, Konzepte für die Analyse sehr großer Datenmengen, sowie neue hochparallele wissenschaftliche Anwendungen für eine Vielzahl verschiedener Disziplinen von der Klimaforschung bis zur Plasmaphysik. Eine zentrale Aufgabe ist die Unterstützung der wissenschaftlichen Nutzer aus ganz Europa. Sie reicht von der Bereitstellung der Rechenzeit, zugeteilt über ein vom JSC unabhängiges, wettbewerbliches und exzellenzorientiertes Peer-Review-Verfahren bis hin zu einer HPC- und fachspezifischen Individualberatung der Anwender.

dauer von 1 s simuliert werden [1]. Eine Molekulardynamik-Simulation dieses Vorgangs würde bei einem typischen Zeitschritt von 2 fs ($=2*10^{-15}$ s) 500 Billionen Schritte benötigen und viele Jahre dauern. Neben der Simulation von Proteinfaltung wird ProFASi außerdem erfolgreich zur Untersuchung der Aggregation amyloidogener Peptide eingesetzt, die im Zusammenhang mit neurodegenerativen Erkrankungen stehen, sowie zur Simulation von nativ ungefalteten Proteinen (IDPs) [2], deren Strukturpräferenzen experimentell bisher praktisch unzugänglich sind, die aber eine wichtige Rolle in vielen zellulären Prozessen wie z. B. der Krebsentstehung spielen. Interessierte Forscher erhalten im Herbst 2016 auf einer vom SimLab Biology organisierten internationalen CECAM-School eine praktische Einführung in die Monte Carlo Simulation mit ProFASi [3].

Zusammen mit Kooperationspartnern bearbeitet das SimLab Biology darüber hinaus ein großes Spektrum an weiteren Themen:

- Mit den Informatikern der benachbarten RWTH Aachen arbeitet das SimLab Biology an neuen Analysewerkzeugen für molekulare Simulationsdaten, um die bis zu 10 Millionen unabhängigen Protein-Konformationen pro Monte-Carlo-Simulation statistisch auswerten zu können.
- In Zusammenarbeit mit der Uni Bonn und der JSC-Abteilung Föderierte Systeme und Daten wurde eine Doktorarbeit zur automatischen Optimierung von wissenschaftlichen Workflows betreut.
- Im BMBF-Projekt „NGSgoesHPC“ war das SimLab Biology an der Parallelisierung von Genom-Assemblierungsmethoden beteiligt. Partner waren hierbei u. a. die Universitäten Köln und Dresden sowie die Firmen Bull und Intel.

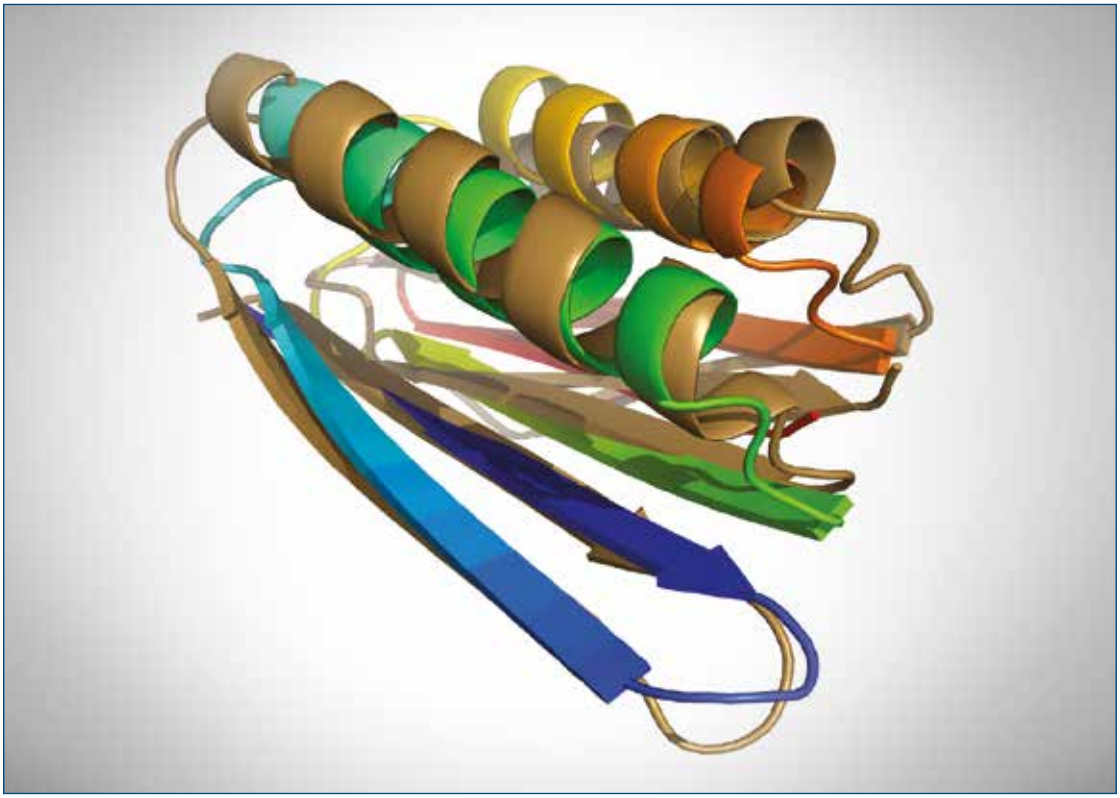


Abbildung 3: Eine durch Simulation mit der Markov Chain Monte Carlo Simulation ermittelte Struktur des TOP7 Proteins (farbig) im Vergleich mit der experimentell bestimmten Struktur (braun).

Quelle: © Sandipan Mohanty

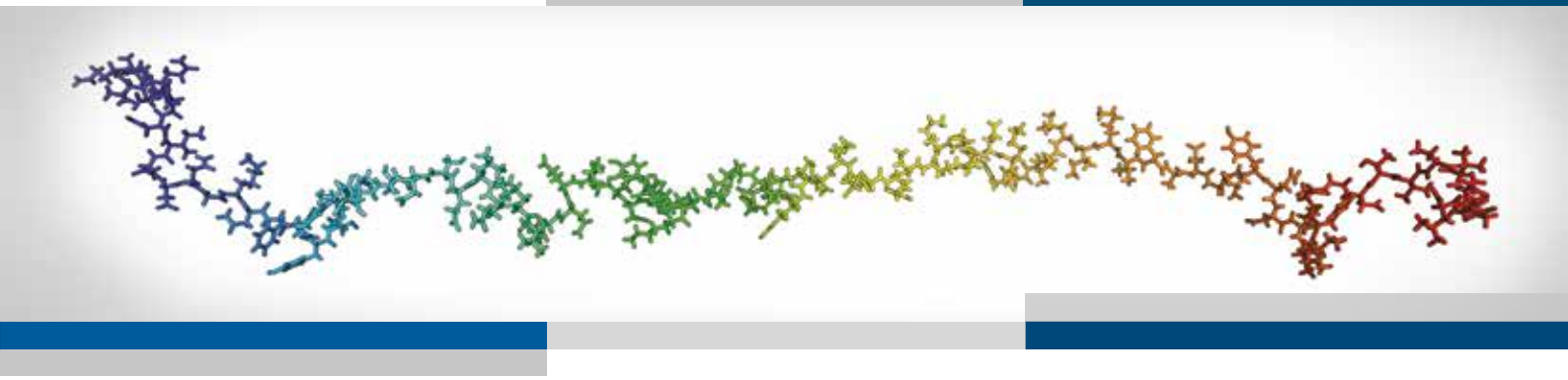


Abbildung 4: Eine ungefaltete Konformation des TOP7 Proteins mit Darstellung aller Atome. Der Farbcode entlang des Moleküls enstpricht dem von Abbildung 3.
Quelle: © Olav Zimmermann

- Einen Schritt in Richtung Zellsimulation macht das SimLab Biology in einem Kooperationsprojekt mit dem benachbarten Institut für Biotechnologie 1 (IBG-1) des Forschungszentrums Jülich (siehe Portrait in systembiologie.de Ausgabe 9). Da eine Zelle aus dem Blickwinkel der molekularen Simulation gigantisch ist, liegt eine Strategie zur Simulation solch großer Systeme mit vielen Molekülen darin, mehrere Simulationstechniken mit unterschiedlichem Detailgrad zu kombinieren. Das SimLab Biology unterstützt daher das IBG-1 bei der Entwicklung eines 3D-Multiskalenmodells zur Simulation zellulärer Prozesse. In dem vom BMBF geförderten Projekt wird die Diffusion der einzelnen makromolekularen Enzyme im dicht gepackten Zellinneren mittels Brownscher Dynamik simuliert, während die weitaus zahlreicheren Metabolitmoleküle kontinuierlich, d. h. als Konzentrationsfeld behandelt werden [4]. Mit der Kopplung der beiden Techniken sollen auf den Supercomputern praktisch relevante Systemgrößen und Zeitskalen simuliert werden, um z. B. die Auswirkung von Immobilisierung, Crowding und Multienzymkomplexen auf enzymatische Reaktionen zu studieren.

NEUE MÖGLICHKEITEN FÜR LEBENSWISSENSCHAFTLER

Supercomputer sind vielseitiger als ihr Ruf und bei rechenintensiven Aufgaben oft energieeffizienter und schneller als andere Alternativen. Zu den bioinformatischen Teilgebieten, die wie z. B. die Phylogenetik und die Proteinstrukturvorhersage von diesen Möglichkeiten bereits Gebrauch machen, werden in den nächsten Jahren weitere hinzukommen. Insbesondere solche Verfahren, die in großen Datenmengen nicht nur suchen, sondern darauf komplexe Berechnungen anwenden wie z. B. beim Training von Vorhersagemethoden mit Hilfe von maschinellen Lernverfahren können von der hohen Rechenleistung und dem schnellen Datentransfer zwischen den Prozessoren profitieren.

Zukünftige Anwendungen, welche die heute noch weitgehend brachliegenden Informationen aus biologischen und medizinischen Publikationen extrahieren, zu riesigen semantischen Netzwerken zusammenfügen und z. B. für die computergestützte Entwicklung von systembiologischen Modellen einsetzen, werden bereits angedacht. Auch erste Modelle einer ganzen Zelle sind bereits beschrieben worden. Dennoch wird selbst die Erstellung eines Modells einer einfachen Zelle, welches sich in der Simulation wie sein reales Gegenstück verhält und somit Vorhersagen auf Systemebene ermöglicht, sowohl die Lebenswissenschaften als auch die Computerwissenschaften noch vor viele Herausforderungen stellen. Beim gegenwärtigen Fortschritt in beiden Disziplinen könnte ein solches Modell aber schon in wenigen Jahren Realität werden. Bis dahin könnte am JSC ein Exaflop-Computer stehen d. h. eine Maschine, die 10^{18} Rechenoperationen pro Sekunde ausführen kann, 200-mal mehr als JUQUEEN, der aktuell schnellste Rechner in Jülich. Das JSC forscht zusammen mit Computer-Herstellern an neuen energieeffizienten Rechner-Konzepten, um den Exaflop-Computer bauen zu können und auch die Simulation einer ganzen Zelle mit vertretbarem Energieaufwand zu ermöglichen.

REFERENZEN:

[1] Mohanty, S., Meinke, J.H., Zimmermann, O. (2013)
Folding of Top7 in unbiased all-atom Monte Carlo simulations.
Proteins 81:1446–1456.
[2] Jónsson, S.A., Mohanty, S., Irbäck, A. (2012)
Distinct phases of free α-synuclein–a Monte Carlo study.
Proteins 80:2169-77.
[3] CECAM school
Atomistic Monte Carlo Simulations of Bio-molecular Systems
Jülich, 19. - 23. September 2016
http://www.cecam.org/workshop-1339.html
[4] Kondrat, S., Zimmermann, O., Wiechert, W., von Lieres, E. (2016)
Discrete-continuous reaction-diffusion model with mobile point-like sources and sinks.
The European Physical Journal E, 39:11.

WEITERE INFORMATIONEN UND KONTAKT:

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation
Jülich Supercomputing Centre (JSC)
52425 Jülich

www.fz-juelich.de/ias/jsc

ANSPRECHPARTNER:

Dr. Olav Zimmermann
Simulation Laboratory Biology
olav.zimmermann@fz-juelich.de
www.fz-juelich.de/ias/jsc/slbio

Foto: O. Zimmermann

neue chancen für die medizin

Medizininformatik wird Forschung stärken und Patientenversorgung verbessern

Interview mit Heyo K. Kroemer

Bei jedem Arztbesuch oder Krankenhausaufenthalt werden medizinische Daten erhoben. Auch in der biomedizinischen Forschung entstehen immer größere Datenmengen. Diese zumeist digital vorliegenden Informationen bestmöglich zu nutzen, ist Ziel des Förderkonzeptes Medizininformatik des Bundesministeriums für Bildung und Forschung. Heyo Kroemer, der Präsident des Medizinischen Fakultätentages, äußert sich im Gespräch über die Chancen, die ein einheitliches Management dieser Daten bieten würde.

Systembiologie.de: *Herr Kroemer, wo steht die Medizininformatik in Deutschland heute?*

Prof. Dr. Heyo Kroemer: Die digitalisierte Krankenversorgung hat ein sehr hohes Potential und ist eigentlich unverzichtbar. Klar ist aber auch, dass es in Deutschland kein einheitliches Datenmanagement gibt. Fast jedes Krankenhaus, fast jeder Standort hat ein eigenes System. Diese Systeme sind zum Teil sehr gut ausgebaut und funktionieren bestens. Aber es kann auch sein, dass eine Klinik ein herausragendes Datenmanagement hat und ein Haus weiter, wird noch auf Papier geschrieben. Das entspricht heutzutage in keiner Weise mehr den Notwendigkeiten und den technischen Möglichkeiten.

Wo liegen derzeit die Probleme?

Das Hauptproblem ist das Fehlen eines einheitlichen Datenmanagements. Es gibt keine anerkannte elektronische Patientenakte und es gibt kein Konzept, das langfristig in den Unikliniken, aber auch zwischen den Kliniken ein einheitliches Datenmanagement ermöglicht.

Wo sehen sie die größten Potentiale der Medizininformatik?

Großes Potential liegt in der Verbesserung der Patientenversorgung. Eine elektronische Patientenakte, in der Hausarzt, Facharzt und Klinikarzt einsehen können, welche Befunde vorliegen und welche Therapien verordnet sind, wäre ein großer Fortschritt. Aber auch die Forschung würde profitieren. Es gibt Unmengen von klinischen Daten, die in den Unikliniken zwar gespeichert sind, aber nicht zu Forschungszwecken verwendet werden können. Wenn die vorhandenen Daten richtig zusammengeführt und ausgewertet werden, werden sich auch die Behandlungsmöglichkeiten deutlich verbessern.

Kann man das an einem einfachen Beispiel darstellen?

Wenn Patienten von einer seltenen Erkrankung betroffen sind, an der nur beispielsweise 50 Menschen in Deutschland leiden, hat der behandelnde Arzt faktisch keine Chance andere Betroffene zu finden, bzw. mit den behandelnden Kollegen Erfahrungen auszutauschen. Mit einer ortsübergreifenden Datenbank bestände diese Möglichkeit.

Ein Hauptanreiz für die Klinik ist die Verbesserung der Versorgungsabläufe. Wie kann man sich das vorstellen?

Wenn ein Patient heutzutage in eine Klinik kommt, veranlasst der behandelnde Arzt Untersuchungen, die unter Umständen schon im Vorfeld stattgefunden haben. Hätte er eine elektronische Patientenakte mit allen relevanten Informationen ließen sich viele Doppeluntersuchungen vermeiden. Das ist einer der Gründe warum die USA diese Akten praktisch flächendeckend durchgesetzt haben.

Heyo Kroemer ist Professor für Pharmakologie und personalisierte Medizin an der Universität Göttingen, Dekan der Universitätsmedizin und Sprecher des Vorstands für Forschung und Lehre. Er sitzt außerdem im Senatsausschuss für Sonderforschungsbereiche der DFG und ist Präsident des Medizinischen Fakultätentages, dem Zusammenschluss der Medizinischen Ausbildungs- und Forschungsstätten Deutschlands (Foto: Irene Böttcher-Gajewski / Universität Göttingen).

Was muss aus Ihrer Sicht getan werden, um die Situation zu verbessern?

Der allererste Schritt ist getan: Wir haben das Problem als solches erkannt. Was wir jetzt brauchen, sind intelligente Lösungsstrategien, die über einzelne Standorte hinausgehen. Und ich glaube, dass das Förderkonzept des BMBF zur Medizininformatik sehr dazu beitragen kann: Es zielt auf einen standortübergreifenden Zusammenschluss von Unikliniken, da keine Einzelantragstellung mehr möglich ist.

Wie groß ist die Bereitschaft gemeinsam an dem Problem zu arbeiten?

Allen Beteiligten ist bewusst, dass wir in diesem Bereich ein großes Defizit haben. Dementsprechend groß ist die Bereitschaft das Problem anzugehen. Und wenn in den kommenden Jahren noch parallele Programme aufgebaut werden, die nicht wie dieses BMBF-Programm überwiegend der Forschung zugute kommen, sondern auf eine Verbesserung der Krankenversorgung abzielen, dann wird man dieses Defizit in endlicher Zeit beheben können.

