



Liebe Leserinnen, liebe Leser,

wir freuen uns, Ihnen nach einer längeren, Corona-bedingten Pause eine neue Ausgabe des Forschungsmagazins ‚Blick in die Wissenschaft‘ in der Ausgabe 44/45 präsentieren zu können.

Die Corona-Pandemie hat auch die Universität Regensburg und alle ihre Mitglieder vor große Herausforderungen gestellt. Dennoch konnten zentrale Zukunftsprojekte weitergeführt und umgesetzt werden. So stellt vor allem die Gründung unserer neuen Fakultät für Informatik und Data Science (FIDS) einen wahren Meilenstein in der Geschichte und Entwicklung der Universität Regensburg dar. Als größtes Strukturprojekt seit der Gründung der Fakultät für Medizin vor 30 Jahren ist unsere Informatikfakultät ein Zukunftsprojekt von weitreichenden Dimensionen. Mit der neuen strategischen Schwerpunktsetzung im Bereich Informatik und Data Science und vor allem auch der Querschnittsorientierung der neuen Fakultät sieht sich die Universität Regensburg sehr gut gerüstet, ihre bisherigen Stärken in diesen Bereichen zu bündeln, weiter auszubauen und zu

vertiefen. Schließlich sind *Digital Transformations* als eines der vier Gestaltungsfelder und Zukunftsthemen in unserem *Hochschulentwicklungsplan 2025* fest verankert. Dieses Gestaltungsfeld adressiert die neue Fakultät ebenso wie den Bereich *Integrated Sciences in Life, Health, and Disease* als ein weiteres Schwerpunktgebiet unserer Universität.

Die Grundsatzbeschlüsse in den Gremien der Universität Regensburg im Sommer und Herbst 2019 zur Einrichtung der neuen Fakultät erfolgten nach einer vorhergehenden Phase intensiver Planungen dann letztlich fast zeitgleich mit der Regierungserklärung des Bayerischen Ministerpräsidenten Dr. Markus Söder am 10. Oktober 2019 und der Verkündung der Hightech Agenda Bayern. Unterstützt und beschleunigt durch die Mittel der Hightech Agenda Bayern konnte der Auf- und Ausbau der Fakultät für Informatik und Data Science zügiger umgesetzt werden, nachdem die neue Fakultät im März 2020 formal gegründet und im Laufe des WS 2021/22 aus sich heraus handlungs- und

funktionsfähig wurde. Im Mai 2022 konnten wir gemeinsam mit Ministerpräsident Dr. Markus Söder und Staatsminister für Wissenschaft und Kunst Markus Blume den offiziellen Kickoff für die Fakultät begehen. Dass dieser komplexe Prozess im Kontext der Herausforderungen der Corona-Pandemie vollzogen und abgeschlossen werden konnte, ist ein Zeichen für die Bedeutung dieser gesamtuniversitär-strategischen Maßnahme und für den Rückhalt für das Großprojekt in der universitären Gemeinschaft.

Im Laufe des Gründungsprozesses ist es gelungen, die verschiedenen Informatiknahen und -interessierten Kräfte der Universität an einen Tisch zu bringen und gemeinsam ein zukunftsorientiertes Konzept für die Fakultät zu entwickeln. Ein externes Gutachten mit hochrangiger Expertise skizzierte und evaluierte 2019 wesentliche inhaltliche Schwerpunkte und Strukturierungen für die neue Fakultät, an denen sich in den Jahren 2019-2021 die von Vizepräsident Prof. Dr. Nikolaus Korber geleitete Gründungskommission in der

konkreten Arbeit zum Aufbau der Fakultät orientierte. In insgesamt 15 Berufungsverfahren wurden die ersten neuen Professuren in der Fakultät zügig besetzt – ein Prozess, der in Kürze abgeschlossen sein wird. Im Besetzungsprozess hat sich vor allem auch gezeigt, wie attraktiv die Neugründung einer Fakultät und die Möglichkeiten zur Mitgestaltung und zum Aufbau neuer Strukturen für Wissenschaftlerinnen und Wissenschaftler sind und wie viel Zukunftspotential von unserer neuen Fakultät ausgeht. So konnten wir zum Wintersemester 2023/24 130 Studierende für den B.Sc. Informatik und den B.Sc. Data Science begrüßen.

