



Mario Albrecht, Ines Hopper-Pfister

# Innovative Bioinformatik für die Lebenswissenschaften

## *Innovative Bioinformatics for the Life Sciences*

Von der Neurowissenschaft über Gesundheitsforschung bis hin zur Bioinformatik: BioTechMed-Graz agiert an der Schnittstelle von Mensch, Medizin und Technik. Die erfolgreiche Kooperation zwischen den drei Grazer Universitäten TU Graz, Karl-Franzens-Universität und Medizinischer Universität setzt neue Impulse am Wissenschaftsstandort Graz. Dank der interuniversitären Forschungsplattform konnte auch die Bioinformatik an der TU Graz weiter gestärkt werden.

*From neuroscience and health research to bioinformatics, BioTechMed-Graz is at the interface of people, medicine and technology. The successful collaboration between three Universities in Graz, i.e. Graz University of Technology, the University of Graz and the Medical University of Graz is creating new momentum for the research location Graz. As part of the inter-university research platform, it is now possible to strengthen bioinformatics research at Graz University of Technology further.*





**Abbildung 1:** BioTechMed: drei Grazer Universitäten auf Erfolgskurs. TU Graz-Vizektor für Forschung Horst Bischof, BioTechMed-Professorin Christine Moissl-Eichinger, Uni Graz-Rektorin Christa Neuper, BioTechMed-Professor Mario Albrecht sowie MedUni Graz-Rektor Josef Smolle (v. l. n. r.)

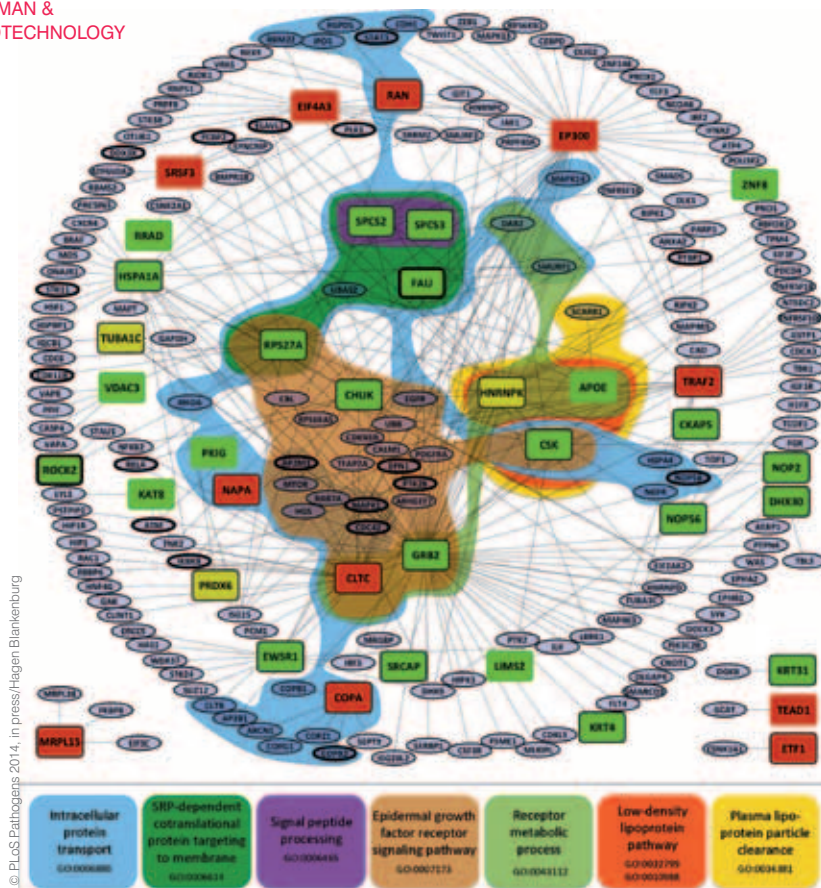
**Figure 1:** BioTechMed: three Graz universities set on a course for success. Graz University of Technology Vice Rector of Research Horst Bischof, BioTechMed Prof. Christine Moissl-Eichinger, Uni Graz Rector Christa Neuper, BioTechMed Prof. Mario Albrecht and MedUni Graz Rector Josef Smolle (from left to right.)