Wo sehen Sie die größten Hürden?

Das wird man im Laufe dieses Programmes sehen. Sicherlich ist der Datenschutz in Deutschland die größte Hürde. Wenn man sich allerdings frühzeitig damit auseinandersetzt, Vorgänge transparent gestaltet und von vornherein die Datenschützer mit einbindet, kann man diese Hürde erfahrungsgemäß erfolgreich nehmen.

Dazu kommt die mangelnde Bereitschaft der Politik, aber auch der einzelnen Universitätskliniken, die Medizininformatik ausreichend zu finanzieren. Im Moment werden lediglich unter einem Prozent des Umsatzes für Informationstechnologie ausgegeben. Will man wettbewerbsfähig werden, müssen diese Ausgaben auf fünf bis sechs Prozent steigen.

Ist der im Förderkonzept veranschlagte Zeitrahmen ausreichend?

Das Förderkonzept ist sehr langfristig angelegt. Und das ist auch gut so. Denn das Programm wird die Abläufe in den Universitätskliniken fundamental ändern. Das können wir nicht von Heute auf Morgen erreichen.

Was sind aus Ihrer Sicht die Stärken des Förderkonzeptes?

Die größte Stärke ist die konsequente Ausrichtung von IT-Strategien über Standorte hinaus. Das hat es in Deutschland noch nicht gegeben. Und ich glaube auch, dass das Programm weit über die IT hinaus Bedeutung haben wird. Was wir im Moment betreiben, die alleinstehende Universitätsmedizin, die an einem Ort alle Technologien vorhält, wird man aus Kostengründen zukünftig nicht mehr aufrechterhalten können. Aus meiner Sicht bietet das Konzept einen guten Anreiz, dieses fundamentale Problem jetzt anzugehen.

Foto: everythingpossible – Fotolia.com

Welche Effekte werden wir außerhalb der Universitätskliniken sehen?

Eine der großen deutschen Schwächen ist praktisch das völlige Fehlen einer Sektor-übergreifenden Kommunikation. Kommt man nach einem längeren Krankenhausaufenthalt beispielsweise in eine Pflegeeinrichtung, wäre es naheliegend, dass auch alle elektronischen Patientendaten weitergegeben werden. Das ist leider nicht so. Hier besteht eine weitere große Herausforderung.

Betrifft die Medizininformatik denn ausschließlich die bisher genannten Bereiche?

Der Austausch medizinischer Daten ist ein sensibler Bereich. Natürlich müssen in der Medizininformatik neben den rein technischen und medizinischen Problemen deshalb auch die juristischen Fragen und ethischen Bedenken diskutiert werden. Genau das stößt das Förderkonzept aber auch an. Es sieht vor, dass diese Aspekte durch begleitende Aktivitäten beleuchtet werden und ein gesamtgesellschaftlicher Diskurs eröffnet wird.

Wo wird Deutschland in zehn Jahren stehen?

Ich bin optimistisch. Wir werden in zehn Jahren in einer Situation sein, dass in der Universitätsmedizin durchgehend elektronische Patientenakten geführt werden, die mit entsprechenden Biobanken verknüpft sind, und dass der Austausch innerhalb und zwischen den Unikliniken problemlos möglich ist.

Das Interview führte Katja Nellissen. Redaktionelle Bearbeitung Marco Leuer und Bettina Koblenz.

Informationen zum „Förderkonzept Medizininformatik" finden Sie unter:
http://www.gesundheitsforschung-bmbf.de/de/medizininformatik.php

Was ist Medizininformatik?

Die Medizininformatik „ist die Wissenschaft der systematischen Erschließung, Verwaltung, Aufbewahrung, Verarbeitung und Bereitstellung von Daten, Informationen und Wissen in der Medizin und im Gesundheitswesen. Sie ist von dem Streben geleitet, damit zur Gestaltung der bestmöglichen Gesundheitsversorgung beizutragen".
(Definition Medizinische Informatik, Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V.)

Kontakt:

Prof. Dr. Heyo K. Kroemer
Vorstand Forschung und Lehre
Dekan der Medizinischen Fakultät
UNIVERSITÄTSMEDIZIN GÖTTINGEN
GEORG-AUGUST-UNIVERSITÄT
Göttingen

www.med.uni-goettingen.de/de/content/ueberuns/dekanat.html

i:DSem – integrative datensemantik in der systemmedizin

Neue Initiative des Bundesministeriums für Bildung und Forschung für innovatives Datenmanagement in der biomedizinischen Forschung

von Christian Rückert

In der Systemmedizin möchte man riesige Mengen von patientenbezogenen Daten zusammenführen und nutzen: Es gilt, die bestmögliche Behandlung für einen bestimmten Patienten zu finden; und zwar vor dem Hintergrund des Wissens, das aus der Forschung und den Erfahrungen mit vielen anderen Patienten stammt. Das ist heute aber noch nicht möglich. Um die systemmedizinische Forschung voranzubringen, müssen eine Vielzahl klinischer Daten, die teils unstrukturiert und in unterschiedlicher Qualität vorliegen, sowie krankheitsrelevante molekulare Daten, strukturiert, zusammengefasst, aufbereitet und schließlich den Ärzten für ihre Entscheidung über die wirksamste Behandlung zur Verfügung gestellt werden.

Integrative Datensemantik – Dienste für die Systemmedizin

Arztpraxen und Krankenhäuser erfassen riesige Mengen an Daten über ihre Patienten: Blutwerte, die Liste der verordneten Medikamente, Vorerkrankungen, Bilder aus Röntgen-, MRT- oder Ultraschalluntersuchungen, pathologische Befunde, Vermerke über Allergien und Unverträglichkeiten und vieles mehr. Während einige dieser Daten digital und strukturiert vorliegen, wie das bei den Blutwerten üblicherweise der Fall ist, kann davon bei anderen keine Rede sein. So sind die Ergebnisse bildgebender Verfahren oft in digitalen Formaten verfügbar, was sie zeigen, deutet aber der Mediziner auf Basis seiner Ausbildung und Erfahrung. Automatisierte Analysen auf der Grundlage von Mustern sind hier noch nicht möglich.

Wichtige Informationen für die Behandlung der Patienten finden sich auch in Ärztebriefen und Berichten. Diese sind jedoch oft nicht digitalisiert oder höchstens als Scan der Akte beigefügt und sind damit inhaltlich für IT-Systeme nicht verwertbar. Selbst wenn sie textlich erfasst sind, so sind sie immer noch in natürlicher Sprache verfasst und somit für eine unmittelbare automatisierte Weiterverarbeitung kaum geeignet.

Über diese klinischen Daten hinaus werden in Zukunft molekulare Untersuchungen eine größere Rolle spielen. Ergebnisse aus Hochdurchsatzanalysen, Genominformationen des Patienten oder von Tumoren, Proteom- und Metabolomanalysen werden dem Systemmediziner der Zukunft Diagnose- und Behandlungshinweise liefern, die heute noch nicht verfügbar sind. Insofern adressiert die Integrative Datensemantik einen fundamentalen Teilaspekt des innovativen Datenmanagements. Sie trägt damit wesentlich zur De-facto-Standardisierung von Datensätzen bei, die für jede Form der Datenintegration unverzichtbar ist.

Um diese vielgestaltigen Daten zu nutzen, müssen sie zunächst für den Computer verwertbar gemacht werden. Strukturierung, semantische Beschreibung und Integration sind da Stichwörter, die noch vor der Analyse und Informationsaggregation stehen. Dazu fehlen gegenwärtig geeignete Integrationskonzepte und -systeme. Ontologien für die verschiedenen Anwendungen müssen entwickelt und Datenstrukturen etabliert werden.

Dieser Problematik nimmt sich die BMBF-Initiative *„i:DSem – Integrative Datensemantik in der Systemmedizin“* an. Eingebettet in das Modul *„Zukunfts- und Querschnittsthemen“* des Forschungs-

Foto: santiago silver - Fotolia.com

und Förderkonzeptes *e:Med*, bietet *i:DSem* die Möglichkeit, auf diesen Innovationsbedarf zu reagieren.

Zugleich liefert die Integrative Datensemantik das Fundament für die Entwicklung technischer und methodischer Innovationen, um die bestehende und insbesondere die zukünftig weiter ansteigende Datenflut in den Lebenswissenschaften inhaltlich effektiver zu verwerten und damit Forschungsergebnisse zum Wohle des Patienten schneller in die klinische Anwendung zu überführen.

i:DSem-Verbünde starten 2016

Ab 2016 wird das Bundesministerium für Bildung und Forschung acht Forschungsverbünde mit etwa 20 Millionen Euro bis zu fünf Jahre fördern. Die ausgewählten Projekte beschäftigen sich mit verschiedenen epidemiologisch bedeutsamen Krankheiten. Gleich fünf Verbünde arbeiten an verschiedenen Ansätzen, die die bessere medizinische Versorgung von Krebspatienten ermöglichen und die systemmedizinische Forschung in diesem Bereich unterstützen sollen.

Je nach Schwerpunkt dieser Krebs-bezogenen Projekte steht die Integration von Sequenzdaten der genetischen Veränderungen der Krebszellen gegenüber gesunden Körperzellen, die Aufklärung von Zell-Signalnetzwerken oder die Nutzbarkeit der Daten für den Mediziner im Fokus. Damit Ärzte während der Behandlung von Krebserkrankungen schnell und intuitiv auf die große Menge bereits existierenden Wissens zugreifen können, müssen nicht nur die Datenbestände für Computersysteme verständlich beschrieben und vorgehalten werden. Auch die Entwicklung von intuitiv zu bedienenden Informations-Präsentations-Systemen für Ärzte wird in den Projekten vorangetrieben. Dies sind interaktive klinische Anwendungen, die dem Arzt am Computer erlauben – beispielsweise basierend auf der Kenntnis genetischer Veränderungen beim einzelnen Patienten – die individuell beste Therapie für den Patienten im Rahmen der bekannten Behandlungsmöglichkeiten auszuwählen.

Zwischen der semantischen Beschreibung der Daten und ihrer Anwendung in der Patientenversorgung liegt noch eine Verarbeitungsebene, die dem Endanwender kaum sichtbar und doch von großer Wichtigkeit ist: Für die Darstellung und Klassifizierung von Entscheidungsgrundlagen und Entscheidungsmöglichkeiten bedürfen die Computer-Anwendungen Phänotyp-Profile, diagnostische Klassifikatoren, Risikoprofile und Krankheitsmodelle. Diese notwendigen Vorarbeiten sowie Instrumente zur Modellsimulation von biologischen Prozessen in der Systemmedizin gehören ebenfalls zu den Projektarbeiten im Rahmen von *i:DSem*.

Neben den onkologischen Erkrankungen beschäftigt sich ein Projekt mit neuro-degenerativen Erkrankungen wie der Demenz. Das Ziel ist, die verfügbaren Patientendaten für möglichst ganzheitliche Analysen zugänglich zu machen. Darauf aufbauend sollen die neu entwickelten Methoden dann einen Beitrag zu einer verbesserten Früherkennung von Demenzerkrankungen und zu einer optimierten Patienten-Stratifizierung liefern. Ferner wird angestrebt, die Identifikation neuer prognostischer Faktoren zu verbessern und im Idealfall auch die Entwicklung neuer Therapien zu fördern.

Ein weiteres Projekt widmet sich Rückenmarksverletzungen: Alleine in den Industrieländern geht man von 250.000 bis 500.000 akuten Rückenmarkspatienten aus, für die bisher keine erfolgreichen therapeutischen Verfahren zur Verfügung stehen. Dies ist vor allem der unüberschaubaren Menge unstrukturierten Wissens geschuldet, die sich in enormen Beständen einschlägiger Forschungsliteratur verbirgt. Das Projekt zielt daher auf die Entwicklung eines Informationssystems, das umfassendes Wissen für Neurowissenschaftler bereitstellt. Um dieses mit Wissen zu füllen, werden zunächst neuartige Verfahren entwickelt, die aus wissenschaftlichen Veröffentlichungen automatisiert Informationen extrahieren. Begleitend hierzu wird ein Bewertungsverfahren entwickelt, das auf dieses Informationssystem zugreift und dem Wissenschaftler oder Arzt bei der Auswahl von erfolgsversprechenden therapeutischen Konzepten behilflich ist. Basierend auf diesem Informationssystem sollen dann klinische Studien entworfen werden, die große Verbesserungen in der Patientenbehandlung versprechen.

Eine ähnliche Vorgehensweise – jedoch mit Anwendung in der Transplantationsmedizin – liegt einem weiteren Projekt zugrunde: Es beschäftigt sich mit der datenschutzkonformen Beschreibung und Verwertung von klinischen Datenbeständen aus unterschiedlichen Quellen, um daraus Vorhersagemodelle zu entwickeln. Dazu soll eine IT-Plattform geschaffen werden, mit deren Hilfe klinische Daten harmonisiert und anonymisiert in einer zentralen Datenbank gesammelt und zum weiteren Gebrauch zur Verfügung gestellt werden. Im weiteren Verlauf des Projektes ist geplant, diese Datenbank für systembiologische Modellierungen zu nutzen, insbesondere auf dem Gebiet der Transplantationsmedizin. Mit diesen Modellen sollen mögliche Komplikationen bei Stammzelltransplantationen vorhergesagt und so die Lebensqualität der Betroffenen verbessert werden.

Neben integrativen datensemantischen Arbeiten im engeren Sinne müssen grundlegende Dienste entwickelt und bereitgestellt werden, die für den Umgang mit personenbezogenen Gesundheitsdaten notwendig sind. Ontologien und Krankheitsmodelle sind genauso im Fokus dieser Projekte wie die Integration verschiedener Datenquellen und ihre optimale Verwertung. Nur wenn die Mediziner am Ende mit geringem Aufwand auf die Daten zugreifen können und sie übersichtlich dargestellt bekommen, werden die Systeme Einzug in den Klinikalltag erhalten.

Von den Grundlagen zur Anwendung

Von Beginn an wurde die Translation der Ergebnisse als wichtiges Element der Fördermaßnahme *i:DSem* angesehen. Die Forschungsverbünde werden nach drei Jahren, am Ende der Entwicklungsphase, evaluiert und gehen nur bei einem positiven Votum in die zweijährige Translationsphase. Schon bei der Einreichung der Projektanträge mussten die Wissenschaftler darlegen, auf welche Weise eine Nutzung ihrer Arbeiten erfolgen kann und eine erste Anwendung als *Proof of Concept* in den Arbeitsplan der Translationsphase einarbeiten.

So ist es wenig überraschend, dass die meisten Verbünde über Kliniken oder Softwareanbieter aus dem klinischen Bereich als Translationspartner verfügen. Auch die Information klinischer Kollegen ist in vielen Fällen Bestandteil der Projekttätigkeit. Symposien, Workshops und Lehrveranstaltungen sind fest mit eingeplant. In den meisten Projekten trägt das medizinische Personal auch schon in der Entwicklungsphase dazu bei, dass die entwickelten Systeme im Anschluss klinisch verwendbar werden.

Kontakt:

Dr. Christian Rückert
Projektträger Jülich
Forschungszentrum Jülich GmbH
c.rueckert@fz-juelich.de

www.ptj.de/idsem

Foto: FZJ, R. U. Limbach

„projekte können auch auf eine gute art scheitern“

Interview mit Olaf Wolkenhauer

Olaf Wolkenhauer war Deutschlands erster Professor für Systembiologie. Seit nunmehr 12 Jahren lehrt er an der Universität Rostock. Im Interview mit systembiologie.de spricht er über seine Forschungspläne für die Zukunft und erinnert sich an seine erste Begegnung mit der Molekularbiologie. Mit seinem Wechsel in die Systembiologie setzte der Ingenieur damals alles auf eine Karte, wie er sagt.

Systembiologie.de: *Vor nunmehr 12 Jahren bekamen Sie den Ruf an die Uni Rostock. Wie hat sich die Systembiologie-Landschaft in Deutschland seitdem verändert?*

Prof. Dr. Olaf Wolkenhauer: Die Systembiologie hat sich inzwischen in der Forschungslandschaft fest etabliert. Das ist maßgeblich den Fördermaßnahmen des Bundesforschungsministeriums zu verdanken. Sie haben die interdisziplinäre Arbeit aktiv unterstützt.

Wenn Sie an die Anfänge zurückdenken. Welchen Stellenwert hatte die Systembiologie damals noch als Forschungsansatz?

Zu diesem Zeitpunkt waren Kooperationen zwischen Medizinern und Biologen mit uns Modellierern noch mit einem hohen Risiko verbunden. Und zudem mit einem großen Fragezeichen versehen, ob man überhaupt eine Chance hatte, in irgendeiner Form gefördert zu werden. Doch die unterschiedlichen Disziplinen in der Systembiologie sind im Laufe der Zeit zusammengewachsen. Wir müssen heute gar nicht mehr großartig argumentieren, dass es Sinn macht, mathematische Modelle zu erstellen. In vielen Projekten ist den Wissenschaftlern von Anfang an völlig klar, dass die Zusammenarbeit für beide Seiten eine gute Sache ist.

Wer geht denn eher auf wen zu. Die Biologen und Mediziner auf die Modellierer oder umgekehrt?

