



GEN-AU Projekt: Bioinformatik Integrationsnetzwerk

GEN-AU Project: Bioinformatics Integration Network

Das BM:BWK hat im Jahr 2001 das österreichische Genomforschungsprogramm GEN-AU ins Leben gerufen. Über neun Jahre (gegliedert in drei Phasen) werden Projekte mit einem Gesamtvolumen von 100 Millionen Euro gefördert. Die Forschungsarbeiten im Rahmen von GEN-AU zielen auf die Aufklärung der Genom-Struktur und auf die Bestimmung der Funktion einzelner Gene ab. Die bei Genomforschungsprojekten anfallenden Daten sind aber sehr umfangreich, enthalten viel wertvolle biologische Information und sind extrem schwer zu interpretieren. Nur unter Zuhilfenahme der Methoden und Konzepte der Informatik und Statistik ist die Analyse solcher Daten und die Beantwortung biologischer Fragen möglich. Die Erfassung, Verarbeitung und Interpretation dieser Daten wird durch die relativ junge Wissenschaftsdisziplin „Bioinformatik“ abgedeckt, die an der Schnittstelle zwischen den molekularen Biowissenschaften und den Informationswissenschaften angesiedelt ist.

Das Institut für Genomik und Bioinformatik (IGB) der TU Graz hat sich im Jahre 2002 mit dem interuniversitären Projekt „Bioinformatik Integrationsnetzwerk“ (BIN I) beworben und die Förderzusage für drei Jahre im Umfang von 1,7 Mio. Euro erhalten. Kürzlich wurde auch der Folgeantrag für das Projekt BIN II für die Jahre 2006-2008 mit einem Volumen von 2,5 Mio. Euro akzeptiert. Als Partner sind das Forschungsinstitut für Molekulare Pathologie (IMP) in Wien, das Institut für Theoretische Chemie der Universität Wien, die Abteilung Strukturbiologie des Instituts für Chemie der Universität Graz und das Tiroler Krebsforschungsinstitut beteiligt.

Das Forschungsprojekt BIN I hatte sich zum Ziel gesetzt, das bioinformatische Knowhow in Österreich zu bündeln und bioinformatische Lösungen zu entwickeln und anzuwenden, die auf die Fragestellungen und Anforderungen der österreichischen – und internationalen – Genom- und Proteomforscher zugeschnitten sind. Spezielles Augenmerk wurde dabei auf die Gebiete

- Entwicklung und Verwendung von Methoden zur Datenbankintegration
- kombinierte Analyse von Genomsequenzen, Genexpressionsdaten und Proteinstrukturdaten
- Untersuchung der RNA Sekundärstrukturen
- Analyse von tertiären Proteinstrukturen
- Funktionelle Vorhersage von Proteineigenschaften aufgrund der Sequenz gelegt.

Die Werkzeuge, die entwickelt wurden, sollten entweder in Form herunterladbarer Programme oder in Form von Webservices für die wissenschaftliche Gemeinde zur Verfügung stehen.

Ein weiterer Schwerpunkt des Projekts war die Vernetzung und Bildung von Partnerschaften mit der Wirtschaft. In diesem Zusammenhang wurde im April 2003 von SUN Microsystems das „SUN Center of Excellence for Bioinformatics“, als erstes seiner Art in Österreich, eröffnet. Mit der Auswahl des Instituts für Genomik und Bioinformatik und des IMP in Wien zeichnet SUN die Forschungsarbeit der beiden Institute in diesem Bereich aus, wobei erstmals die Bioinformatik als neue wissenschaftliche Disziplin unterstützt wird.

Eine wesentliche Voraussetzung für die Erreichung der oben genannten Ziele war der Aufbau einer entsprechenden IT Infrastruktur an den Standorten IMP Wien und IGB in Graz. Am IMP steht nun ein Computercluster mit 140 Prozessoren der Opteron-Klasse (mit 64bit Verarbeitungsbreite und 2,2 GHz Taktfrequenz), etwa 500 GB Hauptspeicher und etwa 7 Terabyte Plattenspeicher zur Verfügung. An der TU Graz wurde mit personeller und räumlicher Unterstützung durch den Zentralen Informatikdienst (ZID) Infrastruktur bestehend aus 5,5 Terabyte SAN Speicher, einem Computing Cluster mit 50 XEON

CPUs mit 100 GB Hauptspeicher, sowie Produktions-, Datenbank- und Entwicklungsservern basierend auf 64 Bit Prozessorarchitektur installiert.

Bei der Entwicklung der Methoden und Applikationen werden modernste Softwaretechnologien (UML, multi-tier Umgebungen wie die Java 2 Enterprise Architecture „J2EE“, Codegenerierung auf Basis der „Model Driven Architecture“) eingesetzt, gepaart mit der schnellen Prototypentwicklung unter Matlab, R und Perl. Durch den Einsatz von J2EE und die Auslagerung von Berechnungen auf die Computing Cluster ist die Skalierbarkeit der Anwendungen bei erhöhter Last gewährleistet.