Im vorliegenden Heft von ‚Blick in die Wissenschaft‘ möchten wir Ihnen nunmehr vor allem die Forschungsaktivitäten der Fakultät für Informatik und Data Science näher vorstellen. Dabei beglückwünsche ich die Fakultät, dass sie bereits eineinhalb Jahre nach ihrer vollständigen Handlungs- und Funktionsfähigkeit und während der weiteren Planungen zum Aufbau und der Ausarbeitung ihrer Studiengänge insbe-

sondere im Master-Bereich ein so vielfältiges Themenheft zu ihren aktuellen Forschungsarbeiten vorlegen konnte.

Das facettenreiche und vielfältige Themenspektrum dieses Sonderhefts illustriert, wie die Fakultät für Informatik und Data Science die an der Universität Regensburg bisher vorhandenen IT-Kompetenzen erfolgreich bündelt und in die Zukunft gerichtet erweitert. Sie ermöglicht die essentielle interdisziplinäre Vernetzung mit der gesamten Universität, von den Geistes- und Sozialwissenschaften bis zu den Natur- und Lebenswissenschaften. Die Beiträge verdeutlichen, wie interdisziplinäre Forschung das Fundament starker methodischer und fachlicher Grundlagen weiterentwickelt und wie die bisherigen Informatik-Schwerpunkte der Universität Regensburg (Computational Science, Informationswissenschaft, Medieninformatik, Wirtschaftsinformatik) erfolgreich in die neue Fakultät überführt werden konnten und die Querschnittsorientierung unterstützen.

Das vorliegende Heft mit seinem Schwerpunkt auf aktuellen Forschungsar-

beiten begleitet den im Wintersemester 2023/24 erfolgten Start der beiden grundständigen Bachelor-Studiengänge Informatik und Data Science. Ein kurzer Überblicksbeitrag zur Lehre in der neuen Fakultät zeigt anschaulich die bereits gewachsene Vielfalt der Informatikstudiengänge und die intensive und gelebte Verbindung von Forschung und Lehre auch an dieser neuen Fakultät.

Unsere neue Fakultät leistet hervorragende Arbeit und ich bin sicher, dass Ihnen die nachfolgenden Seiten einen spannenden Einblick in die verschiedenen Facetten der FIDS geben werden. Ein ganz besonderes Dankeschön möchte ich an dieser Stelle an das gesamte Dekanat der Fakultät für Informatik und Data Science und insbesondere an Forschungsdekanin Prof. in Dr. Meike Klettke richten, die für diese Sonderausgabe die Koordinationsarbeit der vorliegenden Ausgabe federführend übernommen hat.

Prof. Dr. Udo Hebel
Präsident der Universität Regensburg

**Blick in die Wissenschaft
Forschungsmagazin
der Universität Regensburg**

ISSN 0942-928-X

Heft 44/45

31. Jahrgang

Herausgeber

Prof. Dr. Udo Hebel
Präsident der Universität Regensburg

Redaktionsleitung für diese Ausgabe

Prof.in Dr. Meike Klettke / Fakultät für Informatik und Data Science

Redaktionsbeirat

Prof. Dr. jur. Christoph Althammer
Prof. Dr. rer. nat. Ferdinand Evers
Prof. Dr. rer. nat. Stefan Friedl
Prof. Dr. rer. nat. Mark W. Greenlee
Prof. Dr. theol. Andreas Merkt
Prof. Dr. phil. Omar W. Nasim
Prof. Dr. rer. nat. Klaus Richter
Prof. Dr. rer. pol. Daniel Rösch
Prof. Dr. med. Ernst Tamm
Prof. Dr. paed. Oliver Tepner
Prof. Dr. phil. Christiane Heibach