„Innovation durch Kooperation“ lautet wohl das Motto der drei Grazer Universitäten, die sich im Jahr 2012 zum Forschungsverbund BioTechMed-Graz zusammengeschlossen haben. Im Rahmen von BioTechMed-Graz bündeln die drei Big Player ihre bereits vorhandenen Kompetenzen in den vier großen gemeinsamen Forschungsthemen „Molekulare Biomedizin“, „Neurowissenschaften“, „Pharmazeutische und medizinische Technologie“ sowie „Quantitative Biomedizin und Modellierung“. Gemeinsame Forschung für die Gesundheit ist das große Ziel, die durch die optimale Nutzung der Infrastruktur auch eine verstärkte internationale Sichtbarkeit mit sich bringt.

BioTechMed-Graz dient als Plattform, um Wissenschaftlerinnen und Wissenschaftler aus dem biomedizinischen Bereich zusammenzubringen. „Im Rahmen dieser Initiative haben wir auch mehr Möglichkeiten, Spitzenforscherinnen und -forscher nach Graz zu holen“, wie Horst Bischof, Vizektor für Forschung an der TU Graz, unterstreicht. Beispielhaft dafür ist die gemeinsame Ausschreibung und Besetzung von vier neuen Professuren im Rahmen von BioTechMed. Eine dieser vier BioTechMed-Professuren wurde dem Fachbereich Bioinformatik gewidmet. Die Professur wird drei Jahre lang vom Land Steiermark gefördert und ist seit März dieses Jahres an der TU Graz angesiedelt. >

“Innovation through cooperation” is the motto of the three Universities in Graz, which united in the “Forschungsverbund BioTechMed-Graz” in the year 2012. Through BioTechMed-Graz, the three big players bundle the already existing competences into four larger joint research areas “Molecular Biomedicine,” “Neurosciences,” “Pharmaceutical and Medical Technology,” as well as “Quantitative Biomedicine and Modeling.” Combined research activities for health are the big goal, with heightened international visibility through optimal use of the available shared infrastructure.

BioTechMed serves as the platform that allows researchers in the biomedical area to connect. “In the framework of this initiative, we have more possibilities to recruit top scientists to Graz,” says Horst Bischof, Vice Rector of Research of Graz University of Technology. An example for this is the joint recruitment of four new professors via BioTechMed to Graz. One of these four new BioTechMed professorships is dedicated to bioinformatics. This appointment, which has been in effect since March 2014, will be funded for three years by the State of Styria. >



**Abbildung 2 :**  
**Netzwerkdarstellung der Wechselwirkungen humaner Gene und Proteine, die für die Vermehrung des Hepatitis-C-Virus in menschlichen Zellen relevant sind. Die farblich hinterlegten Gene und Proteine sind den unten angeführten Funktionsbeschreibungen zugeordnet.**  
*Figure 2:*  
*Network representation of the interactions of human genes and proteins, which are relevant to the life cycle of the hepatitis C virus in human cells. The functions of the genes and proteins in colored areas are described at the bottom.*

**Aufschwung für die Bioinformatik an der TU Graz**

Die Bioinformatik ist eine junge, interdisziplinäre Wissenschaft und bildet eine Schnittstelle zwischen Biologie, Medizin, Statistik und Informatik. Die Anwendungsbereiche sind breit gefächert und reichen von neuen Methoden zur Diagnose und Therapie genetischer Krankheiten über die Unterstützung bei der Züchtung neuer Nutzpflanzenrasen bis hin zur Erzeugung von Medikamenten auf biologischen Wegen. Durch die ständige Erweiterung der Anwendungsfelder besteht nach wie vor ein hoher Bedarf an zusätzlichen Bioinformatikerinnen und Bioinformatikern weltweit. Universitäten sind daher gefragt, das Lehr- und Forschungsangebot auszuweiten – die TU Graz hat darauf entsprechend reagiert.