In den ersten Jahren musste ich als Modellierer auf die Biologen zugehen – ganz am Anfang ohne Erfolg. Ich wollte das Verhalten von Zellen als Prozesse verstehen, Netzwerke studieren und die Zeitabläufe, die die Funktion von Zellen beeinflussen, untersuchen. Doch dafür haben sich die Biologen anfangs gar nicht interessiert. Das hat sich dann irgendwann erfreulicherweise geändert. Inzwischen ist es so, dass man einfach zusammenkommt und es passt. Und das ist toll.

„Ich werde mich um die Gesetzmäßigkeiten kümmern“

Sie haben ursprünglich Regelungstechnik studiert. Wie sind Sie überhaupt zur Systembiologie gekommen?

Ein Grundinteresse an Biologie war bei mir schon immer da. Ich wollte gerne Biologie studieren, meine Eltern haben mir das aber nicht erlaubt. Wir unterstützen Dich nur, wenn Du Ingenieurwissenschaften studierst, haben sie gesagt. Und das war im Rückblick auch gut so. Denn in meinem Studium habe ich gelernt, dynamische Systeme und zeitliche Prozesse mit mathematischen Modellen zu beschreiben. Und wenn ich mir die damaligen Biologiebücher angeguckt habe, dann waren diese Netzwerke, die wir heute häufig in der Systembiologie untersuchen, der Endpunkt. Das heißt, man hat Experimente gemacht, um zu zeigen, dass einzelne Komponenten miteinander interagieren und damit war Schluss. Für mich, der gelernt hatte, Prozesse zu beschreiben, wäre das aber erst der Startpunkt für eine mathematische Modellierung gewesen. Das erschien mir ganz logisch. Das wurde aber Anfang der 90er Jahre noch nicht gemacht.

Was ist heute ihr Forschungsschwerpunkt? Mit welchen Themen beschäftigen Sie sich derzeit?

Mein Lieblingsthema ist die Zellkommunikation, zum Beispiel in der Krebsforschung. Wir betrachten Zellen nicht mehr isoliert, sondern im Wechselspiel mit ihrer Umgebung. Eine ganz wichtige Entwicklung für die Systembiologie war die Erkenntnis, dass wir Prozesse wie Metastasierung nur dann verstehen, wenn wir

Olaf Wolkenhauer (Quelle: Universität Rostock, ITMZ, Foto: E. Altrichter).

Netzwerke untersuchen. Doch bisher haben wir immer mehr reingezoomt, um molekulare Details zu entdecken. Für mich ist das große Thema für die Zukunft aber das Rauszoomen, die Fähigkeit, die Zelle in ihrem Kontext zu betrachten. Ich werde jetzt wieder einen radikalen Schnitt machen und alles auf eine Karte setzen: Ich werde mich um die Gesetzmäßigkeiten kümmern. Um Krankheitsprozesse zu verstehen, brauchen wir Methoden zum Rauszoomen. Ich möchte den Zusammenhang zwischen konkreten Ergebnissen und dem übergeordneten Thema herstellen. Es geht darum, die vielen Details zusammenzubauen zu einem vollständigen Bild.

„Mein Chef hat sich Sorgen gemacht"

Wie könnte diese Verallgemeinerung aussehen von der Sie sprechen?

Ich möchte das vorhandene Wissen miteinander in Beziehung setzen. Das geschieht bislang schon, aber nur in Review-Artikeln. In diesen Analysen werden die in Zusammenhang stehenden Annahmen verbal formuliert. Etwas Ähnliches würde ich gerne mit mathematischen Interpretationen machen. Wenn der Biologe sagt, dass dies oder jenes etwas anderes reguliert, dann möchte ich dieses Wort „regulieren" übersetzen und in einen formalen, mathematischen Zusammenhang bringen. Das kann grandios schiefgehen. Aber dazu bin ich bereit. Das heißt, wir machen in unserer Gruppe weiterhin Projekte, bei denen wir den Details auf der Spur sind. Aber ich will darüber hinaus eine Vogelperspektive einnehmen. Wenn es überhaupt so etwas wie Gesetzmäßigkeiten in der Biologie gibt, dann sind allgemeinere Vorhersagen möglich. Jedes Modell hat dieses Risiko. Es geht immer um die Frage, wie kann ich in einem Modell Dinge vereinfachen, ohne dass das Modell seine Aussagekraft verliert.

Sie sprachen eben davon, dass sie jetzt wieder alles auf eine Karte setzen und wie einst auf Risiko spielen wollen. Hatten Sie dabei Ihren Wechsel in die Systembiologie im Sinn?

Genau. Ich habe in der Regeltheorie promoviert. Als ich dann zur Systembiologie wechseln wollte, haben sich der Chef des Instituts und andere ältere Kollegen Sorgen gemacht. Sie befürchteten, dass ich mir mit der Ausrichtung hin zur Biologie die Karriere ruiniere. Jetzt brauche ich mir allerdings im Unterschied zu damals keine Sorgen mehr zu machen, weil ich das Glück habe, Professor zu sein. Jetzt habe ich die Freiheit.

Ihr biologisches Interesse vor allem an Krankheitsverläufen hat auch einen persönlichen Hintergrund.

Ja, ganz klar. Ein Schlüsselmoment für mich war die Erkrankung meines Vaters. Die Ärzte sagten, da könne man nichts machen. Durch die Mutation eines Moleküls im Blut wurde seine Lunge langsam zerstört. Ich habe versucht, das zu verstehen. Ich bin in einen Buchladen gerannt und habe dort meine erste Berührung gehabt mit der Biologie von Zellen. Vorher war Biologie für mich gleichbedeutend mit Tieren und Pflanzen und eng mit meiner Faszination für die Natur verbunden. Und dann waren es auf einmal die Zellen, die mich faszinierten. Da war mir plötzlich

klar, Mensch, ich habe das hier studiert, es ging schließlich auch um dynamische Prozesse. Das war für mich die Initialzündung. Ich habe gesehen, mit dem was ich gelernt habe, kann ich auch auf diesem Gebiet etwas ausrichten. Dass die Biologie natürlich viel komplizierter ist als die Ingenieurwissenschaften, musste ich dann auch erfahren.

„Komplexität ist ein Ansporn, kein Frust"

Hat Sie diese Erfahrung jemals frustriert?

Nein, die Natur ist einfach so komplex. Aber diese Vielfalt ist auch der Grund, warum es so schön und so spannend ist, sich mit ihr zu beschäftigen. Mich hat das immer motiviert. Komplexität ist ein Ansporn, kein Frust. Und dazu gehört auch, dass wir in einigen Projekten festgestellt haben, dass die Dinge komplexer sind als wir dachten. Das ist natürlich zunächst einmal negativ, weil wir unser Ziel nicht erreicht haben. Aber es ist trotzdem ein Fortschritt, wenn man sich dieser Komplexität bewusst wird und man weitermachen kann. So können Projekte auch auf eine gute Art und Weise scheitern. Die Gründe des Scheiterns sind auch ein Erkenntnisgewinn.

Was würden Sie einem jungen Wissenschaftler empfehlen, der systembiologisch arbeiten möchte?

Ob es gut ist, einen eigenen Studiengang für Systembiologie einzurichten, weiß ich nicht. Wir unterrichten erst auf dem Master-Level und unsere Erfahrungen damit sind sehr gut. Die Leute studieren zunächst Physik, Informatik, Biologie oder Medizin. Erst dann spezialisieren sie sich und versuchen, über den Tellerrand zu schauen. Das gelingt nicht unbedingt von Anfang an. Ich empfehle also die Spezialisierung über ein Master-Studium oder die Promotion.

Sie haben die Systembiologie auch schon von ihrer philosophischen Seite betrachtet. Was war dabei Ihre wichtigste Erkenntnis?

Für mich ist Arthur Schopenhauer jemand, dessen Arbeiten mich sehr interessieren. Sie helfen mir bei meinen Überlegungen, wie ich an die Dinge herangehe. Die Philosophen nehmen eine Vogelperspektive ein und beobachten so, wie sich die Wissenschaften entwickeln. Es kann durchaus sein, dass wir in ein paar Jahren über viele Dinge, die wir in der Systembiologie gemacht haben, lachen, beispielsweise über die Untersuchung einzelner Signalwege. Es könnte naiv erscheinen zu glauben, dass man ein isoliert betrachtetes Teilsystem dazu nutzen kann, ein komplex interagierendes, großes Ganzes zu verstehen. Und die Philosophie hilft einem, diese Vogelperspektive einzunehmen. Deshalb finde ich es fruchtbar, mit Wissenschaftsphilosophen zusammenzuarbeiten, weil diese einem zeigen, welche Irrwege in anderen Fächern wie der Physik gegangen wurden und welche Rolle die Modellierung dort spielt. Die Beschäftigung mit der Philosophie ermöglicht die Reflektion über das eigene Handeln. Im Alltag sind wir leider oft so beschäftigt, dass kaum noch Zeit bleibt für diese Reflektion.

Das Gespräch führten Melanie Bergs und Gesa Terstiege.

Kontakt:

Prof. Dr. Olaf Wolkenhauer
Systems Biology and Bioinformatics
Universität Rostock
olaf.wolkenhauer@uni-rostock.de

www.sbi.uni-rostock.de

Quelle: Universität Rostock, ITMZ, Foto: E. Altrichter

„die leber ist erste wahl für wissenschaftler und unser netzwerk“

Interview mit Peter Jansen
Programmdirektor des Forschungsnetzwerks LiSyM – Systemmedizin der Leber

Forschung und Arbeit mit Patienten gingen bei Peter Jansen immer Hand in Hand. Seit den 70er Jahren hat sich der Mediziner der Leber verschrieben. Als Programmdirektor repräsentiert und koordiniert der Niederländer seit Beginn des Jahres das neue Forschungsnetzwerk LiSyM. Im Gespräch mit systembiologie.de berichtet er über seine Ziele für die kommenden Jahre und erklärt seine Faszination für das Organ Leber.

Systembiologie.de: *Was hat Sie an der Aufgabe als Programmdirektor des Netzwerks LiSyM gereizt?*

Prof. Dr. Peter Jansen: Mit meiner Arbeit als Programmdirektor betrete ich nach Jahrzenten in der Leberforschung noch einmal Neuland. Ich bin es wie viele meiner Kollegen aus der klinischen Forschung nicht gewohnt, mit Modellierern zusammenzuarbeiten. Wir haben *in vitro* Experimente auf der einen und klinische Daten auf der anderen Seite. Dazwischen liegt immer eine Lücke, weil wir nicht wissen, ob und wie sich die Ergebnisse aus dem Labor auf den Menschen übertragen lassen. Die Modellierer können hier eine Verbindung schaffen. Diese Zusammenarbeit hat mich besonders gereizt und ich bin sehr gespannt, wie sie die medizinische Forschung am Ende weiterbringen kann. Ich bin mir schon jetzt sicher, dass es ein

Die menschliche Leber fasziniert Peter Jansen seit den siebziger Jahren. *„Es gibt so viel Wissen und Daten über die Leber, dass die Forschung hier nie zum Stillstand kommt.“*, sagt er.

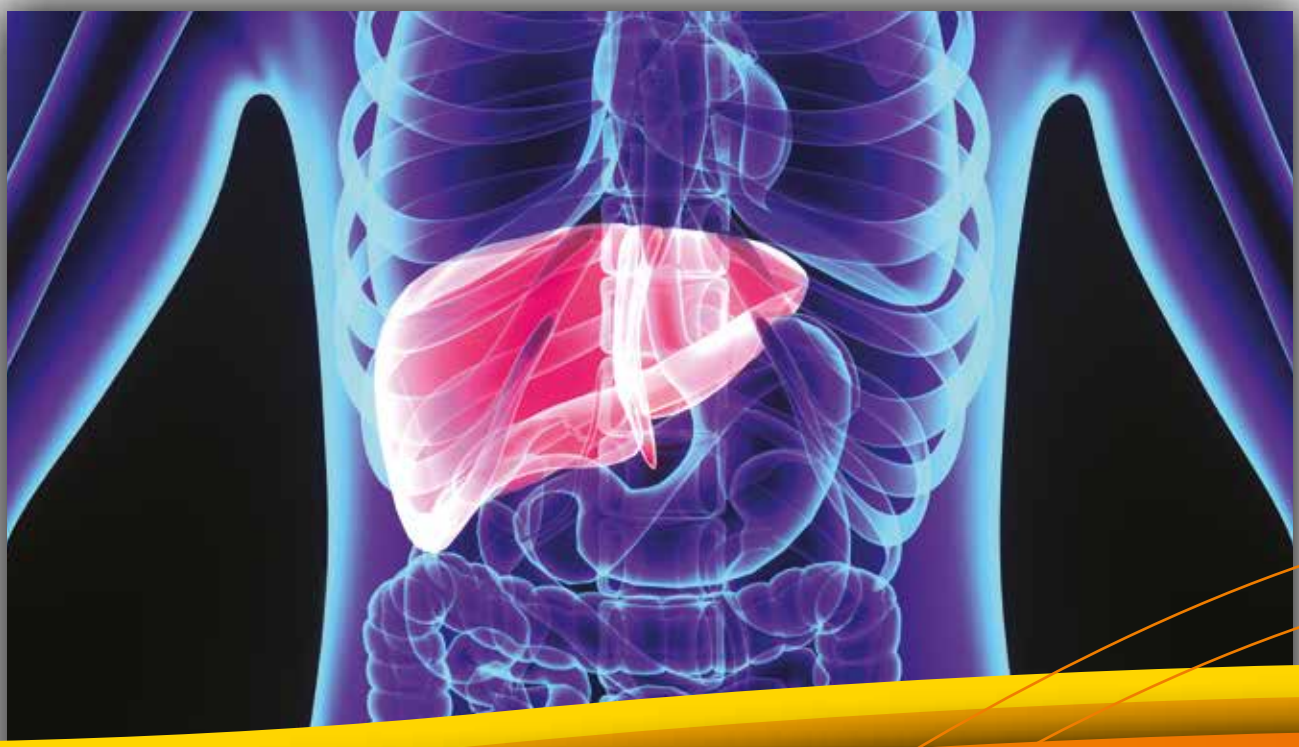


Foto: 7activestudio – Fotolia.com

Der niederländische Medizinprofessor Peter Jansen erwartet ganz neue Einblicke durch seine Aufgabe als Programmdirektor des Forschungsnetzwerks LiSyM (Foto: Academic Medical Center in Amsterdam).

sehr interessanter wissenschaftlicher Diskurs werden wird. Das Bundesforschungsministerium hat hier etwas ganz Neues geschaffen.

Was sind Ihre Ziele für die kommenden Jahre? Was wollen Sie im Netzwerk erreichen?

Mein Wunsch wäre, dass Toxikologen, Medikamentenentwickler und Kliniker das Netzwerk LiSyM als Prototypen für die medizinische Forschung der Zukunft betrachten. Bisher ist es oft sehr enttäuschend, wenn man die Ergebnisse der molekularen *in vitro* Forschung oder von Tierversuchen direkt auf die Klinik und damit auf den Menschen überträgt. Es gibt große Unterschiede in der Wirkung eines Medikaments. *In vitro* Experimente und Tierversuche wecken daher leider oft unrealistische Erwartungen an klinische Erfolge. Die Modellierer können diese Lücke hoffentlich schließen und zu einem besseren Verständnis bei der Übertragung von Ergebnissen auf der molekularen Ebene auf den ganzen Organismus beitragen. In Zukunft kann man dann vielleicht sogar ganz auf Tierversuche verzichten.

Sie sind bereits seit den 70er Jahren in der Leberforschung aktiv. Was fasziniert Sie so an diesem Organ?

Alles wird faszinierend, sobald man mehr darüber weiß. Ich habe jetzt fast mein ganzes Leben mit Leberforschung verbracht. Ich weiß mittlerweile also eine ganze Menge darüber und es wird jeden Tag spannender. Es gibt so viele biologische Signalwege, die man in der Leber untersuchen kann. Nahezu die ganze Biochemie ist aus Leberexperimenten entstanden. Es gibt so viel Wissen und Daten über die Leber, dass die Forschung hier nie zum Stillstand kommt. Auch bei LiSyM zeigt sich, wie wichtig die zahlenreichen Datensätze und Studien zur Leber sind. Über andere Organe ist längst nicht so viel Wissen verfügbar. Die Leber ist damit erste Wahl für Wissenschaftler und für unser Netzwerk.

In Ihrem Beruf schlagen Sie die Brücke zwischen Forschung und medizinischer Praxis. Wie befruchten sich diese beiden Bereiche Ihrer Arbeit gegenseitig?

Wenn ich die Patienten treffe und ihre Probleme sehe, dann ergeben sich gleich so viele klinische Fragestellungen, die ich lösen möchte. Ich schaue mir die Patienten und ihre Labordaten an und frage mich, warum bestimmte Enzyme erhöht sind oder bestimmte Therapien nicht anschlagen. Und wenn man mit Forschern im Labor zusammenarbeitet, ergeben sich noch ganz neue Fragestellungen. Als ich in den 80er Jahren angefangen habe, haben sich die Forscher im Labor nicht dafür interessiert, was in der Klinik passiert. Das hat sich inzwischen zum Glück

geändert. Wenn die Grundlagenforscher jetzt mit uns Ärzten zusammenarbeiten, dann verstehen sie, dass es in der Klinik viele spannende Fragen gibt, an denen sie arbeiten können. Für sie ist es auch eine große Motivation zu wissen, dass sie mit ihrer Forschung Menschen helfen können. Für die Ärzte ist es wiederum spannend, die Mechanismen hinter Krankheiten aufzuklären, die sie jeden Tag sehen. Forscher und Ärzte motivieren sich somit gegenseitig.