Universität Regensburg
93040 Regensburg
Telefon +49 941 9432300
Telefax +49 941 9433310

Verlag

Universitätsverlag Regensburg GmbH
Leibnizstraße 13, 93055 Regensburg

Telefon +49 941 78785-0
Telefax +49 941 78785-16

info@univerlag-regensburg.de
www.univerlag-regensburg.de
Geschäftsführer: Dr. Albrecht Weiland,
Felix Weiland M.A.

Abonnementsservice

bestellung@univerlag-regensburg.de

Anzeigenleitung

Larissa Nevecny
MME-Marquardt
info@mme-marquardt.de

Herstellung

Universitätsverlag Regensburg GmbH
info@univerlag-regensburg.de

**Einzelpreis € 7,00
Doppelheft € 14,00**

Jahresabonnement

bei zwei Ausgaben pro Jahr

€ 10,00 / ermäßigt € 9,00

Für Schüler, Studierende und Akademiker/innen im Vorbereitungsdienst (inkl. 7% MwSt.) zzgl. Versandkostenpauschale € 1,64 je Ausgabe. Bestellung beim Verlag. Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

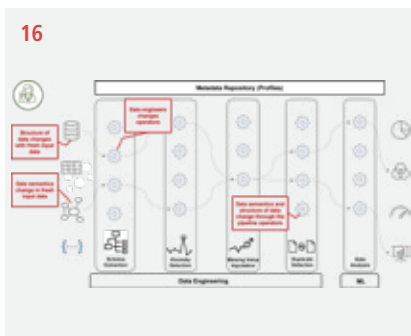
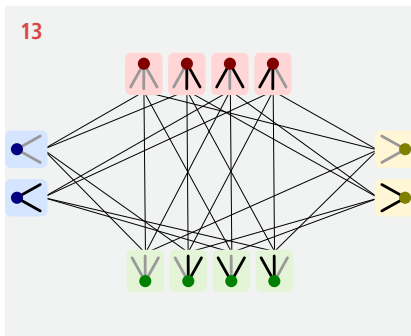
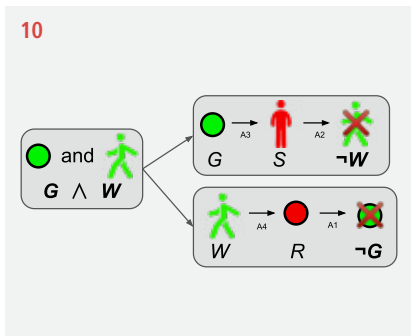
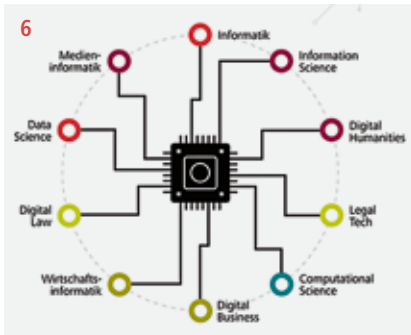


Rohstoffe
Transporte
Produktion

g CO₂e
492
Pro Produkt

CO₂-Emissionen
ausgeglichene

Inhalt



Einleitung 5
Florian Erhard, Bernd Heinrich, Meike Klettke, Christian Wolff

Lehre an der Fakultät für Informatik und Data Science 6
Florian Erhard, Udo Kruschwitz, Bernd Heinrich, Christian Wolff

Automatisches Beweisen: Methoden und Anwendungen 10
Julie Cailler, Philipp Rümmer