„Durch die Initiative BioTechMed ist es uns gelungen, die Professur Bioinformatik mit Mario Albrecht an der TU Graz einzurichten, parallel dazu konnte auch eine Professur für Computational Biotechnology mit Christoph Sensen besetzt werden“, freut sich Horst Bischof über die Neuzugänge und die damit verbundene Stärkung der Bioinformatik an der TU Graz. Während Christoph Sensen sich mit der Auswertung von Hochdurchsatzexperimenten aus der Genomforschung und der Entwicklung von Bioinformatikpipelines für solche Experimente auseinandersetzt (mehr über seine Forschungen auf Seite 15 f.), widmet sich Mario Albrecht einem kompletären Bereich der Bioinformatik.

**Bioinformatics revival at Graz University of Technology**

Bioinformatics is a relatively young, interdisciplinary research area, at the intersection of biology, medicine, statistics and informatics. The application domains of research results from bioinformatics are widespread, covering new diagnostic methods and therapeutic approaches for genetic diseases, the support of plant breeding, and the creation of new pharmaceutical compounds through biological production. Through the continual expansion of the application areas, there is a high demand for additional scientists with bioinformatics skills worldwide. Universities are therefore challenged to expand the teaching as well as the research in the area of bioinformatics, and Graz University of Technology is now responding to this with the recruitment of two new professors.

“Through BioTechMed, we were able to install the bioinformatics professorship for Mario Albrecht, in parallel, we were also able to recruit Christoph Sensen for a professorship for computational biotechnology”, reports Horst Bischof enthusiastically about the efforts to strengthen the field of bioinformatics at Graz University of Technology. While Christoph Sensen focuses on the analysis of high-throughput genomics experiments and the development of new bioinformatics analysis pipelines (see also page 15 f.), Mario Albrecht concentrates on a complementary field of bioinformatics research.

**New bioinformatics methods**

Mario Albrecht develops new bioinformatics methods that can be used to filter the most important facts from the vast experimental data collections in biology and medicine, with the goal of discovering and visualizing hidden molecular interactions. The main focus is on the integrative data analysis and creation of network models, in particular, with the goal of better understanding the cause and course of diseases. Through this, the bioinformatics professorship will contribute to all four areas of BioTechMed-Graz, but first and foremost to the area of molecular and quantitative biomedicine. The large quantities of genomic data, which are available today, have to be connected to functional knowledge about cellular processes, as well as quantitative results from high-throughput omics experiments such



```

1 11 21 31 41
2 ARTFFVGGNF K L N G K K E I V E R L N A I P E V E V V I C P P A Y L D Y
2 ARTPFVGGNW K M N G T K A E A K E L V E A L K A . K L P D D V E V V V A P P A V Y L D T A

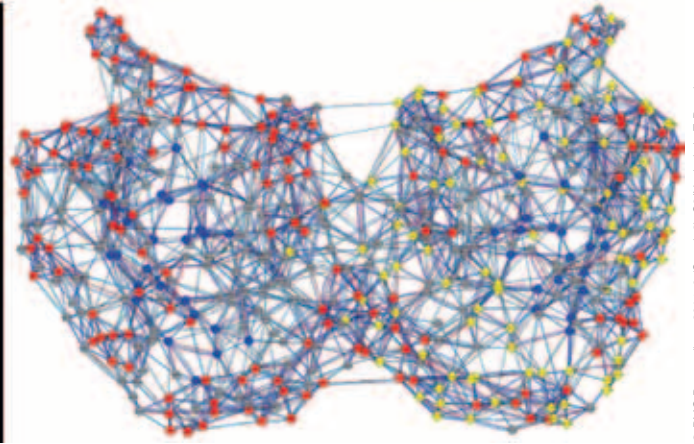
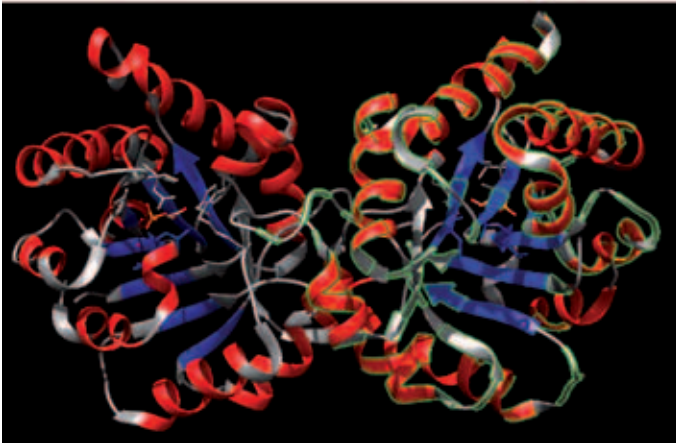
51 61 71 81 91
52 L V K K P V V G A Q N A Y L K A G A F T G E S V D I K D V G A K V I L G H S E R R
51 E A L K G S K I K V A A Q N C Y K E A K G A F T G E I S P E M L K D L G A D Y V I L G H S E R R H

101 111 121 131 141
102 F H E D D K F I A D K K E A L G S V G V I L C I B E T L E E K K A G K T L D V V E R Q L A V
101 F G E T D E L V A K K V A H A L E H G L K V I A C I G E T L E E R E A G N T E E V V F R Q T K A L

151 161 171 181 191
152 E E V K D . W N V V V A Y E P V W A I S T G L A A T P E D A Q D I H A I R K L A K I G D K
151 A S L G D E W K N V V I A Y E P V W A I S T E K T A T P E Q A Q E V H A F I R K W L A E N V S A E

201 211 221 231 241
201 A E L R I L Y G S S A G N A V F K D K A D D G F L V G G A S L K P E F L D I I N S R N
201 A E S V R I L Y G S S V K P A N A K E L A A O P D I D G F L V G G A S L K P E F L D I I N S R N