Sie sind nicht nur LiSyM-Programmdirektor, sondern stehen auch weiterhin regelmäßig am Krankenbett – denken Sie auch an Ruhestand? Oder fragt Ihre Familie manchmal danach?

Im Gegenteil, ich habe gerade sogar eine weitere Funktion übernommen. Ich bin jetzt auch Gastprofessor in London und Maastricht. Meine Familie unterstützt mich bei allen Projekten, weil sie weiß, dass es genau das ist, was ich machen möchte.

Außerdem finde ich, dass 65 definitiv kein Alter für einen Wissenschaftler ist, um in Rente zu gehen. Im Austausch mit jungen Wissenschaftlern kann ich meinen Beitrag leisten, da machen sich die Jahrzehnte Erfahrung bezahlt. Ich denke, ich kann ihnen ein guter Ratgeber sein. Solange ich das machen kann, genieße ich es.

Das Gespräch führten Melanie Bergs und Gesa Terstiege.

Kontakt:

Prof. Dr. Peter Jansen
Programmdirektor Forschungsnetz LiSyM
p.l.jansen@amc.uva.nl

www.Hepaconsult.com

Forschungsnetz Systemmedizin der Leber (LiSyM)

Anknüpfend an die bisher veröffentlichen Maßnahmen im Bereich der Systemmedizin fördert das Bundesforschungsministerium seit dem Januar 2016 das Forschungsnetz „Systemmedizin der Leber“ (Liver Systems Medicine - LiSyM) im Rahmen des übergeordneten Forschungs- und Förderkonzeptes „e:Med – Maßnahmen zur Etablierung der Systemmedizin“. Das Forschungsnetz vereint nationale Kompetenzen aus den Bereichen der anwendungsorientierten Grundlagenforschung, der klinischen Forschung und der Systembiologie für die Leberforschung. Insgesamt sind 27 Projekte beteiligt, darunter vier Nachwuchsgruppen. In vier Forschungsschwerpunkten wird der Verlauf krankheitsbedingter Veränderungen der Leberfunktion bis hin zum Organversagen auf den verschiedenen Organisationsebenen (Zelle-Gewebe-Organ-Organismus) untersucht. Ziel ist die Entwicklung von Multi-Skalen-Modellen, die eine Aufklärung der grundlegenden Schlüsselmechanismen bei Lebererkrankungen ermöglichen. Basierend auf aktuellen Fortschritten in der klinischen Leberforschung soll mit diesem Ansatz eine Einordnung der Patienten in Risikogruppen ermöglicht sowie optimierte Behandlungsverfahren entwickelt werden. Durch die Umsetzung der Maßnahme werden die Forschungsergebnisse aus dem von 2010-2015 geförderten Kompetenznetz „Die Virtuelle Leber“ für eine Anwendung in der Klinik weiterentwickelt. Das Bundesforschungsministerium stellt hierfür 20 Millionen Euro für die kommenden fünf Jahre zur Verfügung.

woran man sich erinnert

Die epigenetische Grundlage des Gedächtnisses aus systembiologischer Sicht

von Tonatiuh Pena Centeno, Ramon Vidal, Magali Hennion und Stefan Bonn

Eines der wesentlichen Merkmale, die den Menschen von anderen Lebewesen unterscheidet, sind seine hochentwickelten geistigen Fähigkeiten. Grundlegende Intelligenz, unser Gedächtnis und Gefühle sind für uns von so großer Bedeutung, dass man sagen könnte, sie definieren wer wir sind. Allerdings wissen wir noch nicht, wie all das funktioniert. Was das Gedächtnis betrifft, so ist noch nicht völlig klar, wie die Speicherung und Erhaltung von Erfahrungen geschieht.

Unter Wissenschaftlern herrscht jedoch Konsens darüber, dass das Gedächtnis mit der sogenannten *synaptischen Plastizität* im Zusammenhang steht, der Fähigkeit, bestehende neuronale Verbindungen zu verstärken und neue Verbindungen zu bilden. Neue Studien stützen die Hypothese, dass Veränderungen der Genaktivität in den Gehirnzellen, die über den genetischen Code hinausgehen, d. h. *epigenetische* Veränderungen, mit dieser Plastizität assoziiert sind und eine wichtige Rolle bei der Enkodierung des Gedächtnisses spielen (Abbildung 1). In diesem Artikel werden neue systembiologische Erkenntnisse über die mechanistische Rolle von epigenetischen Veränderungen bei Lern- und Gedächtnisprozessen besprochen.

Das Gedächtnis mit einem „Schweizer Taschenmesser“ untersuchen

Während klar ist, dass Gedächtnisprozesse durch die Verstärkung und Abschwächung neuronaler Verbindungen innerhalb von Zellnetzwerken gesteuert werden, weisen genetische und pharmakologische Untersuchungen darauf hin, dass epigenetische Faktoren bei Lern- und Gedächtnisprozessen eine Rolle spielen (Lopez-Atalaya & Barco 2014). Der Begriff *Epigenetik* beschreibt die Erforschung von Veränderungen des Phänotyps eines Organismus, die nicht auf Veränderungen der zugrunde liegenden DNA-Sequenz beruhen. Epigenetische Modifikationen, die zu phänotypischen Veränderungen führen, treten während der Lebenszeit eines Organismus regelmäßig und natürlicherweise auf und können durch Umweltfaktoren wie Alter, Krankheiten oder sogar Lebensstil ausgelöst werden; ferner können sie erblich sein oder nicht. Derzeit sind mindestens drei epigenetische Mechanismen bekannt: DNA-Methylierung, Histonmodifikationen und Gene-Silencing durch nicht-kodierende RNA. Während bei der Erforschung der Auswirkungen epigenetischer Veränderungen im Hinblick auf verschiedene biologische Prozesse, unter anderem Embryonalentwicklung, Altern und Erkrankungen, große Fortschritte erzielt wurden, so ist über ihre Rolle bei gedächtnisbezogenen Prozessen bisher nur wenig bekannt.

Dieser „Wissensmangel“ kann durch die Anwendung systembiologischer Ansätze, des „Schweizer Taschenmessers“ der Biologie, bei der Erforschung des Lernens und der Gedächtnisbildung erheblich reduziert werden. Mit modernen Hochdurchsatz-Methoden lassen sich genomweite DNA-Veränderungen und Veränderungen von Histonmodifikationen in bestimmten Hirnarealen und spezifischen Zelltypen zu verschiedenen Zeitpunkten vor und nach Lernvorgängen sichtbar machen (Halder *et al.*, 2016). Die Analyse von genomweiten Chromatin-Modifikationen und RNA-Expressionsdaten in Kombination mit veröffentlichten strukturierten Daten (z. B. Datenbanken) und unstrukturierten Daten (z. B. Forschungsartikel) kann dann dazu beitragen, besser zu verstehen, wie Erinnerungen gebildet und gespeichert werden.

Welche Erkenntnisse können also aus systembiologischen Ansätzen der Hirnforschung gewonnen werden?

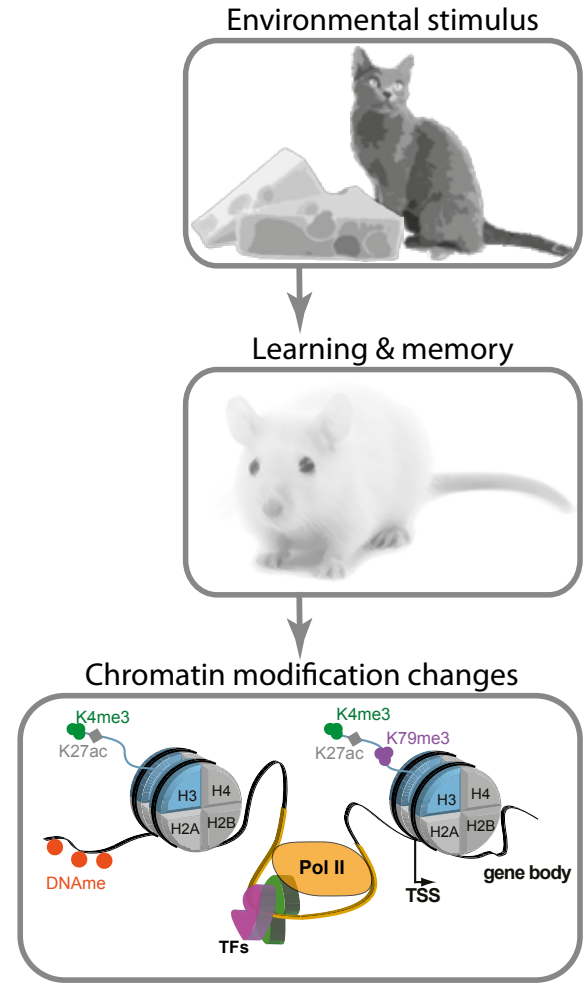


Abbildung 1: Umweltbedingte Stimuli führen zu einer Anpassung unseres Verhaltens. Am Ursprung dieser Anpassung ist die Fähigkeit sich auf neue umweltbedingte Situationen einzustellen und zu lernen. Die Entwicklung und Erhaltung des Gedächtnisses setzt Änderungen in den epigenetischen Modifikationen des Chromatins voraus (Grafik: Stefan Bonn).

Ein wenig Aufmerksamkeit bitte

Bereits seit Langem ist bekannt, dass Modifikationen von Histonen, Proteine um die sich die DNA im Zellkern windet, an bestimmten Resten eine wichtige Rolle bei der Regulation der Genexpression spielen (Abbildung 1). Manche davon wurden mit aktiven Genzuständen in Verbindung gebracht, andere mit repressiven Genzuständen und einige mit Lernprozessen. Welche Veränderungen von Histonmodifikationen finden während Lernprozessen statt und könnten sie ein molekulares Korrelat des Gedächtnisses sein? Überraschenderweise und im Gegensatz zu den meisten Entwicklungsdaten korrelieren genomweite Veränderungen von Histonmodifikationen nur schwach mit Veränderungen der Genexpression während des Lernens. Andererseits ist frühes Lernen bei einem Großteil der Gene von einer „globalen" Zunahme von Histonmodifikationen, die mit Aktivität in Verbindung stehen, und einer „globalen" Abnahme von Histonmodifikationen, die mit Inaktivität in Verbindung stehen, begleitet. Dieser Effekt ist von kurzer Dauer und tritt während der Gedächtnisbildung auf (1 Stunde nach dem Lernen), allerdings nicht während der Aufrechterhaltung des Gedächtnisses (4 Wochen nach dem Lernen). Obwohl diese Ergebnisse bedeuten, dass Histonmodifikationen bei der Aufrechterhaltung des Gedächtnisses keine große Rolle spielen, könnten die beobachteten globalen „aktivierenden" Veränderungen von Histonmodifikationen Genome in eine Art aufmerksamen oder vorbereiteten Zustand versetzen. Dieser zeitabhängige Wechsel könnte ein molekulares Korrelat von Aufmerksamkeit sein, da er Hirnareale möglicherweise auf zukünftige eingehende Stimuli „vorbereitet" und die Gedächtnisbildung stärkt.

Ein bleibender Eindruck

Während die beobachteten lernbezogenen Veränderungen von Histonmodifikationen möglicherweise Ihre Aufmerksamkeit erregt haben, könnten die Erkenntnisse zu Veränderungen der DNA-Methylierung bei der Gedächtnisbildung und -aufrechterhaltung einen bleibenden Eindruck bei Ihnen hinterlassen (Abbildung 2). Während Histonmodifikationen von relativ kurzer Dauer sind und mit der Gedächtnisbildung einhergehen, scheint die DNA-Methylierung wichtig dafür zu sein, um Erinnerungen für längere Zeit aufzubauen und zu bewahren. Langfristige Veränderungen der DNA-Methylierung in Neuronen sind daher eng mit der differentiellen Expression und dem Splicing von Zielgenen verbunden. Diese Zielgene sind an der Gestaltung der synaptischen Plastizität und der Verschaltung von Neuronen beteiligt und geben so weitere Hinweise auf die funktionelle Bedeutung der DNA-Methylierung für Lern- und Gedächtnisprozesse.

Arbeitsgruppe von Dr. Stefan Bonn im Januar 2016 (Foto Daniel Riester, DZNE-G).

Mehr als nur Neuronen und „Kitt“

Wenn man an das Gehirn und an Gedächtnisbildung denkt, dann denkt man automatisch an Neuronen. Und es stimmt ja auch, dass die Neuronen für einen Großteil der Schwerstarbeit zuständig sind, die das Gehirn leistet. Wie ist es jedoch mit der Rolle von anderen Zelltypen, insbesondere von Gliazell-Populationen? Sind sie wirklich nur der „Kitt“, der das Gehirn zusammenhält? Zumindest aus epigenetischer Sicht kann diese Frage klar mit Nein beantwortet werden, da auch in nicht-neuronalen Zellen in einem überraschend großen Ausmaß sowohl genspezifische als auch globale Veränderungen von Histonmodifikationen stattfinden. Insbesondere während der Gedächtnisbildung zeigen nicht-neuronale Zelltypen kurzfristige Veränderungen von Histonmodifikationen, was auf eine bisher noch nicht bekannte funktionelle Rolle dieser Zellen bei Lernprozessen hindeutet.

Quo Vadis?

Systembiologische Forschungsansätze werden in der Neurowissenschaft nicht häufig angewandt und einer der Gründe dafür könnte eine gesunde Skepsis gegenüber *„high input – low output“*-Technologien sein. Die neuere Forschung zeigt jedoch, dass die Systembiologie weitere Erkenntnisse liefern kann, indem sie einen tieferen Einblick in die Netzwerke der Genregulation, die dem Gedächtnis zugrunde liegen, ermöglicht. Dieses Verständnis der epigenetischen Grundlagen des Lernens und der Gedächtnisprozesse bei gesunden Menschen kann nun als Referenzpunkt

Abbildung 2:

Die Aufrechterhaltung des Gedächtnisses geht mit Änderungen der DNA-Methylierung (5mC) in kortikalen Hirnarealen einher. Wenn wir lernen, wird an spezifischen Promotor und Enhancer Regionen die DNA-Methylierung geändert. Manche dieser Änderungen bleiben für Monate bestehen und stimmen in Raum und Zeit mit der Position des Gedächtnisses überein (Grafik: Stefan Bonn).

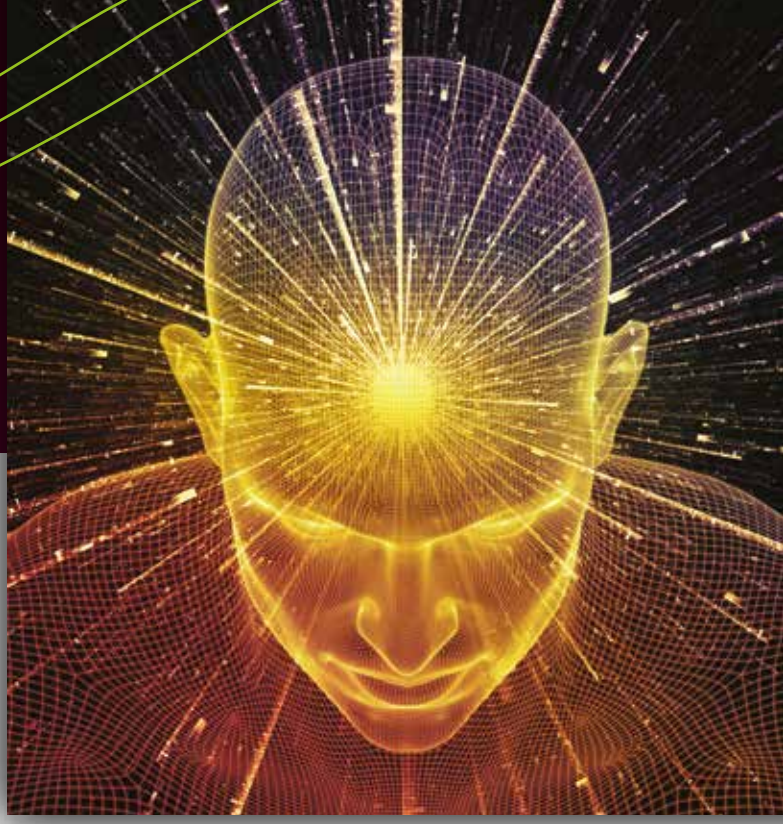


Foto: agsandrew -Fotolia.com

für den Vergleich mit Krankheitszuständen wie der Alzheimer-Erkrankung oder mit natürlichen Alterungsprozessen dienen. Interessant wäre es auch, die epigenetischen Veränderungen in gedächtnisbildenden Zellnetzwerken zu verstehen, indem man genetische Markierungstechniken mit Einzelzell-Sequenzierungen und Analysemethoden der Bioinformatik kombiniert.

Projektpartner:

Dr. Bettina Schmid
Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) München,
Munich Cluster for Systems Neurology (SyNergy), München, Deutschland.