Von den vielfältigen Forschungsergebnissen, die im Laufe von BIN I entstanden sind, können folgende hervorgehoben werden:

- Mit Hilfe der in-silico Vorhersage von Proteinfunktionen konnte eine bislang noch nicht beschriebene Lipase gefunden werden. Dabei wurden Hidden Markov Modelle und genomübergreifender Vergleich von Proteinsequenzen eingesetzt [5].
- Ein neuartiger Algorithmus zum Vergleich von konservierten RNA Sekundärstrukturen erlaubte die Identifizierung von über tausend sogenannter „non-coding“ RNAs, die eine bisher noch unbekannt Funktion der Regulation der Vorgänge in den Zellen spielen. Einige der identifizierten ncRNAs wurden bereits in vitro nachgewiesen [4].
- Mit dem „Microarray Analysis and Retrieval System“ (MARS) wurde eine Applikation entwickelt, die den gesamten Workflow von Microarray Experimenten abdeckt. Die anfallenden Daten werden lückenlos durch das System verwaltet. Die nachfolgende Auswertung wurde gegenüber bisherigen Methoden wesentlich vereinfacht und ausgebaut [1].
- Französische Wissenschaftler am INSERM in Paris konnten in Zusammenarbeit mit Bioinformatikern am IGB einen stichhaltigen Nachweis für den Zusammenhang zwischen Immunsystemfunktionen und der Ausbreitung von Krebs erbringen (Abb. 1). Zu diesem Zweck wurden alle relevanten klinischen Daten von rund 1000 Dickdarmkrebs-Patienten sowie an die 400 Gewebeproben, die innerhalb eines Zeitraumes von 15 Jahren entnommen und untersucht wurden, in einer Datenbank verwaltet und mit Hilfe statistischer Methoden zueinander in Beziehung gesetzt [2].

Weitere Ergebnisse des BIN I Projekts wurden in insgesamt 46 begutachteten Artikeln veröffentlicht, davon 16 mit Autorschaft des Instituts für Genomik und Bioinformatik.

Ein weiterer wichtiges Ziel des Projektes war die Aus- und Weiterbildung von Wissenschaftlern in der Bioinformatik. Dies wurde mit der Durchführung von Diplomarbeiten und Dissertationen und der Bereitstellung von Arbeitsplätzen für Gastwissenschaftler erreicht. Über das GEN-AU Mobility Programm konnten Mitarbeiter des BIN Projekts bis zu einem Jahr im Ausland zum Kennenlernen neuer Methoden verbringen.

Aufbauend auf den Ergebnissen von BIN I wird im Folgeprojekt BIN II, das Anfang dieses Jahres gestartet wurde, an der Entwicklung und Anwendung weiterer bioinformatischer Lösungen gearbeitet. Neu entwickelte experimentelle Methoden und Erweiterungen bestehender Methoden erfordern die Anpassung der vorhandenen Algorithmen und das Verfolgen völlig neuer Ansätze in der Auswertung und Visualisierung. Als Partner neu hinzugekommen sind das Center for Integrative Bioinformatics in Wien, das Institut für Molekulare Strukturbiologie der Universität Wien und die private Universität für medizi-

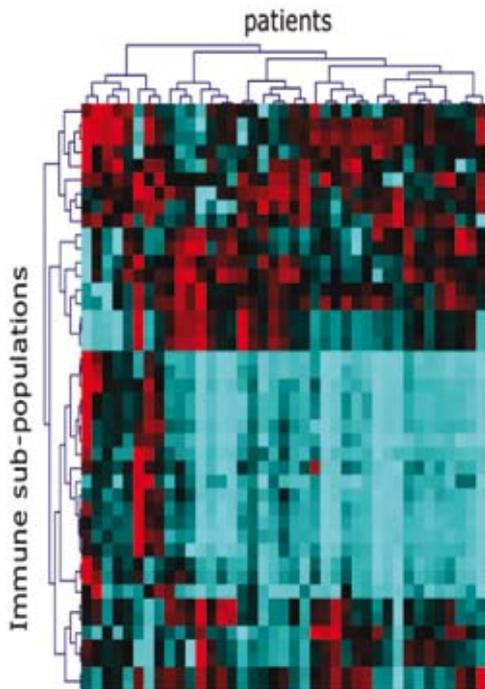


Abb. 1: Hierarchisches Clustering nach Immunsystemgruppen von Dickdarmkrebsbiopsien. Es sind klar vier große Gruppen mit unterschiedlichem Immunmarkern erkennbar.