```



© BMC Proceedings 2014, 8 (Suppl. 2): S2/Nabzhoda T. Doncheva

### Neue Bioinformatikmethoden

Mario Albrecht entwickelt neue Bioinformatikmethoden, um die wichtigsten Informationen aus der großen experimentellen Datenflut der Biowissenschaften und Medizin herauszufiltern und verborgene molekulare Beziehungen zu entdecken und zu visualisieren. Der Fokus der Forschung liegt dabei auf der integrativen Datenanalyse und der Erstellung von Netzwerkmodellen, insbesondere um Ursachen und Verlauf von Krankheiten besser zu verstehen. So wird die Bioinformatik-Professur zu allen vier Forschungsthemen von BioTechMed-Graz beitragen, insbesondere zur molekularen, quantitativen Biomedizin. Die heutzutage in großen Mengen produzierten Genomikdaten sind oft mit Wissen über die Funktion zellulärer Vorgänge sowie mit quantitativen Omik-Hochdurchsatzdaten wie der Transkriptomik, Metabolomik und Proteomik zu verknüpfen. Durch diese Integration von genomischen Sequenzdaten mit Funktionswissen und dynamischen Daten wie zum Beispiel aus Genexpression, metabolischen Stoffwechselwegen und Proteininteraktionen wird es erst möglich, biologische Prozesse im Detail zu analysieren und zu modellieren. Um die Daten zu verarbeiten und hinsichtlich funktioneller Zusammenhänge zu interpretieren, entwickelt die Bioinformatik effiziente Informationssysteme und Algorithmen. Daher stehen im Rahmen der Professur nicht nur die methodischen Anwendungen der Bioinformatik in Biologie und Medizin im Fokus, sondern auch die >

as transcriptomics, metabolomics or proteomics. Through this integration of genomic sequences with knowledge about function as well as dynamic data such as from gene expression, metabolic pathways or protein interactions, it will become possible to analyze biological processes in detail and create models. Bioinformaticians develop efficient algorithms and information systems to process and interpret the data with the aim of elucidating the functional relationships. Therefore, the bioinformatics professorship is not only focused on the application of bioinformatics in the biological and medical field, but also on the computational aspects, such as the development of efficient network algorithms, database searches, knowledge technologies, modeling methods, and the visual analysis of molecular data. The developed bioinformatics methods are applied in close collaboration with biologists and medical researchers and implemented as software packages and web services so that they can be utilized easily by researchers in the life sciences. To this end, the bioinformatics agenda also includes training and teaching modules.