Prof. Dr. Christian Haass
Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) München,
Biomedizinisches Zentrum, Ludwig-Maximilians-Universität München,
Munich Cluster for Systems Neurology (SyNergy), München, Deutschland.

Prof. Dr. Andre Fischer
Forschungsgruppe Epigenetik neurodegenerativer Erkrankungen
Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE)
Klinik für Psychiatrie und Psychotherapie, Universitätsmedizin Göttingen,
Göttingen, Deutschland.

Referenzen:

Lopez-Atalaya JP, Barco A. (2014) Can changes in histone acetylation contribute to memory formation? Trends Genet. Dec;30(12):529-39.

Halder R, Hennion M, Vidal R, Shomroni O, Rahman R, Rajput A, Pena Centeno T, van Bebber F, Capece V, Garcia Vizcaino J, Schuetz A-L, Burkhardt S, Benito E, Navarro Sala M, Bahari Javan S, Haass C, Schmid B, Fischer A, Bonn S. (2016) DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. Nat Neurosci. 19(1):102-10.

Kontakt:

Foto: S. Bonn

Dr. Stefan Bonn
stefan.bonn@dzne.de

Dr. Magali Hennion,
Dr. Tonatiuh Pena Centeno,
Dr. Ramon Vidal
Forschungsgruppe Computergestützte Analyse biologischer Netzwerke
Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE)
Göttingen

http://www.dzne.de/standorte/goettingen/forschergruppen/bonn.html

AptaBodies

DNA-Aptamere als Alternative für Antikörper beim Western Blotting

von Jasmin Dehnen und Frieda Anna Sorgenfrei

Artikel siehe auch Laborjournal Heft 12/2015, Seite 56 - 57

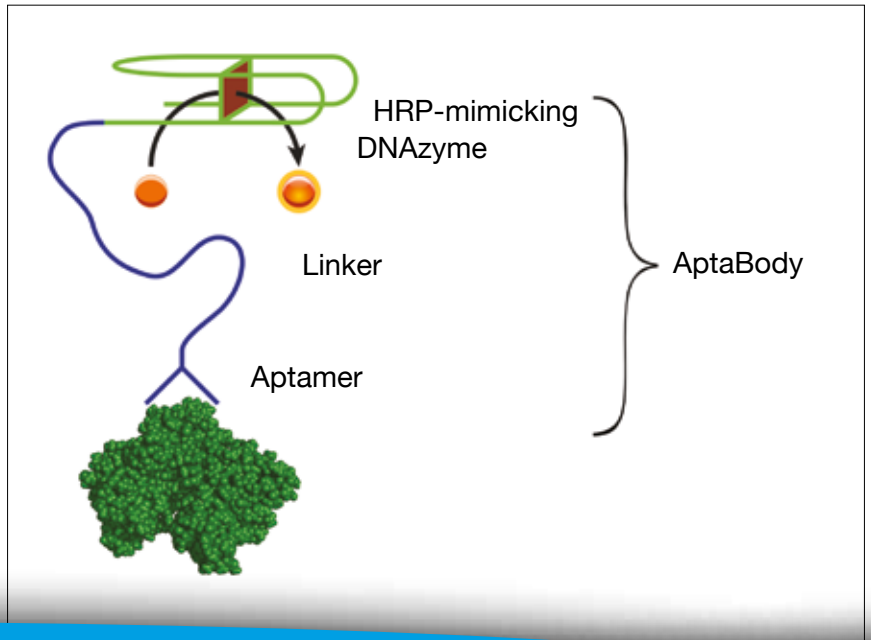
Was erhält man, wenn man ein Protein-bindendes Aptamer an ein DNAzym hängt, das die Reaktion der Meerrettich-Peroxidase imitiert? Einen Apta-Body für Western Blots!

Jeder Molekularbiologe kennt ihn, die meisten haben ihn schon mal um Rat gefragt und alle fluchen über ihn. Nein, gemeint ist nicht der Chef – die Rede ist vom Western Blot. Bei diesem trennt der Experimentator die Proteine in einem Polyacrylamid-Gel auf und transferiert sie anschließend auf eine Membran. Diese inkubiert er schließlich mit einem Protein-spezifischen Antikörper, der an das gewünschte Protein bindet. Meist ist für die Detektion noch ein zweiter Antikörper nötig, der an den Primärantikörper andockt.

Seit Harry Towbin 1979 den Western Blot als Methode zur Proteindetektion vorstellte bereitet uns dieser sowohl Sorge als auch Freude. Freude, weil das Protokoll einfach und die Anwendungen vielfältig sind. Sorge, weil er meist doch nicht so funktioniert, wie es im Lehrbuch steht. Über die Jahre verfeinerten verschiedene Gruppen das Western Blot-Protokoll mit dem Ziel, seltene Proteine zu detektieren oder die langwierigen Wasch- und Inkubationszeiten zu umgehen.

Ein kritischer Punkt des Western Blots, der Biowissenschaftlern nicht nur Kopfschmerzen bereitet, sondern ihnen auch das Geld aus der Tasche zieht, ist der Antikörper. Fast jedes Labor kennt Fälle, in denen Antikörper nicht tun, was sie sollen, und unspezifisch oder überhaupt nicht binden. Das frustriert den Forscher

Abbildung 1: AptaBody



Der AptaBody bindet über das Aptamer an ein Protein und katalysiert mit dem HRP-mimicking DNAzym die chemilumineszente Reaktion von Luminol und Wasserstoffperoxid (Foto: iGEM Team Heidelberg).

Die Heidelberger AptaBody-Crew mit Mentor Roland Eils (v.r) beim iGEM Finale in Boston (Foto: iGEM Team Heidelberg).

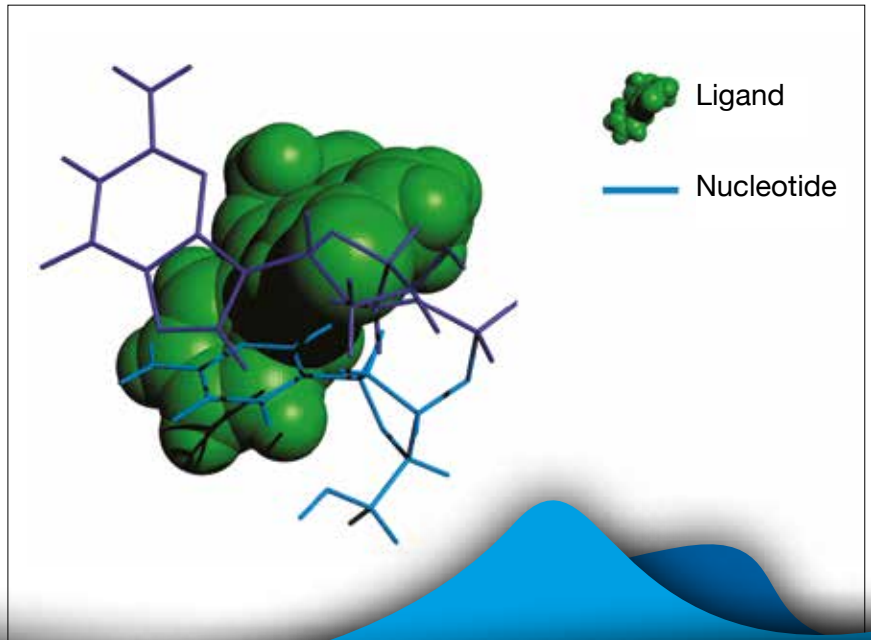
gleich zweifach: Er hat sich nicht nur experimentell im Kreis bewegt, sondern auch noch für viel Geld ein Reaktionsgefäß mit einer nutzlosen Proteinlösung erstanden, das zudem noch einen der hart umkämpften Plätze im Freezer blockiert.

Nach jedem Fehlversuch geht die Suche nach einem funktionierenden Antikörper weiter und die zähe Prozedur vom Antikörper bestellen bis zum Detektieren der Proteine auf der Membran beginnt von Neuem. So ist nicht nur schnell Ebbe in der Kasse, auch die drei Jahre Promotionszeit sind im Nu vorbei. Ein markierter primärer Antikörper sorgt vielleicht für einen besseren Schlaf, das böse Erwachen kommt aber spätestens, nachdem man die satte Rechnung für diesen gesehen hat. Erschwert wird das Ganze durch die nicht unerhebliche Zahl an Proteinen, für die noch kein passender Antikörper entwickelt wurde. Wünschenswert wäre deshalb eine günstige, schnelle und zuverlässige Alternative zu Antikörpern. Aber wie soll diese aussehen?

Exotisches DNAzym

Diese Frage stellten sich zehn Studenten der Universität Heidelberg, die mit ihrem Lösungsansatz am diesjährigen iGEM *(international Genetically Engineered Machine)*-Wettbewerb in Boston teilnahmen. Während der Projektfindungs-Phase tauchte die Gruppe tief in die Welt der Moleküle ein und konzentrierte

Abbildung 2: MAWS - Berechnung neuer Aptamere



Die Software MAWS rechnet anhand von Entropie Minimierung Nukleotid für Nukleotid ein neues Aptamer gegen einen Liganden aus (Foto: iGEM Team Heidelberg).

Frieda Anna Sorgenfrei (Foto: iGEM Team Heidelberg)

sich schließlich auf funktionelle RNA-Moleküle, etwa Ribozyme, die spezifische Reaktionen katalysieren. Das Heidelberger iGEM-Team musste jedoch schnell erkennen, dass die Arbeit mit RNA ihre Tücken hat, weil diese sehr leicht degradiert. Gut, dass Biologen schon vor einiger Zeit DNAzyme – die DNA-Pendants zu Ribozymen – entwickelt haben, die deutlich stabiler sind als klassische Ribozyme.

Im Laufe der Recherche gingen den Studenten die abstrusesten funktionellen Moleküle ins Netz. Der größte Fang war die 17-Nukleotide-lange, katalytisch aktive DNA, Meerrettich-Peroxidase (HRP) mimicking DNAzym, die Dipankar Sens Gruppe von der Simon Fraser University in Kanada 1998 entdeckte (Travascio *et al., Chemistry and Biology,* 9, 505–17). Die klassische HRP katalysiert die chemilumineszente Reaktion von Luminol und Wasserstoffperoxid, bei der blaues Licht entsteht. Das Enzym wird deshalb häufig für die Detektion von Western Blots kovalent an Sekundärantikörper geknüpft. Aber wie schafft es ein 17 Nukleotide langes DNA-Fragment, die gleiche Reaktion zu katalysieren wie HRP?

Das DNAzym bildet eine G-Quadruplex Sekundärstruktur, die das Binden von Hämin in dessen Zentrum ermöglicht. Das so komplexierte Hämin katalysiert letztendlich die gleiche Reaktion wie HRP. Warum also nicht das HRP-mimicking DNAzym als Signalgeber des geplanten Antikörperersatzes verwenden?

Blieb noch die Frage zu klären, wie der neuartige Antikörperersatz das anvisierte Protein erkennt. Hier fiel die Wahl des Heidelberger Teams auf eine weitere Klasse funktioneller DNA-Moleküle: Aptamere. Aptamere sind kurze Nukleinsäuren, die an nahezu beliebige Targets binden. Für den Einsatz beim Western Blot benötigte die Studentengruppe natürlich Protein-bindende Aptamere.

Einfaches AptaBody Protokoll

Hierzu koppelte sie das HRP-mimicking DNAzym an ein Aptamer, das His-Tags erkennt (siehe Abbildung 1). Das Resultat war ein kurzer DNA-Einzelstrang, der über das 5‘-Ende His-getaggte Proteine bindet und mit dem 3‘-Ende Luminol aktiviert. Mit diesem sogenannten AptaBody konnten die Studenten Proteine im Zelllysat von Escherichia coli detektieren. Das Protokoll für die Verwendung von AptaBodies bei Western Blots ist einfach und schnell: Zunächst kocht man den AptaBody, damit er beim anschließenden Abkühlungsprozess seine Sekundärstruktur einnimmt. Anschließend pipettiert man Hämin hinzu.

Der *International Genetically Engineered Machine* (iGEM) Wettbewerb

Seit über zehn Jahren wird der international führende iGEM *(International Genetically Engineered Machines)* Wettbewerb am renommierten Massachusetts Institute of Technology (MIT) in Boston ausgetragen. Es ist der weltweit größte Wettbewerb der Synthetischen Biologie und findet jährlich statt.

Für den Wettbewerb arbeiten Studenten- und Schülerteams den Sommer über an einem frei gewählten Forschungsprojekt und präsentieren anschließend ihre Ergebnisse während des sogenannten Giant Jamboree in Boston. Dort werden Medaillen und Hauptgewinne in über zehn Kategorien verliehen. Ziel des Wettbewerbes ist es, neue biologische Systeme zu entwickeln, die einen Lösungsansatz für alltägliche Probleme bieten oder einen Beitrag zur Grundlagenforschung leisten. Dafür erhalten die Teams eine Zusammenstellung von DNA-Sequenzen. Diese biologischen Module, sogenannte Biobricks, können in Modellorganismen zu neuen Systemen kombiniert werden. Darüber hinaus organisieren die Teilnehmer Veranstaltungen, um die Aufmerksamkeit der Öffentlichkeit auf das große Potential der Synthetischen Biologie zu lenken. Der iGEM Wettbewerb erfährt zunehmende Beliebtheit. 2015 nahmen 280 Teams aus Nord- und Südamerika, Europa, Asien und Australien teil.

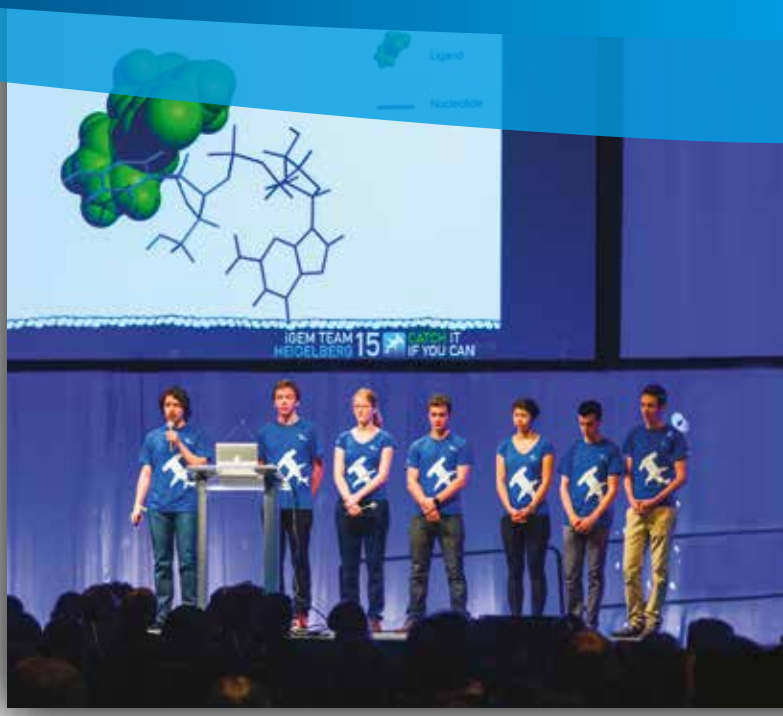
Damit ist die meiste Arbeit auch schon getan. Während der Inkubationszeit von AptaBody und geblotteter Membran, heißt es für den Biowissenschaftler Abwarten und Teetrinken. Das anschließende Protokoll ist für den eingefleischten Western Blotter nichts Neues. Nach einem kurzen Waschschritt, der ungebundene AptaBodies entfernen soll, wird eine Luminol/Wasserstoffperoxid-Lösung auf den Ansatz pipettiert. Der AptaBody funktioniert also wie ein markierter primärer Antikörper – ist aber um ein Vielfaches günstiger.

AptaBodies wären natürlich ziemlich eingeschränkt, wenn sie nur His-Tags erkennen würden. Wünschenswert wären AptaBodies, die auch an Ziel-Proteine ohne His-Tag binden. Die Voraussetzung hierfür sind Protein-spezifische Aptamere, die an das HRP-mimicking DNAzym fusioniert werden. Aptamere mit hoher Spezifität und Affinität für einen Protein-Liganden lassen sich routinemäßig mit dem SELEX-Verfahren selektionieren. Bei diesem inkubiert man eine zufällige DNA-Bibliothek mit dem Ziel-Protein und wäscht danach ungebundene Sequenzen ab. Alle anderen Sequenzen mutiert man, um die Bindung zu optimieren, und gibt sie wieder zu den Proteinen. Diese Schritte wiederholt man, bis man Aptamere mit einer maximalen Affinität zu den Zielproteinen erhält.

Aber mal ehrlich: das SELEX-Verfahren ist mühsam, verschlingt viel Zeit und führt nicht immer zum Ziel, da es stark vom Ausgangs-DNA-Pool abhängt. Gute Gründe für die Heidelberger Studenten nach einer Alternative zur traditionellen Methode zu suchen. Anstatt weiter an der Bench zu stehen, entwickelten sie eine Software, die neue Aptamere in silico generiert. Der hierzu verwendete Algorithmus basiert auf dem Prinzip der Entropie-Minimierung. Er berechnet auf Basis einer bekannten 3D-Struktur des Ziel-Proteins die optimalsten Aptamer-Kandidaten. Der vom Heidelberger iGEM-Team ausgesuchte Name für die entwickelte Software MAWS (*Making Aptamers without SELEX* siehe Abbildung 2) erinnert nicht zufällig an JAWS – den starken, weißen Hai in Steven Spielberg's erstem Blockbuster.