Algorithmen und Komplexitätstheorie 13
Radu Curticapean

Evolution in Datenbanken und Data Engineering Workflows 16
Meike Klettke

IoT-basiertes Prozessmanagement – Mobile Benutzerführung in der digitalen Fabrik 19
Stefan Schönig

Cyber Threat Intelligence: Gemeinschaftliche IT-Sicherheit durch den Austausch von Informationen 23
Johannes Grill, Daniel Schlette, Günther Pernul

Kann man den Entscheidungen Künstlicher Intelligenz trauen? Zu den Auswirkungen unsicherer Daten auf die Entscheidungen Neuronaler Netze 26
Thomas Krapf, Bernd Heinrich

Mensch vs. Maschine: Wettbewerb und Kooperation mit künstlicher Intelligenz in digitalen Märkten 30
Andreas Schauer, Daniel Schnurr

Notfallpläne für den Ernstfall testen 34
Maria Leitner

Maschinelles Lernen mit Anwendungen in den Naturwissenschaften 37
Merle Behr, Markus Schmitt

Automatisierte, KI-basierte Analyse von Bilddaten:

Der Lehrstuhl für Bildverarbeitung

Dorit Merhoff

40

Die Genome des Menschen – Forschungsschwerpunkte der Arbeitsgruppe für Algorithmische Bioinformatik

Birte Kehr

43

Algorithmen zum Entschlüsseln der Genregulation

Francisca Rojas Ringeling, Stefan Canzar

46

Mit Hilfe von Daten Immunprozesse entschlüsseln:

Der Lehrstuhl Computational Immunology

Florian Erhard

49

Maschinelles Lernen enthüllt den verborgenen Prozess der Tumorentstehung

Linda Hu, Andreas Lösch, Rainer Spang

52

Allgegenwärtige Mensch-Maschine-Interaktion: Entwicklung, Forschung und Infrastruktur der Medieninformatik

Raphael Wimmer, Johanna Bogon, Niels Henze, Christian Wolff

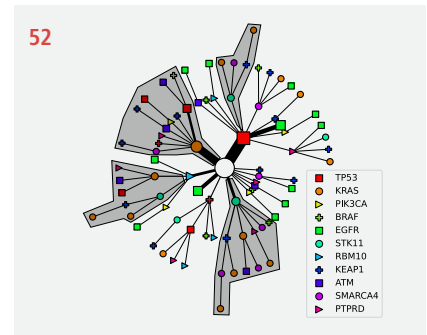
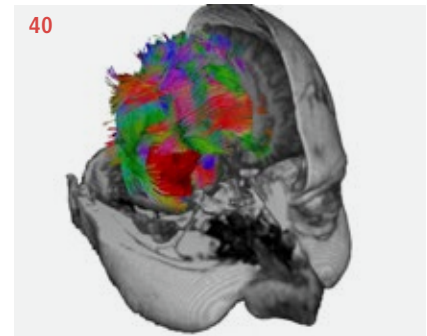
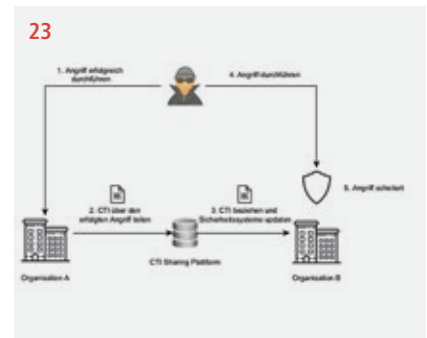
54

Wissen aus dem Internet – Genug, genau, geprüft und geeignet?

Informationswissenschaft in Regensburg

Udo Kruschwitz, Bernd Ludwig, David Elsweiler

60



SOFTWAREENTWICKLER (M/W/D) GESUCHT

Voll- oder Teilzeit (30 -40 Std.) | Alle Fachbereiche

Hallo, **wir sind mb Support.**

Seit 20 Jahren unterstützen wir die Versicherungswirtschaft mit unseren Lösungen, u.a. Konzerte und Kunstwerke weltweit zu versichern. Als inhabergeführtes **Familienunternehmen** arbeiten wir auf Augenhöhe in einer familiären Atmosphäre. Als stark wachsendes und profitables Unternehmen bieten wir gleichzeitig **attraktive Benefits**.