### Molecular and medical bioinformatics

Especially for medical research, where the topic of personalized medicine is beginning to play a larger role, it is still a big challenge to integrate molecular data with functional information about biological processes in healthy and diseased >

**Abbildung 3:** Visuelle Analytik von Sequenzvariationen unterschiedlicher Triosephosphat-Isomerasen (scTIM und dTIM). Die oben dargestellten Sequenzunterschiede sind unten in der Proteinstruktur (links) und der entsprechenden Netzwerkrepräsentation (rechts) abgebildet. **Figure 3:** Visual analysis of sequence variations of triosephosphate isomerases (scTIM and dTIM). The sequence differences displayed in the upper part are mapped to the protein structure (left) and the network representation (right) in the lower part.



## **Mario Albrecht: BioTechMed-Professor für Bioinformatik BioTechMed Professor of Bioinformatics**

Mario Albrecht studierte Informatik an der Universität in Passau, an der University of Manchester und an der Universität in Bonn. 2006 promovierte er in Bioinformatik an der Universität des Saarlandes. Von 1999 bis 2002 war er als wissenschaftlicher Mitarbeiter am Fraunhofer-Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI) in Sankt Augustin tätig, danach am Max-Planck-Institut für Informatik in Saarbrücken. Von 2008 bis 2012 war Albrecht an diesem Max-Planck-Institut Leiter der Forschungsgruppe „Molekulare Netzwerke in der Medizinischen Bioinformatik“. Von Jänner 2012 bis März 2014 war er Professor für Bioinformatik am Institut für Biometrie und Medizinische Informatik der Universitätsmedizin Greifswald, bevor der 39-Jährige als Professor für Bioinformatik an die TU Graz wechselte.

*Mario Albrecht studied informatics at the University of Passau, the University of Manchester and the University of Bonn. He gained his doctorate in bioinformatics at Saarland University in 2006. From 1999 to 2002 he was research associate at the Fraunhofer Institute for Algorithms and Scientific Computing SCAI in St. Augustin, following this at Max Planck Institute for Informatics at Saarbrücken. From 2008 to 2012 he was head of the research group "Molecular Networks in Medical Bioinformatics" at Max Planck Institute. From January 2012 to March 2014 he was professor of bioinformatics at the Institute of Biometrics and Medical Informatics at the Medical University of Greifswald before the 39-year-old was appointed professor of bioinformatics at Graz University of Technology.*

informatischen Aspekte wie effiziente Netzwerkalgorithmen, Datenbanksuchen, Wissenstechnologien und Modellierungsmethoden sowie die visuelle Analytik der molekularen Daten. Die entwickelten Bioinformatikmethoden werden außerdem in engen Kooperationen mit Biologinnen und Biologen sowie Medizinerinnen und Medizinern angewandt und in einsatzfähige Software und Webservices umgesetzt, sodass sie von Forscherinnen und Forschern in den Lebenswissenschaften leicht verwendet werden können. Hierfür sollen auch geeignete bioinformatische Trainingskurse und Lehrveranstaltungen angeboten werden.

### **Molekulare und medizinische Bioinformatik**

Gerade in der aktuellen medizinischen Forschung, in der die personalisierte Medizin immer wichtiger wird, ist es noch eine große Herausforderung, molekulare Daten zusammen mit Informationen über die Funktion biologischer Vorgänge im gesunden und kranken Zellzustand zur Interpretation individueller Variationen des menschlichen Genoms zu verwenden. Das Ziel der Bioinformatik ist es hierbei, krankheitsrelevante Genomvariationen wie zum Beispiel Sequenzmutationen zu identifizieren und ihre funktionelle Auswirkung auf die Krankheit möglichst genau abzuschätzen. Hierzu werden bioinformatische Methoden benötigt, die aus vorhandenen Daten und bekanntem biologischen Wissen mögliche molekulare Effekte ableiten können und zuverlässige funktionelle Vorhersagen generieren. Dies ist für den Arzt bzw. die Ärztin zur verlässlichen Diagnose, Prognose und Therapieplanung von Krankheiten von größter Wichtigkeit und erfolgreiche Bioinformatikmethoden werden immer mehr Verwendung in der klinischen Praxis finden.