Nur noch Oligos bestellen

Funktionieren soll das Ganze in der Praxis folgendermaßen: Der von Antikörpern für Western Blots enttäuschte Biowissenschaftler gibt die Struktur des Zielproteins in MAWS ein. Die Software generiert verschiedene Aptamere, die spezifische Epitope des Proteins erkennen. Anschließend verknüpft der Forscher die erzeugten Aptamere mit dem HRP-mimicking DNAzym. Die hierfür benötigten Oligos lässt er synthetisieren und hat die einsatzfähigen AptaBodies am nächsten Tag auf der Bench.



Die AptaBody-Crew beim iGEM Finale in Boston
(Foto: Justin Knight, the iGEM Foundation)

Weitere Experimente sollen die Funktion der AptaBodies genauer beschreiben und sie zum weitläufig anwendbaren Werkzeug für die molekulare Biologie machen. Western Blots mit kurzen DNA-Fragmenten, die hundertmal günstiger sind als Antikörper – das klingt fast zu gut, um wahr zu sein und doch sind die Studenten auf dem besten Weg genau das zu erreichen. Nicht umsonst belegte die Heidelberger Crew beim iGEM-Wettbewerb in der Gesamtwertung den dritten Platz.

Kontakt:

Frieda Anna Sorgenfrei
Studentin der Universität Heidelberg,
Fach Molekulare Biotechnologie (M.Sc)
frieda.sorgenfrei@bioquant.uni-heidelberg.de

Jasmin Dehnen
Studentin der Universität Heidelberg,
Fach Biowissenschaften (B.Sc)
jasmin-dehnen@gmx.de

http://2015.igem.org/Team:Heidelberg und
http://2015.igem.org/Team:Heidelberg/Project/AB

Foto Dehnen: J. Dehnen; Foto Sorgenfrei: A. Sorgenfrei

events

Computational Genomics approaches to Precision Medicine

12. – 23. September 2015, Berlin

AUFGEPASST: EINE SUMMER SCHOOL ZUR COMPUTER-GENOMIK

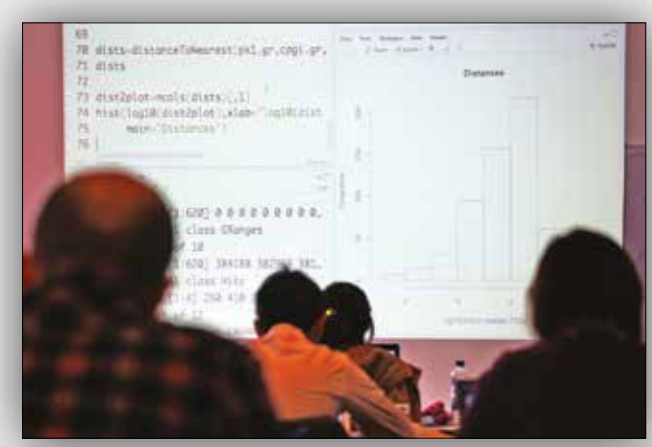
Sowohl die Biologie als auch die Medizin nutzen mehr und mehr Hochdurchsatzmethoden, die immer größere Datenmengen produzieren. Um diese zu verarbeiten und zu analysieren sind Fähigkeiten nötig, die eine eigene Disziplin darstellen. Für Wissenschaftler, die diese Experimente durchführen, ist es besonders wichtig die Analyse ihrer Daten auch zu verstehen.

Das Berlin Institut für Medizinische Systembiologie (BIMSB) am Max-Delbrück-Centrum für Molekulare Medizin (MDC) hat dazu ihre Summer School Reihe mit einem Kurs ergänzt, der sich genau mit diesem Thema beschäftigt. Hochrangige Wissenschaftler aus dem Memorial Sloan Kettering Cancer Institute, der New York University und der Universität Zürich sind gekommen um Ansätze zur Analyse von Sequenzierdaten der nächsten Generation, wie RNA-Seq und Chip-Seq, zu vermitteln.

„Hervorragende Einrichtung... extrem gut organisiert... exzellente Mitarbeiter... Hochinteressant, gerade auch für einen experimentellen Biologen“ – von den Teilnehmern kamen nur Superlative zur Beschreibung der „Computational Genomics“ Summer School.

Altuna Akalin, Leiter der Bioinformatik Technologie Plattform am BIMSB, hatte diesen Kurs organisiert, unterstützt durch die Stiftung Charité. Er hat sogar schon Pläne den erfolgreichen Kurs weiterzuführen und bietet die nächste Summer School „Computational Genomics approaches to Precision Medicine“ vom 12.-23. September, 2016 an – unterstützt durch das BMBF (unter Vorbehalt). Diesmal werden wieder Informatiker, Biologen und besonders Mediziner aufgefordert sich für den Kurs zu bewerben.

Weitere Informationen gibt es auf:
http://compgen2016.mdc-berlin.de.



Impressionen der letztjährigen „Computational Genomics“ Summer School (Fotos: Grietje Krabbe/MDC).

COMPUTATIONAL GENOMICS APPROACHES TO PRECISION MEDICINE

CONFIRMED LECTURERS

ALTUNA AKALIN
Max Delbrück Center,
Berlin Institute for Medical Systems Biology

UWE OHLER
Max Delbrück Center,
Berlin Institute for Medical Systems Biology

NIKOLAUS RAJEWSKY
Max Delbrück Center,
Berlin Institute for Medical Systems Biology

NICHOLAS D. SOCCI
Memorial Sloan Kettering Cancer Center,
New York, USA

MARK ROBINSON
University of Zurich, Switzerland

ROLAND SCHWARZ
University of Cambridge, UK

CHRIS E. MASON
Weill Cornell Medical College
New York, USA

DAVIDE RISSO
UC Berkeley, USA

12-23 SEP 2016 | Berlin Germany | Berlin Institute for Medical Systems Biology, Max Delbrück Center

Application Deadline 01 July 2016

COURSE MODULES

- Introduction to R & Bioconductor
- Statistics and Exploratory Data analysis
- Introduction to Next-gen sequencing
- RNA-seq analysis
- ChIP-seq analysis
- Variant calling and annotation
- Data integration and predictive modeling
- Metagenomics and human health
- Cancer classification based on HT-seq data

http://compgen2016.mdc-berlin.de

MDC MAX DELBRÜCK CENTER FOR MOLECULAR MEDICINE IN THE HELMHOLTZ ASSOCIATION
BIMSB THE BERLIN INSTITUTE FOR MEDICAL SYSTEMS BIOLOGY

Bundesministerium für Bildung und Forschung

Konferenzbericht

8th International Conference on Systems Biology of Human Disease – SBHD 2015
06. - 08. Juli 2015, Heidelberg

MIT HILFE DER SYSTEMBIOLOGIE KRANKHEITEN BESSER VERSTEHEN

von Cornelia Depner

Zum dritten Mal fand bei schönstem Sommerwetter vom 06. bis 08. Juli 2015 die internationale Konferenz „Systems Biology of Human Disease" (SBHD) am Deutschen Krebsforschungszentrum (DKFZ) in Heidelberg statt. Die jährlich stattfindende Konferenz wurde vor einigen Jahren von Prof. Peter Sorger von der renommierten Harvard Medical School in Boston ins Leben gerufen und hat sich seither zu einer deutsch-amerikanischen Veranstaltung entwickelt, die alternierend in Boston und Heidelberg stattfindet. Unterstützt wurde die Konferenz 2015 unter anderem von der Schweizer Systembiologie-Initiative SystemsX und von dem vom BMBF ins Leben gerufenen Systemmedizinnetzwerk e:Med. Die rund 200 Teilnehmer und Teilnehmerinnen stellten die aktuellen Ergebnisse ihrer Forschung in 43 Vorträgen und 81 Poster-Präsentationen vor, um neueste krankheitsrelevante Forschungsergebnisse aus der Systembiologie zu präsentieren und zu diskutieren. Das Organisationskomitee gab auch ausgewählten Nachwuchswissenschaftlern die Möglichkeit ihre Arbeit in einem 20-minütigen Vortrag zu präsentieren. In 5-minütigen Kurzvorträgen konnten Nachwuchswissenschaftler zudem auf ihre Poster aufmerksam machen.

Übergreifendes Thema der Konferenz war die Nutzung von mathematischen Methoden und Computermodellen zur Erfassung und Untersuchung von komplexen biologischen Systemen auf allen Ebenen, vom Genom bis hin zum gesamten Organismus. Dabei bot die Konferenz einen breiten Überblick über die Anwendung systembiologischer Forschung in der Medizin, von systematischer

SBHD 2015 Preisverleihung

v. l. n. r.: Andreas Raue von Merrimack Pharmaceuticals, Preisträger Karsten Rippe (DKFZ Heidelberg), Konferenzleiter Roland Eils (DKFZ/ Universität Heidelberg), Preisträger Kazuki Tainaka (Universität Tokio), Georg Draude von der Chroma Technology GmbH

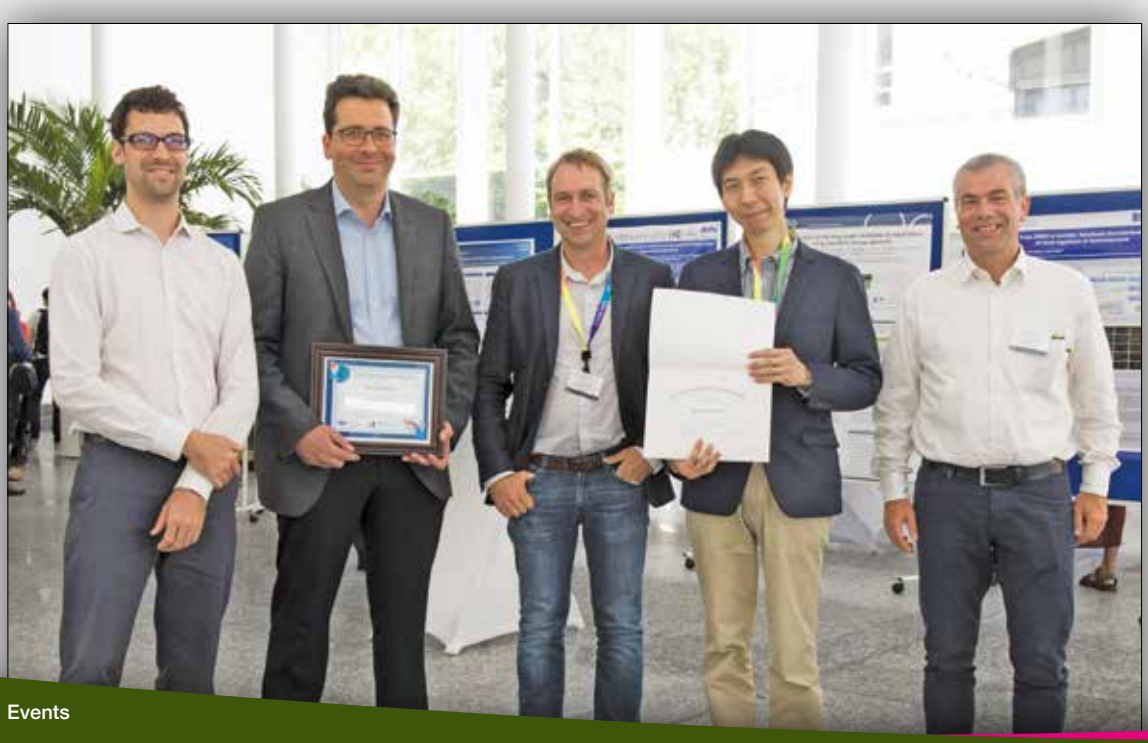
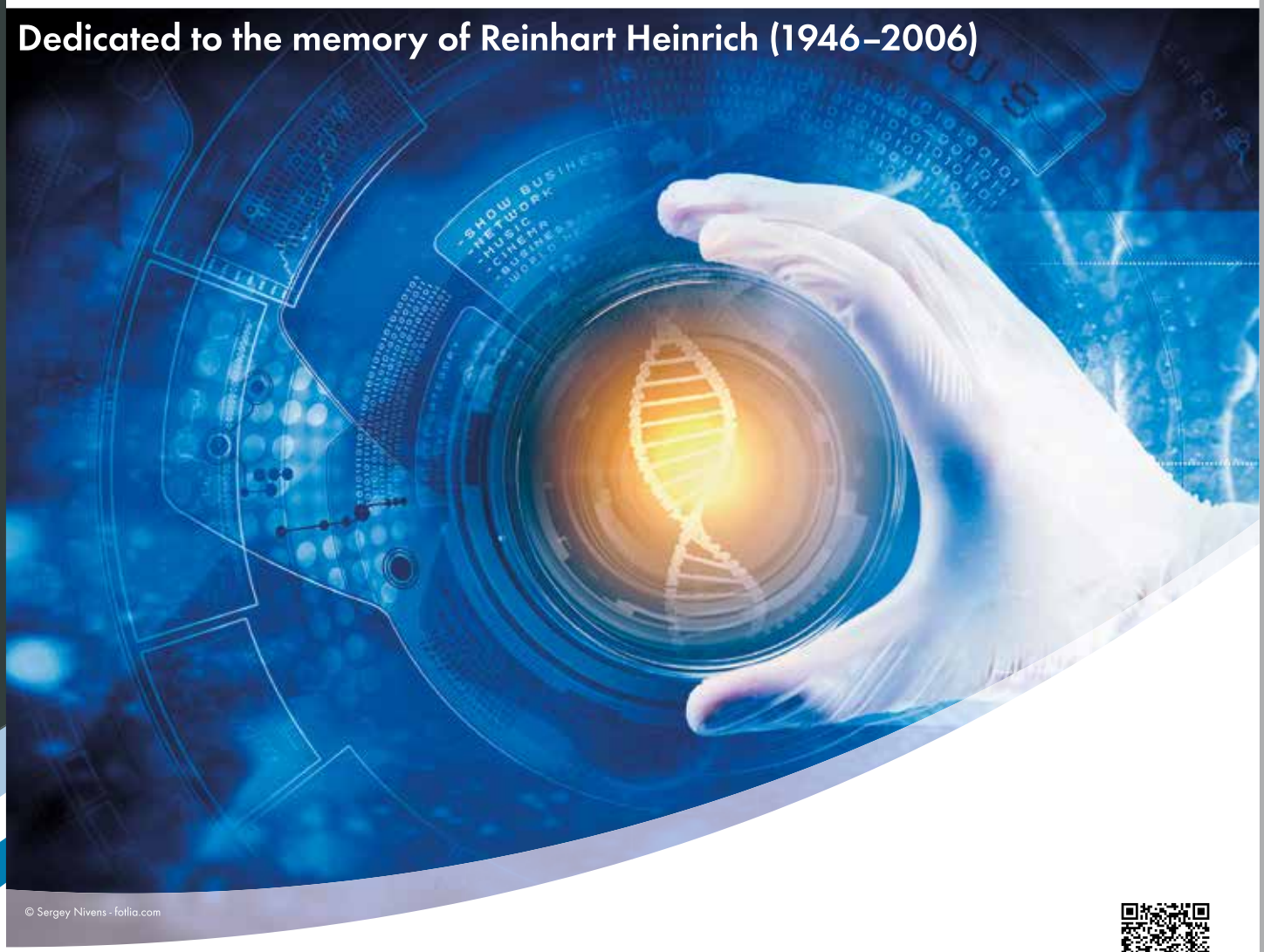


Foto: Yan de Andres

ISGSB 2016 JENA

INTERNATIONAL STUDY GROUP FOR SYSTEMS BIOLOGY

Dedicated to the memory of Reinhart Heinrich (1946–2006)



© Sergey Nivens - fotlia.com

04–07 OCTOBER 2016

Friedrich Schiller University of JENA, GERMANY

Abstract Deadline 31 May 2016

© FSU Jena

© FSU Jena

Main themes

1. Infection modelling
2. Regulatory interactions and signalling
3. Plant physiology and development
4. Biological thermodynamics
5. Optimality principles
6. Multiscale systems medicine
7. Metabolic pathway analysis
8. Towards whole-cell models

www.isgsb2016.de

Auditorium während der Vorträge (Foto: Redaktion systembiologie.de).

Posterpräsentationen im Foyer des Kommunikationszentrums am DKFZ (Foto: Redaktion systembiologie.de).

Pharmakologie, Computational Biology und Netzwerkrekonstruktion bis hin zu Proteomik, Einzelzell-Transkript- und Proteinanalyse sowie der mathematischen Modellierung von Therapieresistenzmechanismen.

Auch Themen wie die Datenverarbeitung und das Datenmanagement wurden diskutiert. Einen Ansatz mit großem Potential für die medizinische Forschung zeigte Professor Eytan Ruppin, von der University of Maryland, auf. Sein Team entwickelte eine Methode, mit der die immer größer werdenden Datenschätze, die bei der Analyse der Veränderungen von Tumorgenomen entstehen, genutzt werden können. Durch einen systematischen Vergleich der Daten von vielen Patienten können patientenspezifische und potentielle Angriffspunkte für Therapien vorhergesagt und überprüft werden.