nische Informatik in Hall/Tirol. Folgende Aktivitäten sind geplant:

- Gemeinsame Analyse von heterogenen Daten wie z.B. Geneexpressionsdaten, Daten aus Literatursuche und Proteininteraktionsdaten
- Erweiterung der Vorhersage von Proteinfunktion aufgrund der Sequenz
- Analyse von RNA Struktur und Kinetik
- Vergleich und Klassifikation von 3D Proteinstrukturen
- Analyse der Evolution genomischer Sequenzen
- Verwaltung, Analyse und "Mining" von Massenspektroskopiedaten
- Ausbau und Wartung der bioinformatischen Serviceleistungen
- Ausbildung und Weiterbildung von hoch-qualifiziertem Personal durch die Betreuung von DissertantInnen, Entwicklung und Abhaltung von Kursen, sowie Betreuung von postgradualen WissenschaftlerInnen. Damit sollte langfristig auch die Abwanderung von heimischen Talenten verhindert werden und im Gegenzug Österreich als Forschungsstandort für ausländische Wissenschaftler attraktiv gemacht werden (brain gain statt brain drain). Um diesem Anspruch gerecht zu werden wurde ein internationales PhD Programm ausgeschrieben (<http://genome.tugraz.at/binphd.php>), für das sich bereits Studenten aus der ganzen Welt beworben haben.

Weitere Informationen finden Sie auf unserer Homepage: <http://genome.tugraz.at> und auf der Homepage des GEN-AU Programms: <http://www.gen-au.at>.

Weiterführende Literatur (Auszug):

- [1] Maurer M, Molidor R, Sturn A, Hartler J, Hackl H, Stocker G, Prokesch A, Scheideler M, and Trajanoski Z. MARS: microarray analysis, retrieval, and storage system. *BMC Bioinformatics* 6(1):101, 2005
- [2] Pages F, Berger A, Camus F, Sanchez-Cabo F, Costes A, Molidor

R, Mlecnik B, Kirilovsky A, Nilsson M, Damotte D, Meatchi T, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH, and Galon J. Effector memory T cells, early metastasis, and survival in colorectal cancer. *New England Journal of Medicine* 353(25):2654-2666, 2005

- [3] Vogl C, Sanchez-Cabo F, Stocker G, Hubbard SJ, Wolkenhauer O, and Trajanoski Z. A fully Bayesian model to cluster gene expression profiles. *Bioinformatics* 21(Suppl 2):ii130-ii136, 2005
- [4] Washietl S, Hofacker IL, Lukasser M, Hüttenhofer A, and Stadler PF. Genome-wide mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in human. *Nature Biotechnology* 23(11):1383-1390, 2005
- [5] Zimmermann R, Strauss JG, Haemmerle G, Schoiswohl G, Birner-Gruenberger R, Riederer M, Lass A, Neuberger G, Eisenhaber F, Hermetter A, and Zechner R. Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. *Science* 306(5700):1383-1386, 2004

GEN-AU Project: Bioinformatics Integration Network

The Institute for Genomics and Bioinformatics at Graz University of Technology coordinates the project „Bioinformatics Integration Network“ (BIN I) and its follow-up project BIN II, which are funded in the context of the Austrian Genome Research Programme GEN AU. Together with partners from the Research Institute for Molecular Pathology, the University of Vienna, the University of Graz and the Tyrolean Cancer Research Institute BIN I has successfully addressed the following goals:

- To provide an environment for building bioinformatics capabilities in Austria, through establishing bioinformatic services, research, networking, training for both researchers and industry, and support for commercialization.
- To create an integrative research programme, using an Open Source platform of freely available source code, by establishing connections across multiple information resources—such as data on genomics, proteins, and clinical medicine - as well as across multiple disciplines, including mathematics, statistics, physics, computer science and life sciences.
- To establish bioinformatics infrastructure and support genomic initiatives at academic institutions in Austria.
- To develop and apply computational tools for combined analyses of genome sequences, gene expression data, and 3D structural data.

In the course of BIN I 46 papers in peer-reviewed journals have been published, 16 of them with authorship of the Institute for Genomics and Bioinformatics.

Based on the results of BIN I the follow-up project BIN II will expand and maintain the bioinformatic services, will develop new methods to analyze biological data and will train highly-qualified personnel in the field of bioinformatics. With additional partners from the Center for Integrative Bioinformatics and the private University for Medical Informatics in Tyrol, research will cover the following areas:

- Combined analysis of heterogeneous data including gene expression, literature and protein interaction data
 - Prediction of protein function based on its sequence.
 - RNA structure and kinetics analysis
 - Comparison and classification of 3D protein structures
 - Analysis of genomic sequence evolution
 - Management, analysis, and mining of mass spectroscopy data
- Additionally an international PhD programme will be initiated to attract talented students to Austria as a location for research