Zur Repräsentation von biologischem Wissen eignen sich nicht nur Datenbanken, sondern insbesondere auch Netzwerke. Die Vielzahl von Wechselwirkungen unterschiedlicher Moleküle (u. a. Gene, Proteine, Metaboliten, Lipide und Nukleinsäuren) wird am besten durch Interaktionsnetzwerke dargestellt. Mittels Netzwerkalgorithmen der Informatik lassen sich diese Interaktionsnetzwerke näher analysieren und zur Interpretation molekularer

*cell states for the interpretation of individual variants of the human genome. The objective of bioinformaticians in this case is to identify genomic variations (for example, sequence mutations) relevant to the disease and to determine their functional impact. To achieve this, bioinformatics approaches are needed that can be used to derive potential molecular effects from existing data and biological knowledge and to generate reliable functional predictions. This is very important for the medical doctor for the dependable diagnosis, prognosis, and the planning of the therapy. Therefore, successful bioinformatics methods will be used more and more in clinical practice.*

*Not only databases, but especially networks are capable of representing biological knowledge. The multitude of interactions amongst molecules (e.g. genes, proteins, metabolites, lipids or nucleic acids) is best represented by interaction networks. The detailed analysis of these networks can be achieved through the use of network algorithms from computer science, which allows the interpretation of molecular high-throughput data (Figure 2). At the same time, networks can be used to model biological mechanisms on the molecular level, for example, the impact of sequence variations on protein function (Figure 3). For diseases, subnetworks can point to potential biomarkers. Thus the development of suitable innovative network methods for the study of dynamic processes in cells is an important task of current bioinformatics research.*

*Finally, bioinformatics becomes increasingly more important in the design and prioritization of new biological experiments through the generation of functional hypotheses. Time and resources can be saved through reliable predictions about cellular processes and their function on the molecular level. This requires sophisticated computer methods that calculate ranked lists of genes, proteins or molecular processes with respect to their relevance for the disease (or phenotype),*

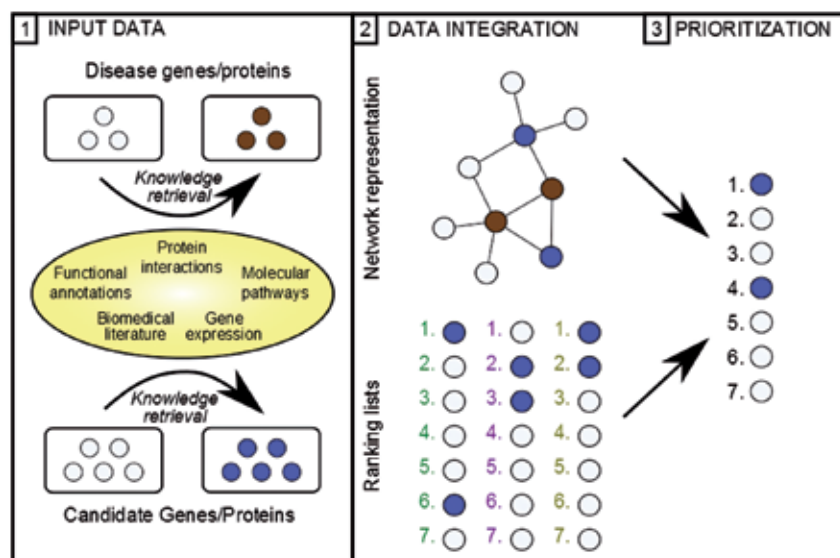


Hochdurchsatzdaten anwenden (Abb. 2). Gleichzeitig können mittels Netzwerken biologische Mechanismen auf molekularer Ebene modelliert werden, zum Beispiel die Auswirkung von Sequenzvariationen auf Proteinfunktionen (Abb. 3). Bei Krankheiten dienen Teilnetzwerke auch als mögliche Biomarker. Daher ist die Entwicklung geeigneter innovativer Netzwerkmethoden eine wichtige Aufgabe der aktuellen bioinformatischen Forschung zu dynamischen Vorgängen in der Zelle.