Eine besondere Ehre wurden Privatdozent Dr. Karsten Rippe vom Deutschen Krebsforschungszentrum und Dr. Kazuki Tainaka von der Universität Tokio zuteil.

Karsten Rippe erhielt den von Merrimack Pharmaceuticals gesponserten „**CSB2 – Preis in Systembiologie**“ für seine interdisziplinären Arbeiten über die strukturelle Organisation unseres Erbguts und deren krankhafte Veränderungen in Krebszellen. Forschungsdaten aus Rippes Abteilung zeigten erstmals, wie Krebszellen durch fehlregulierte epigenetische Modifikationen bestimmte DNA-Reparaturenzyme dazu missbrauchen, Chromosomenenden wieder zu verlängern und dass sich diese Krebszellen dadurch unendlich teilen können. In gesunden Zellen verkürzen sich die Chromosomenenden, die sogenannten Telomere, bei jeder Zellteilung ein wenig. Wenn diese Telomere nicht mehr vorhanden sind, kann sich die Zelle nicht mehr teilen. Dieser Mechanismus ist in Krebszellen ausgeschaltet, da die Chromosomenenden durch DNA-Reparaturproteine stabilisiert und die fehlenden Chromosomenstücke wieder angefügt werden. Mit Hilfe eines systembiologischen Ansatzes und der Auswertung von Mikroskopie-Aufnahmen sowie Sequenzierungsdaten konnten 29 Proteine ermittelt werden, die an dem komplexen Prozess der alternativen Telomer-Verlängerung beteiligt sind. Diese Proteine sind nun Grundlage für die Suche nach neuen Therapieformen, um die Telomer-Verlängerung zu verhindern.

Kazuki Tainaka erhielt den „**Anne Heidenthal Prize for Fluorescence Research**“, gesponsert von der Chroma Technology Corp., für die Entwicklung der CUBIC Methode *(Clear, Unobstructed Brain Imaging Cocktails and Computational Analysis)*. Die CUBIC Methode ist ein neu entwickeltes Verfahren, das durch eine Kombination von chemischer Gewebe-Entfärbung und Nutzung der sogenannten Lichtscheiben-Fluoreszenzmikroskopie (LFSM) extrem detaillierte Aufnahmen vom Inneren einzelner Organe und sogar gesamter Organismen ermöglicht. Dabei wird ein Bestandteil des Hämoglobins, des eisenhaltigen Proteinkomplexes unserer roten Blutkörperchen, mit Hilfe von Amino-Alkoholen ausgespült, sodass das Licht bei der Mikroskopie nicht mehr absorbiert und das Gewebe somit besser sichtbar wird. Durch mehrfache Wiederholung dieses Auswaschprozesses können die Gewebe nahezu durchsichtig gemacht werden. Diese Methode ermöglicht ein neues Verständnis der dreidimensionalen Struktur von Organen und kann für weitere Studien auf Organ- und Zellebene im Kontext verschiedener Krankheiten genutzt werden.

Neben den Vorträgen und Auszeichnungen kam auch die soziale Komponente bei der Konferenz nicht zu kurz. Die Teilnehmer konnten auf einem Neckarschiff ein wunderbares Barbecue und Musik genießen. Highlight der Neckarfahrt war der musikalische Beitrag von Professor Uri Alon vom Weizman Institut in Israel, der seine Erfahrungen in der Forschung in einem Song verarbeitete und zur Freude aller Gäste mit der Gitarre vortrug.

Die 9. SBHD-Konferenz findet vom 14.- 16. Juni 2016 am Broad Institute in Boston, USA statt.
Mehr Informationen auf Seite 85 in diesem Heft und unter www.sbhd2016.org.

INTERNATIONAL CONFERENCE ON THE SYSTEMS BIOLOGY OF HUMAN DISEASE

JUNE 14–16, 2016

REGISTRATION AND DETAILS AT:

sbhd2016.org

Early registration deadline: MARCH 12, 2016
Regular registration deadline: MAY 14, 2016

Poster abstract submission deadline: MAY 14, 2016
All poster abstract submissions will be reviewed and the top two submissions will be invited to give full-length talks. Several more posters will be selected for "lightning talks" designed to preview the poster presentation.

BROAD INSTITUTE
415 MAIN STREET
CAMBRIDGE, MA 02142

CONFIRMED SPEAKERS:

Leonidas Alexopoulos: National Technical University of Athens
Bernd Bodenmiller: University of Zurich
Markus Covert: Stanford University
Hana El-Samad: University of California, San Francisco
Suzanne Gaudet: Dana-Farber Cancer Institute
Alexander Hoffmann: University of California, Los Angeles
Andrew Hopkins: University of Dundee
Galit Lahav: Harvard Medical School
Kathyrn Miller-Jensen: Yale University
Julio Saez-Rodriguez: RWTH-Aachen University
Xia Yang: University of California, Los Angeles

NIGMS NATIONAL CENTERS FOR SYSTEMS BIOLOGY PRESENTATIONS:

Center for RNA Systems Biology: University of California, Berkeley
Center for Complex Biological Systems: University of California, Irvine
San Diego Center for Systems Biology: University of California, San Diego
Center for Systems and Synthetic Biology: University of California, San Francisco
Laboratory of Systems Pharmacology: Harvard Medical School
Systems Biology Center New York: Icahn School of Medicine at Mount Sinai
Center for Systems Biology: Institute for Systems Biology, Seattle
MIT Synthetic Biology Center: Massachusetts Institute of Technology
Virtual Physiological Rat: University of Michigan
New Mexico Spatiotemporal Modeling Center: University of New Mexico
Center for Systems Biology of Retrotransposition: New York University School of Medicine
META Center for Systems Biology: University of Oregon
Stanford Center for Systems Biology: Stanford University

SPONSORED BY:

HHS Harvard Program in Therapeutic Science
HARVARD MEDICAL SCHOOL
dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION
BioQuant
HELMHOLTZ ASSOCIATION
SystemsX.ch The Swiss Initiative in Systems Biology
SBC@MIT

Neues aus der EU-Förderung: ERACoSysMed und ERASysAPP

Europäische Forschungs- und Entwicklungsförderung zur Implementierung systembiologischer und systemmedizinischer Herangehensweisen

Das Förderinstrument „ERA-NET“ ist eine Initiative, die die EU-Kommission ins Leben gerufen hat. Ziel ist es, durch gemeinsame Aktivitäten die Forschung in wichtigen Themenbereichen europaweit zu koordinieren. Sie wird von den nationalen Förderorganisationen der einzelnen Länder getragen.

ERACoSysMed

von Sylvia Krobitsch

Das ERA-NET „ERACoSysMed – Förderung transnationaler Forschungsprojekte in der Systemmedizin zur Implementierung systembiologischer Ansätze in klinischer Forschung und medizinischer Praxis“ ist Teil des Rahmenprogramms Horizont 2020 der Europäischen Kommission. Es wird von der EU-Kommission mit 4,88 Mio. EUR über einen Zeitraum von fünf Jahren gefördert.

ERACoSysMed ist ein 2015 gestartetes ERA-NET-Konsortium, das 15 nationale Förderorganisationen und Ministerien aus 13 Ländern gebildet haben. Ziel ist es, die europäische Systemmedizin weiter zu etablieren. Hierbei spielt die von dem europäischen Konsortium CASyM zur Entwicklung einer gesamteuropäischen Implementierungsstrategie für die Systemmedizin gestaltete Roadmap eine wichtige Rolle. Innerhalb der fünfjährigen Laufzeit von ERACoSysMed sind insgesamt drei Förderbekanntmachungen geplant, wobei die erste Bekanntmachung 2015 unter Mitbeteiligung von EU-Geldern (ERA-NET Cofund) durchgeführt wurde. Eine zweite und dritte Bekanntmachung ist für die Jahre 2016/2017 und 2018/2019 vorgesehen.

Gefördert werden transnationale Verbünde, an denen Arbeitsgruppen aus mindestens drei unterschiedlichen Ländern beteiligt sind. Die an der Bekanntmachung beteiligten Länder fördern hierbei die eigenen Arbeitsgruppen innerhalb eines Verbundes. In der ersten Runde liegt der Schwerpunkt gezielt auf sogenannten Demonstrator-Projekten. Diese Forschungsprojekte sollen den sozioökonomischen Nutzen des systemmedizinischen Ansatzes anhand konkreter klinischer Fragestellungen belegen (Stichwort 4P-Medizin – personalisiert, präventiv, prädiktiv und partizipierend). Die Begutachtung der eingereichten Projektantragsskizzen erfolgte mittels eines zweistufigen Verfahrens. Insgesamt wurden neun Verbundprojekte mit zwölf deutschen Beteiligungen ausgewählt und zur Förderung empfohlen. Die beantragte Fördersumme auf europäischer Ebene beläuft sich auf etwa 12,7 Mio. EUR. Die ersten Projekte mit einer Laufzeit von drei Jahren werden voraussichtlich im zweiten Quartal 2016 starten.

Kontakt:

Dr. Sylvia Krobitsch
Projektträger Jülich
Forschungszentrum Jülich
s.krobitsch@fz-juelich.de

www.eracosysmed.eu
www.ptj.de/eracosysmed

Foto: S. Krobitsch

ERASys*APP*

von K. Zsuzsanna Nagy

Das ERA-NET zum Thema „Angewandte Systembiologieforschung im europäischen Forschungsraum“ mit dem Projektkürzel ERASys*APP* wurde von der Europäischen Kommission mit 2 Mio. EUR im Zeitraum vom 01.01.2013 bis 31.12.2015 gefördert.

Im ERASys*APP*-Konsortium arbeiteten 16 nationale Förderorganisationen aus 13 Ländern zusammen. Das gemeinsame Ziel war, die Ergebnisse aus der systembiologischen Forschung in Europa verstärkt in der Biotechnologie, in der Industrie und im Bereich Gesundheit anzuwenden. Zu diesem Zweck haben die ERASys*APP*-Partner in den Jahren 2013 und 2014 gemeinsam zwei länderübergreifende Bekanntmachungen zur angewandten systembiologischen Forschung veröffentlicht. Aus allen eingereichten Projektskizzen wurden 12 internationale Vorhaben ausgewählt, welche über drei Jahre mit insgesamt etwa 16 Mio. EUR gefördert werden. Deutsche Wissenschaftler beteiligen sich an allen Verbundvorhaben; insgesamt sind 21 deutsche Forschungsgruppen Teil von ERASys*APP*-Projekten. Sie werden mit 6,7 Mio. EUR vom Bundesforschungsministerium unterstützt.

Die Bandbreite der geförderten Projekte ist groß: So werden Vorhaben zur Optimierung verschiedener Produktionsprozesse in Mikroorganismen oder Pflanzen unterstützt, aber auch Projekte zur Metallgewinnung mittels biotechnologischer Verfahren, zur Entwicklung antiviraler Wirkstoffe oder Vorhaben zur Leberkrebsforschung erhalten Fördermittel. Zusätzlich hat ERASys*APP* Networking-Workshops veranstaltet, um Universitäten und Industrie stärker zu vernetzen und die Anwendungsorientierung systembiologischer Forschung weiterhin zu fördern. Um einen Rahmen zur effektiven Zusammenarbeit und Vernetzung von Systembiologen in Europa zu gewährleisten, aktualisierte ERASys*APP* zudem die strategische Forschungsagenda aus dem Jahr 2008 für dieses Forschungsfeld.

Ein weiterer Punkt war die Nachwuchsförderung: Auch hier hat ERASys*APP* einen wichtigen Beitrag zur Aus- und Weiterbildung der nächsten Systembiologen-Generation geliefert. So wurden unter anderem zahlreiche Workshops und Kurse angeboten und ein Bildungsportal zur Systembiologie aufgebaut. Dieses hält Links zu Graduiertenprogrammen, webbasierte Lernmaterialien und eine Plattform zum Austausch von Unterrichtsmaterial bereit (https://www.erasysapp.eu/training-and-exchange/mobility-of-researchers). Zudem hat ERASys*APP* – gemeinsam mit der ESFRI-Maßnahme ISBE (Infrastruktur für Systembiologie in Europa) – den Grundstein für das europäische Daten- und Modellmanagement Projekt FAIRDOM (http://fair-dom.org/) gelegt. Mit Hilfe einer zentralen Plattform, Tools und verschiedenen Diensten unterstützt FAIRDOM systembiologische Projekte bei der Verwaltung und Archivierung von Forschungsdaten. Ein Kurzfilm über die Vorteile von strukturiertem Datenmanagement steht auf youtube unter https://www.youtube.com/watch?v=PWutnWBfUSw zur Verfügung. Nach dem Auslaufen der eigenständigen ERASys*APP*-Fördermaßnahme wurde im Horizont 2020-Programm NMBT für die Jahre 2016/2017 eine COFUND-Maßnahme für Biotechnologie ausgeschrieben, auf die sich ERASys*APP* gemeinsam mit ERASynBio und ERA-IB bewirbt.

Kontakt:

Dr. K. Zsuzsanna Nagy
Projektträger Jülich
Forschungszentrum Jülich
k.nagy@fz-juelich.de

www.erasysapp.eu
www.ptj.de

Foto: K. Zs. Nagy

Willkommen auf systembiologie.de!

Wenn Sie mehr über die Systembiologie erfahren möchten, besuchen Sie unsere Homepage.

Das erwartet Sie:

- Spannende Geschichten aus dem Forschungsalltag – Erfahren Sie mehr über aktuelle Projekte
- Systembiologen im Portrait – Lernen Sie die Gesichter hinter der Forschung kennen
- Umfassende Veranstaltungsübersicht zur Systembiologie – Verpassen Sie keinen wichtigen Termin
- Informationen über aktuelle Fördermaßnahmen – Bleiben Sie stets auf dem Laufenden
- Aktive Mitgestaltung – Schlagen Sie uns Ihr Thema vor

Wir freuen uns auf Ihren Besuch auf unserer Homepage!

systembiologie.de scholae – Sonderausgabe für Schüler erschienen

Die Systembiologie ist ein noch junges Wissenschaftsgebiet und den meisten Schülern heute noch kein Begriff. Um schon in der Schule das Interesse an interdisziplinärer Forschung zu wecken und gezielt künftige Generationen von Studierenden anzusprechen, wurde die Sonderausgabe *systembiologie.de scholae* konzipiert. Sie soll frühzeitig einen Einblick in das vielfältige Forschungsfeld der Systembiologie geben und stellt aktuelle Forschungsergebnisse in einer direkt im Unterricht der Oberstufe einsetzbaren Form dar.

Hierfür haben Mitarbeiter des Deutschen Krebsforschungszentrums und des Projektträgers Jülich gemeinsam mit dem Team des Gläsernen Labors aus Berlin ausgewählte Forschungsbeiträge aus den Magazinen von *systembiologie.de* für die Schule aufbereitet, so dass Schüler, aber nicht zuletzt auch ihre Lehrer, systembiologische Forschungsansätze aus verschiedenen Blickwinkeln betrachten und verstehen können. Neben einer allgemeinen Einführung in die Systembiologie wurden dabei Inhalte aus früheren Ausgaben des Magazins ausgesucht, die auch im Biologie-Lehrplan eine Rolle spielen, wie etwa Krebs, Stammzellen und Epigenetik. Außerdem gibt die Sonderausgabe mit zwei Originalartikeln aus dem Magazin *systembiologie.de* und einem Interview auch Einblicke in die unmittelbare Forschungsarbeit der Systembiologen. Eine Liste von deutschen Universitäten, an denen Systembiologie studiert werden kann, rundet das Heft ab.

systembiologie.de scholae hat das Ziel, systembiologische Erkenntnisse zur Verfügung zu stellen, Diskussionen unter Schülern zu fördern und das Bewusstsein für wissenschaftliche und medizinische Probleme zu erhöhen sowie Lehrer bei der Vermittlung von systembiologischem Wissen zu unterstützen. Jedes Kapitel enthält entsprechende Arbeitsaufträge für die Schüler. Mögliche Lösungsvorschläge und zum Teil weiterführende Hinweise befinden sich in dem beigefügten Lehrerheft *systembiologie.de scholae/Didaktik*.

impressum

Die Sonderausgabe *systembiologie.de scholae* wurde aus Mitteln des Bundesministeriums für Bildung und Forschung und der Helmholtz-Gemeinschaft finanziert und wird kostenlos abgegeben. Einzelexemplare und Klassensätze können auf Anfrage unter folgender Kontaktadresse beim Gläsernen Labor Berlin-Buch bezogen werden: **d.giese@bbb-berlin.de**

Online abrufbar ist das Heft unter:
www.systembiologie.de/de/magazin

Titelseite: systembiologie.de scholae – Sonderausgabe für Schüler
© LANGEundPFLANZ, Titelbild: shotstudio – Fotolia.com

systembiologie.de

Das Magazin für systembiologische Forschung in Deutschland – Ausgabe 10, April 2016

systembiologie.de ist ein halbjährlich erscheinendes Magazin mit Informationen aus der deutschen Systembiologieforschung.

ISSN 2191-2505

Herausgeber:

systembiologie.de wird herausgegeben von der Helmholtz Gemeinschaft, Querschnittsthema Systembiologie und Synthetische Biologie, dem Virtual Liver Network, dem Projektträger DLR und dem Projektträger Jülich.