Wer bist du? Jedes unserer rund 60 Teammitglieder bringt seine eigene Persönlichkeit, Geschichte und Perspektive mit. Für uns zählt, was dich interessiert, was dich antreibt, wie du bist. Du entscheidest, wie du bei uns mitgestalten möchtest. **Wir freuen uns auf dich.**

Die gesamte Stellenausschreibung findest du auf www.mbsupport.de/karriere.

www.mbsupport.de | Friedenstraße 18 | 93053 Regensburg | +49 941 942 60 0 | mb Support GmbH



Die Genome des Menschen

Forschungsschwerpunkte der Arbeitsgruppe für Algorithmische Bioinformatik

Prof. Dr. Birte Kehr

Im Jahr 2000 verkündeten der amerikanische Präsident Bill Clinton und der britische Premierminister Tony Blair die Entschlüsselung des menschlichen Genoms. Damit war ein Meilenstein der Wissenschaft erreicht, der seitdem aus der modernen biomedizinischen Forschung auf molekularer Ebene nicht mehr wegzudenken ist. Dennoch war die zunächst veröffentlichte DNA-Sequenz weit davon entfernt, das Genom eines jeden Menschen abzubilden. Einerseits war sie nicht vollständig, denn es fehlten die schwer zu entschlüsselnden Telomer- und Centromerregionen. Andererseits gibt es Millionen von Unterschieden zwischen den Genomen zweier beliebiger Menschen mit Ausnahme eineiiger Zwillinge, so dass eine lineare DNA-Sequenz immer nur das Genom eines einzelnen Menschen darstellen kann. Heute ist klar, dass sich die Wissenschaft nicht allein auf eine lineare DNA-Sequenz stützen kann, sondern dass zusätzlich die genomische Variation abgebildet werden muss. Hierzu trägt die Forschung der Arbeitsgruppe für Algorithmische Bioinformatik bei.

Variation in der Genomsequenz ist verantwortlich für die sichtbare Diversität unter uns Menschen, beispielsweise für unsere Augenfarbe oder Körpergröße, aber ebenso auch für Unterschiede, die zunächst nicht sichtbar sind. Denn unser Genom beeinflusst, wie hoch unser Risiko ist, Krankheiten zu bekommen, und wie gut ein Patient auf eine medizinische Behandlung anspricht. Je besser daher unsere Fähigkeiten werden, Variationen in der Genomsequenz zu erfassen, umso

besser können wir die Hintergründe von Krankheiten aufklären und umso mehr von diesem Wissen kann in die biomedizinische Forschung und letztendlich in die Medizin einfließen.

Bis es jedoch routinemäßig möglich wird, die Variation in Patientengenomen annähernd vollständig zu erfassen, gibt es noch große Herausforderungen zu bewältigen. Beispielsweise sind die technologischen Möglichkeiten, um DNA-Sequenz aus einem DNA-Molekül auszulesen, begrenzt. Aktuell können nur relativ kurze Sequenzabschnitte am Stück »sequenziert« werden, die anschließend mithilfe von Methoden der Informatik wieder zusammengesetzt werden. Um die Variation in einem Genom zu erfassen, werden Millionen solcher kurzer Sequenzabschnitte, die in der Fachsprache einfach »Reads« genannt werden, erzeugt. Die schiere Größe der Genomdatensätze erfordert einen sparsamen Umgang mit Rechenzeit und Speicherverbrauch. Gleichzeitig müssen Fehleraten der Sequenzierungstechnologien in der Analyse der Daten sorgfältig behandelt werden. Unter anderem damit beschäftigt sich die aktuelle Forschung im Feld der Genomik und Bioinformatik.