Schließlich spielt die Bioinformatik durch die Generierung funktioneller Hypothesen eine immer bedeutsamere Rolle bei der Auswahl und Priorisierung weiterer Experimente. Durch verlässliche Vorhersagen über zelluläre Prozesse und ihre molekulare Beeinflussung können im Labor viel Zeit und enorme Kosten eingespart werden und so die Erfolge von Experimenten optimiert werden. Hierfür benötigt man ausgefeilte Computermethoden, die aus diversen Datenmengen und zusätzlichem biologischen Wissen Ranglisten von Genen, Proteinen oder molekularen Prozessen im Hinblick auf ihre Relevanz für eine Krankheit (oder allgemein einer zellulären Situation) berechnen (Abb. 4). Oft sind auch völlig unterschiedliche Ranglisten miteinander zu vergleichen und zu integrieren, um eine finale Priorisierung vornehmen zu können. Auch auf diesem Forschungsgebiet, das seine informatischen Wurzeln z. B. bei der Priorisierung von Webseiten hat, gibt es noch einige Herausforderungen zu lösen, um die besonderen biologischen Gegebenheiten in den Ranglistenalgorithmen zu berücksichtigen.

### Grazer Forschungsarbeiten in Bio- und Medizininformatik

Insgesamt lässt sich beobachten, dass die Bedeutung der Bioinformatik in den letzten Jahren aufgrund der generierten Datenmengen und der damit möglichen Anwendungen in den Lebenswissenschaften stark zunahm und sich die medizinische Bioinformatik auch immer mehr mit der klinischen Medizininformatik verzahnt. Daher organisiert Mario Albrecht zusammen mit seinem Kollegen Christoph Sensen ab Jänner eine universitätsübergreifende Vortragsreihe verschiedener Grazer Institutionen, in der die Methoden und Werkzeuge der Bio- oder Medizininformatik möglichen Benutzerinnen und Benutzern mit ihren jeweiligen Vor- und Nachteilen vorgestellt werden sowie in Graz vorhandene, wichtige Datenressourcen präsentiert werden. ■



© Mario Albrecht – TU Graz

using diverse datasets and additional biological knowledge as input (Figure 4). It is often necessary to compare and integrate disparate ranking lists to achieve the final prioritization. In this research area, which initially emerged for the ranking of web pages, there are several unsolved challenges, especially the use of biological information in the ranking lists.

### Research in Graz on bioinformatics and medical informatics

Overall, it can be stated that the importance of bioinformatics has increased vastly in recent years, particularly, due to the large datasets that are now being generated and their applicability in the life sciences. The medical aspects of bioinformatics are also more and more interconnected with clinical medical informatics. Therefore, Mario Albrecht and Christoph Sensen will organize a new inter-university seminar series within Graz, starting in January, where methods and tools for bioinformatics and medical informatics, as well as important data resources from Graz, are introduced to potential users, and their pros and cons will be discussed. ■

**Abbildung 4:** Dreistufiges Priorisierungsverfahren zur Identifizierung von weiteren Kandidatengenen/-proteinen für Erkrankungen, für die bereits ein paar Krankheitsgene/-proteine bekannt sind: 1. Wissensverarbeitung und Funktionsannotation. 2. Datenintegration mittels Netzwerken oder Ranglisten. 3. Priorisierung der finalen Rangliste.

**Figure 4:** Three-step prioritization approach for the identification of additional candidate genes/proteins for diseases for which some of the genes/proteins involved are already known: 1. Knowledge processing and functional annotation. 2. Data integration using networks or ranking lists. 3. Prioritization of the final ranking list.