Redaktion:

Chefredakteur: Prof. Dr. Roland Eils (DKFZ/Universität Heidelberg)
Redaktionelle Koordination: Dr. Cornelia Depner (DKFZ Heidelberg)
Redaktion:
Johannes Bausch (Virtual Liver Network, Universität Freiburg), Melanie Bergs (PtJ), Dr. Cornelia Depner (DKFZ Heidelberg), Dr. Jan Eufinger (DKFZ Heidelberg), Dr. Marco Leuer (PT-DLR), Dr. Angela Mauer-Oberthür (BioQuant, Universität Heidelberg), Dr. Yvonne Pfeiffenschneider (PtJ), Dr. Julia Ritzerfeld (DKFZ Heidelberg) und Dr. Gesa Terstiege (PtJ).

Anschrift:
Redaktion systembiologie.de
c/o Deutsches Krebsforschungszentrum (DKFZ)
Abteilung Theoretische Bioinformatik - B080
Im Neuenheimer Feld 580, D-69120 Heidelberg

Der Inhalt von namentlich gekennzeichneten Artikeln liegt in der Verantwortung der jeweiligen Autoren. Wenn nicht anders genannt, liegen die Bildrechte der in den Artikeln abgedruckten Bilder und Abbildungen bei den Autoren der Artikel. Die Redaktion trägt keinerlei weitergehende Verantwortung für die Inhalte der von den Autoren in ihren Artikeln zitierten URLs.

Gestalterische Konzeption und Umsetzung:
LANGEundPFLANZ Werbeagentur GmbH, Speyer (www.LPsp.de)
Druck:
Werbedruck GmbH Horst Schreckhase, Spangenberg (www.schreckhase.de)

PEFC zertifiziert
Dieses Produkt stammt aus nachhaltig bewirtschafteten Wäldern und kontrollierten Quellen
www.pefc.de

Aboservice:
Das Magazin wird aus Mitteln der Helmholtz-Gemeinschaft und des Bundesministeriums für Bildung und Forschung (BMBF) gefördert. Diese Veröffentlichung ist Teil der Öffentlichkeitsarbeit der unter Herausgeber genannten Initiativen. Sie wird kostenlos abgegeben und ist nicht zum Verkauf bestimmt.

Wenn Sie das Magazin abonnieren möchten, füllen Sie bitte das Formular auf www.systembiologie.de aus oder wenden sich an:

Redaktion systembiologie.de, c/o Abteilung Theoretische Bioinformatik B080
Deutsches Krebsforschungszentrum (DKFZ)
Im Neuenheimer Feld 580, D-69120 Heidelberg
abo@systembiologie.de

wir über uns

die systembiologie.de-Redaktion stellt sich vor

systembiologie.de möchte die Erfolge der deutschen Systembiologie auf anschauliche Weise einem breiten Publikum zugänglich machen. Erstellt wird das zweimal jährlich auf Deutsch und einmal jährlich auf Englisch erscheinende Magazin gemeinsam durch die Helmholtz Gemeinschaft, Querschnittsthema Systembiologie und Synthetische Biologie, dem Virtual Liver Network, dem Projektträger Jülich und dem DLR Projektträger. Finanziert wird das Magazin aus Mitteln der Helmholtz-Gemeinschaft und des Bundesministeriums für Bildung und Forschung (BMBF).

Die Redaktionsmitglieder von systembiologie.de:

v.l.n.r. stehend: Roland Eils (DKFZ/Universität Heidelberg), Yvonne Pfeiffenschneider (PtJ), Johannes Bausch (Virtual Liver Network), Angela Mauer-Oberthür (BioQuant, Universität Heidelberg), Kai Ludwig (LANGEundPFLANZ, Speyer), Cornelia Depner (DKFZ Heidelberg), Jan Eufinger (DKFZ Heidelberg).
v.l.n.r. sitzend: Gesa Terstiege (PtJ), Melanie Bergs (PtJ), Julia Ritzerfeld (DKFZ Heidelberg), Marco Leuer (PT-DLR).

Foto: Tobias Schwerdt / DKFZ

kontakt

Helmholtz Gemeinschaft, Querschnittsthema Systembiologie und Synthetische Biologie
Koordination: Prof. Dr. Roland Eils
Wissenschaftliches Projektmanagement:
Dr. Cornelia Depner, Dr. Jan Eufinger, Dr. Julia Ritzerfeld
c/o Deutsches Krebsforschungszentrum (DKFZ) Heidelberg
Abteilung Theoretische Bioinformatik – B080
Im Neuenheimer Feld 580; D-69120 Heidelberg
Email: j.eufinger@dkfz.de, c.depner@dkfz.de, j.ritzerfeld@dkfz.de
www.helmholtz.de/systemsbiology und www.helmholtz.de/syntheticbiology

Virtual Liver Network / LiSyM – Liver Systems Medicine
Programmdirektor: Dr. Adriano Henney / Prof. Dr. Peter Jansen
Wissenschaftliches Projektmanagement: Johannes Bausch
Universität Freiburg; Physikalisches Institut
Hermann-Herder-Str. 3; D-79104 Freiburg
Email: johannes.bausch@virtual-liver.de
www.virtual-liver.de

BioQuant – Universität Heidelberg
Direktorium: Prof. Dr. Roland Eils, Prof. Dr. Hans-Georg Kräusslich, Prof. Dr. Robert B. Russell
Geschäftsleitung: Dr. Angela Mauer-Oberthür
Im Neuenheimer Feld 267; D-69120 Heidelberg
Email: angela.oberthuer@bioquant.uni-heidelberg.de
www.bioquant.uni-heidelberg.de

Projektträger Jülich
Forschungszentrum Jülich GmbH
Lebenswissenschaften, Gesundheit, Fachhochschulen (LGF)
Ansprechpartner:
Dr. Yvonne Pfeiffenschneider, Dr. Gesa Terstiege, Melanie Bergs
Molekulare Lebenswissenschaften (LGF-2)
D-52425 Jülich
Email: y.pfeiffenschneider@fz-juelich.de, g.terstiege@fz-juelich.de, m.bergs@fz-juelich.de
www.ptj.de

DLR Projektträger
Gesundheitsforschung (OE20)
Ansprechpartner:
Dr. Marco Leuer, Ursula Porwol
Heinrich-Konen-Str. 1; D-53227 Bonn
Email: marco.leuer@dlr.de, ursula.porwol@dlr.de
www.dlr-pt.de

EMBL 2016

Conferences

3 - 6 APR | EMBO | EMBL Symposium
Tumour Microenvironment and Signalling
A. Berns, P.P. Pandolfi, B. Pauly | EMBL Heidelberg, Germany

20 - 22 APR | EMBL Conference
The Epitranscriptome
M. Frye, D. O'Carroll | EMBL Heidelberg, Germany

8 - 11 MAY | EMBO | EMBL Symposium
New Model Systems for Linking Evolution and Ecology
D. Tautz, D. Weigel | EMBL Heidelberg, Germany

18 - 20 MAY | EMBL Conference
BioMalPar XII: Biology and Pathology of the Malaria Parasite
G. Christophides, F. Osier, A. Scholzen, T. Voss
EMBL Heidelberg, Germany

29 MAY - 1 JUN | EMBO | EMBL Symposium
Microtubules: From Atoms to Complex Systems
M. Dogterom, C. Janke, A. Musacchio, M. Steinmetz
EMBL Heidelberg, Germany

3 - 5 JUN | EMBL Conference
Hematopoietic Stem Cells: From the Embryo to the Aging Organism
K. Kissa, C. Lancrin, C. Lo Celso, C. Robin
EMBL Heidelberg, Germany

12 - 15 JUN | EMBL Conference
CTLS 2016 - Core Technologies for Life Science
P. England, R. Pepperkok, S. Shorte
EMBL Heidelberg, Germany

26 - 29 JUN | EMBO | EMBL Symposium
Innate Immunity in Host-Pathogen Interactions
Z. Chen, W-D. Hardt, N. Pariente, F. Randow
EMBL Heidelberg, Germany

5 - 7 JUL | EMBL Conference
Lifelong Learning in the Biomedical Sciences
M. Hardman, C. Janko, C. Johnson
EMBL Heidelberg, Germany

24 - 26 JUL | EMBL Conference
Microfluidics 2016
C. Merten, S. Quake | EMBL Heidelberg, Germany

27 - 30 AUG | EMBL Conference
Transcription and Chromatin
D. Duboule, E. Furlong, A. Shilatifard, M. Timmers
EMBL Heidelberg, Germany

31 AUG - 3 SEP | EMBO Conference
Chemical Biology 2016
M. Köhn, J. Overington, C. Schultz | EMBL Heidelberg, Germany

7 - 10 SEP | EMBO | EMBL Symposium
Actin in Action: From Molecules to Cellular Functions
B. Baum, J. Faix, P. Lenart, D. Mullins, F. Nedelec, C. Sykes
EMBL Heidelberg, Germany

14 - 17 SEP | EMBL–Wellcome Genome Campus Conference
Proteomics in Cell Biology and Disease Mechanisms
A.-C. Gavin, A. Lamond, M. Mann | EMBL Heidelberg, Germany

25 - 27 SEP | EMBL–Wellcome Genome Campus Conference
Big Data in Biology and Health
E. Birney, B. Grossman, J. Korbel, C. Relton
EMBL Heidelberg, Germany

5 - 8 OCT | EMBO | EMBL Symposium
The Complex Life of mRNA
A. Ephrussi, N. Sonenberg, J. Steitz, D. Tollervey
EMBL Heidelberg, Germany

12 - 15 OCT | EMBO | EMBL Symposium
Organoids: Modelling Organ Development and Disease in 3D Culture
M. Bissell, J. Knoblich, E. Schnapp
EMBL Heidelberg, Germany

19 - 23 OCT | EMBO Conference
Experimental Approaches to Evolution and Ecology Using Yeast and Other Model Systems
J. Berman, M. Dunham, J. Leu, L. Steinmetz
EMBL Heidelberg, Germany

12 - 15 NOV | EMBO Conference
From Functional Genomics to Systems Biology
E. Furlong, F.C.P. Holstege, N. Rajewsky, M. Walhout
EMBL Heidelberg, Germany

20 - 23 NOV | EMBO Conference
Molecular Machines: Integrative Structural and Molecular Biology
J. Briggs, T. Carlomagno, G. Kleywegt, D. Panne, D. Svergun
EMBL Heidelberg, Germany

Courses

4 - 7 APR, 3 - 6 OCT | EMBL-EBI Course
Introduction to Next Generation Sequencing
T. Hancocks, J. Randall, M. Rossello | EMBL-EBI Hinxton, UK

17 - 23 APR | EMBL Course
High-Accuracy CLEM: Applications at Room Temperature and *in cryo*
R. Mellwig, M. Schorb | EMBL Heidelberg, Germany

18 - 22 APR | EMBL-EBI Course
Advanced RNA-Seq and ChiP-Seq Data Analysis
G. Rustici | EMBL-EBI Hinxton, UK

20 - 21 APR | EMBL Course
Transgenic Animals
J. Dreyer-Lamm, Y. Petersen, S. Stobrawa
EMBL Heidelberg, Germany

24 APR - 1 MAY | EMBO Practical Course
***In vivo* Plant Imaging**
M. Heisler, A. Maizel, R. Simon | EMBL Heidelberg, Germany

9 - 13 MAY, 11 - 15 JUL, 21 - 25 NOV | EMBL Courses
NGS: Enrichment Based Targeted Resequencing
V. Benes, J. Dreyer-Lamm, A. Heim
EMBL Heidelberg, Germany

9 - 13 MAY | EMBL-EBI Course
Networks and Pathways
T. Hancocks, S. Orchard | EMBL-EBI Hinxton, UK

21 - 26 MAY | EMBO Practical Course
Structural Characterisation of Macromolecular Complexes
T. Forsyth, D. Hart, M. S. López, M. Marcia, C. Petosa, D. De Sanctis | EMBL Grenoble, France

23 - 24 MAY, 4 - 5 JUL, 28 - 29 NOV | EMBL Courses
NGS: Whole Genome Sequencing Library Preparation
V. Benes, J. Dreyer-Lamm, A. Heim
EMBL Heidelberg, Germany

23 - 25 MAY | EMBL Course
Computational Aspects of High-Throughput Screening
J. Dreyer-Lamm, T. Hancocks, A. Hersey, J. Lewis, U. Uhrig
EMBL Heidelberg, Germany

23 - 27 MAY | EMBL-EBI Course
Bioinformatics of Plants and Plant Pathogens
P. Kersey | EMBL-EBI Hinxton, UK

29 MAY - 3 JUN | EMBO Practical Course
Non-Neuronal Optogenetics
H. Janovjak, S. De Renzis, J. Toettcher
EMBL Heidelberg, Germany

6 - 10 JUN | EMBL-EBI Course
Bioinformatics for Principal Investigators
S. Morgan | EMBL-EBI Hinxton, UK

5 - 11 JUN | EMBL Course
Bioimage Data Analysis
K. Miura, P. Paul-Gilloteaux, S.Tosi
EMBL Heidelberg, Germany

13 - 17 JUN | EMBL-EBI-Wellcome Genome Campus Course
Summer School in Bioinformatics
T. Hancocks, S. Morgan | EMBL-EBI Hinxton, UK

19 - 23 JUN | EMBO Practical Course
Computational Biology: Genomes to Systems
P. Bork, F. Ciccarelli, J. Korbel, R. Krause
EMBL Heidelberg, Germany

20 - 24 JUN | EMBL Course
Fundamentals of Widefield and Confocal Microscopy and Imaging
F. Eich, J. Marquardt, S. Terjung | EMBL Heidelberg, Germany

27 JUN - 1 JUL | EMBL-EBI Course
Cancer Genomics
G. Rustici | EMBL-EBI Hinxton, UK

28 JUN - 1 JUL, 4 - 7 OCT | EMBL Courses
Whole Transcriptome Data Analysis
V. Benes, R. Calogero | EMBL Heidelberg, Germany

3 - 8 JUL | EMBL Course
Advanced Fluorescence Imaging Techniques
F. Eich , R. Pepperkok, S.Terjung
EMBL Heidelberg, Germany

3 - 8 JUL | EMBL-EBI-Wellcome Genome Campus Course
***In silico* Systems Biology**
L. Emery | EMBL-EBI Hinxton, UK

11 - 15 JUL | EMBL Course
Quantitative Proteomics
J. Krijgsvield, M. Savitski | EMBL Heidelberg, Germany

28 AUG - 5 SEP | EMBO Practical Course
Cryo-Electron Microscopy and 3D Image Processing
J. Briggs, B. Boettcher, L. Passmore, C. Sachse, H. Stahlberg
EMBL Heidelberg, Germany

29 AUG - 2 SEP | EMBL Course
Chromatin Signatures During Differentiation
J. Dreyer-Lamm, P. Grandi, K.-M. Noh
EMBL Heidelberg, Germany

12 - 14 SEP | EMBL-EBI Course
Metagenomics Bioinformatics
H. Denise, L. Emery, A. Mitchell | EMBL-EBI Hinxton, UK

12 - 20 SEP | EMBO Practical Course
Protein Expression, Purification and Characterization
C. Loew, R. Meijers, A. Parret | EMBL Hamburg, Germany

19 - 23 SEP | EMBL-EBI Course
Structural Bioinformatics
T. Hancocks, G. Kleywegt, C. Orengo | EMBL-EBI Hinxton, UK

19 - 24 SEP | EMBL Course
Extracellular Vesicles: from Biology to Biomedical Applications
J. Dreyer-Lamm, A. Hendrix, E. Nolte-'t Hoen
EMBL Heidelberg, Germany

17 - 24 OCT | EMBO Practical Course
Solution Scattering from Biological Macromolecules
A. Kikhney, D. Svergun | EMBL Hamburg, Germany

17 - 23 OCT | EMBO Practical Course
High-Throughput Microscopy for Systems Biology
J. Ellenberg, D.W. Gerlich, B. Neumann, R. Pepperkok
EMBL Heidelberg, Germany

7 - 11 NOV | EMBL-EBI-Wellcome Genome Campus Course
Resources for Computational Drug Discovery
T. Hancocks | EMBL-EBI Hinxton, UK

9 - 10 NOV | EMBL Course
Microinjection into Adherent Cells
J. Dreyer-Lamm, S. Stobrawa | EMBL Heidelberg, Germany

28 NOV - 2 DEC | EMBL-EBI Course
Biological Interpretation of Next Generation Sequencing
G. Rustici | EMBL-EBI, Hinxton UK

4 - 9 DEC | EMBL-EBI-Wellcome Genome Campus Course
Proteomics Bioinformatics
L. Emery | EMBL-EBI, Hinxton UK

7 - 11 DEC | EMBL Course
Microbial Communities: Modelling Meets Experiments
R. Mahadevan, K. Patil, K. Sasaki
EMBL Heidelberg, Germany

For full event listing please visit our website
www.embl.org/events
@emblevents

We would like to thank the members of the EMBL ATC Corporate Partnership Programme:
Founder Partners: Leica Microsystems, Olympus
Corporate Partners: BD, Boehringer Ingelheim, GSK, Illumina, Thermo Fisher Scientific
Associate Partners: Eppendorf, Merck, Nikon, Sanofi

EMBL