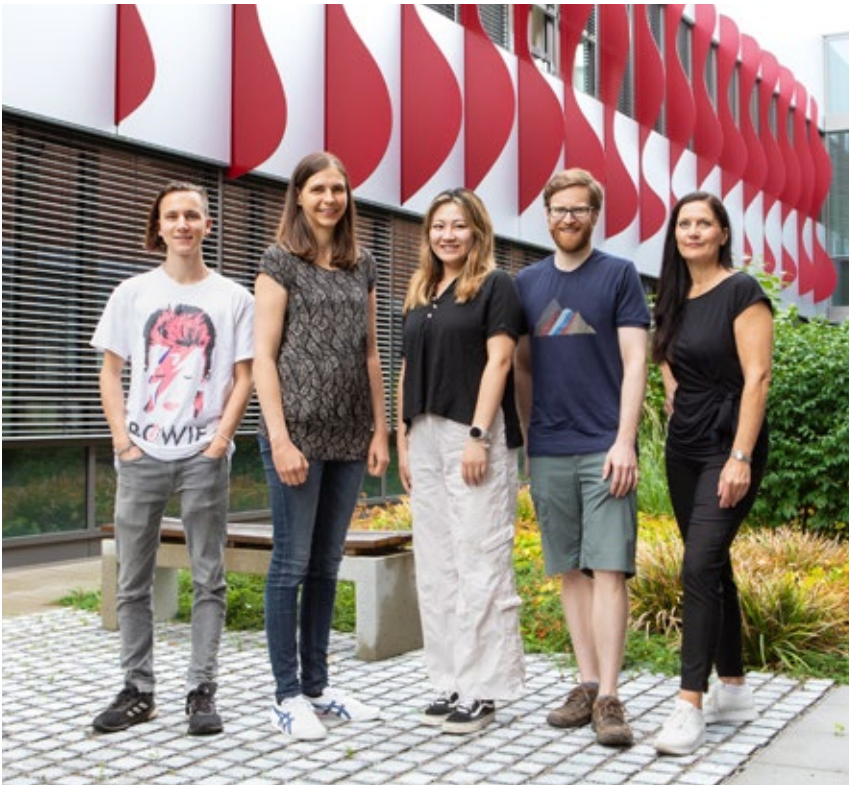
Bioinformatiker stellen sich Genome als lange Texte bestehend aus den Zeichen A, C, G und T vor. Hinter den Zeichen verbirgt sich die biochemische Zusammensetzung von DNA, die Nukleotide. Das menschliche Genom besteht aus gut 3 Milliarden aneinandergeschlossenen Nukleotiden, belegt also als Textdatei etwa 3 Gigabyte an Speicher auf einem Computer und würde ausgedruckt ein ganzes Regal voller Bücher füllen. Es ist in fast jeder Zelle unseres Körpers zu finden und ist dort in Chromosomen or-

ganisiert. Jedes Chromosom ist ein langes DNA-Molekül, aufgewickelt und verpackt in Proteine.

Im menschlichen Genom liegt jedes Chromosom – und damit alle 3 Milliarden Nukleotide – in zwei Kopien vor, da jeweils ein Chromosomensatz von der genetischen Mutter und ein Chromosomensatz vom genetischen Vater vererbt wird. Diese Kopien sind aktuellen Schätzungen zufolge zu etwa 99.6% identisch (NHGRI 2023), sofern unsere Eltern nicht nahe miteinander verwandt sind. Auf die Länge des menschlichen Genoms bedeuten 0.4% Abweichungen jedoch immer noch mehrere Millionen Unterschiede, die als Varianten bezeichnet werden. Natürlich findet man nicht nur zwischen den an uns von unseren Eltern vererbten Kopien des Genoms Millionen von Varianten, sondern auch beim Vergleich der Genome zweier beliebiger Menschen. Mindestens ein gewisser Anteil dieser Varianten kann physiologische Unterschiede zwischen uns Menschen erklären.

Die größte Zahl der Varianten sind Einzelnukleotidvarianten (engl. Single nucleotide variants, kurz SNVs), bei denen sich die Sequenz in genau einem Zeichen (Nukleotid) unterscheidet. Kleine Insertionen und Deletionen von 1–50 Nukleotiden werden als »Indels« bezeichnet. Beim Vergleich zweier menschlicher Genome findet man Millionen von SNVs und Hunderttausende Indels. Varianten, die einen Sequenzabschnitt von mehr als 50 Nukleotiden betreffen, bezeichnet man als »Strukturvarianten«. Zu ihnen zählen große Deletionen, Insertionen, Duplikationen, Inversionen, Translokationen und komplexe Varianten, an denen mehrere Sequenzabschnitte be-

Bildnachweis: UKR, Universitätsklinikum Regensburg



1 Die Arbeitsgruppe für Algorithmische Bioinformatik am Leibniz-Institut für Immuntherapie.

teilt sind. Aufgrund ihrer Größe betreffen Strukturvarianten weit mehr Nukleotide des menschlichen Genoms als SNVs und Indels zusammen, obwohl schätzungsweise nur 20–30 Tausend Strukturvarianten pro Genom vorkommen (Audano et al. 2019).

Die Arbeitsgruppe für Algorithmische Bioinformatik (Abb. 1) entwickelt neue Ansätze, um genomische Variation auf der Grundlage von Genomsequenzdaten zu beschreiben. Insbesondere interessiert uns, an welchen Positionen im Genom Strukturvarianten liegen und wie sie aussehen (Detektion) und, ob sie auf keiner, einer oder beiden Kopien der Chromosomen vorliegen (Genotypisierung). Dabei formulieren wir die konkreten Fragestellungen nach Möglichkeit als bekannte algorithmische Optimierungsprobleme, um sie dann mit etablierten Verfahren der Informatik lösen zu können. Fehlerwahrscheinlichkeiten

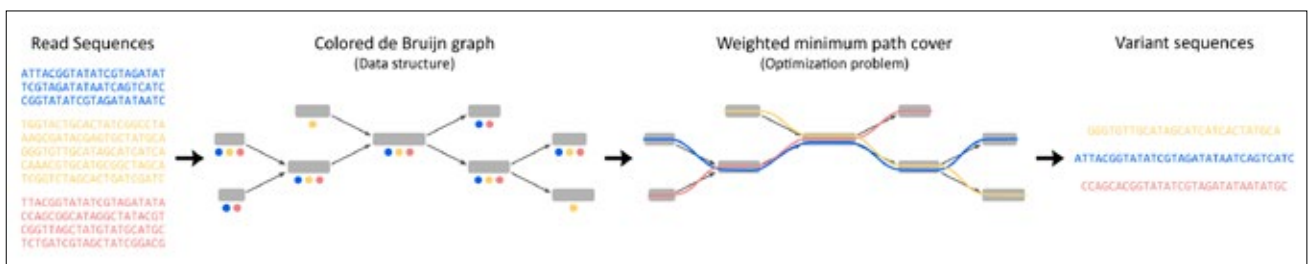
und andere Unsicherheiten in den Daten beschreiben wir mit statistischen Modellen und lassen sie als Gewichte in die Optimierungsprobleme einfließen.

Ein Beispiel für unsere Arbeiten ist die Skalierung von bestehenden Ansätzen für die Detektion und Genotypisierung von Strukturvarianten auf die stetig wachsende Zahl sequenzierter Genome (Krannich et al. 2022, Niehus et al. 2021). Wir wollen nutzen, dass die Daten von vielen Genomen sich gegenseitig ergänzen und die Evidenz für viele Varianten mit jedem analysierten Genomdatensatz wächst. Dadurch, dass die allermeisten genomischen Varianten vererbt werden, kommt ein großer Anteil von ihnen in mehr als einer Person vor. Jedes Kind erbt die Variation von seinen Eltern, die diese von den Großeltern und diese wieder von ihren Eltern geerbt haben, und so weiter. Als Folge der Ver-

erbungsregeln gibt es über alle Kontinente hinweg Varianten, die häufiger oder weniger häufig vorkommen. Für eine vererbte Variante gilt also, dass es umso wahrscheinlicher wird, sie in mehr als einem Genom zu finden, je mehr Genome wir gleichzeitig nach Varianten durchsuchen. Je mehr Daten (Reads) eine Variante wiederum beschreiben, umso sicherer können wir uns sein, dass es sich tatsächlich um eine Variante und keine Sequenzierungsfehler handelt. Varianten können durch eine Skalierung auf viele Genome also mit größerer Sicherheit und damit umfassender erfasst werden. Die Skalierung setzen wir um, indem wir die Read-Sequenzen in anderen Datenstrukturen repräsentieren als die Read-Sequenzen einzelner Genome, denn es muss ja zusätzlich abgebildet werden, welche Read-Sequenz aus welchem Genom stammt. Auf diesen Datenstrukturen formulieren wir wiederum andere Optimierungsprobleme, so dass schlussendlich neue Detektions- und Genotypisierungsansätze entstehen (Abb. 2).

Je nach Art der genomischen Varianten bieten sich andere Möglichkeiten, sie zu detektieren und genotypisieren, und entsprechend unterschiedlich sehen auch die Lösungsansätze aus. Beispielsweise erfordert die Beschreibung großer Insertionsvarianten, bei denen zusätzliche, unbekannte Sequenz beobachtet wird, dass diese zusätzliche Sequenz aus den kurzen Read-Sequenzen zusammengesetzt wird. Eine Deletionsvariante, bei der ein Sequenzabschnitt fehlt, ist dagegen durch eine Anfangs- und Endposition in einem Referenzgenom vollständig beschrieben.

Auch die Sequenzierungstechnologie, mit der die zugrundeliegenden Daten erzeugt werden, spielt eine große Rolle für die Ausgestaltung eines neuen Ansatzes. Technologien, die kurze Reads von 150 Zeichen auslesen, haben niedrige Fehlerraten und die Fehler, die sie machen, sind größtenteils Substitutionen. Diese Daten sind in vielen Projekten aus Kostengründen die erste Wahl und sind entsprechend in gro-



2 Die Suche nach genomischen Insertionsvarianten formuliert als das gewichtete Minimum Path Cover Optimierungsproblem.

Bildnachweis: Birte Kehr

ßer Zahl verfügbar. Technologien, die längere Reads von 10.000 bis über 100.000 Zeichen an einem Stück auslesen können, haben noch deutlich höhere Fehlerraten, wobei die meisten Fehler kurze Insertionen und Deletionen sind. Ein Großteil der heute bekannten Strukturvarianten konnte erst mit solch langen Reads detektiert werden und auch die eingangs genannten Telomer- und Centromerregionen wurden mit diesen Daten erstmals im Jahr 2022 zugänglich (Nurk et al. 2022).

Fest steht, dass all diese Variation in einer linearen DNA-Sequenz nicht abgebildet werden kann. Im Forschungsfeld wird seit über 10 Jahren diskutiert, welche Darstellung des menschlichen Genoms einschließlich der in der Bevölkerung existierenden Variation am geeignetsten ist. Sich verzweigende Graphdatenstrukturen scheinen unabdingbar, konnten sich aber in der Breite noch nicht durchsetzen. Ein Grund dafür mag sein, dass jegliche Ansätze zur Analyse von Sequenzierungsdaten auf ei-

nem linearen Referenzgenom basieren und für die neue Datenstruktur angepasst oder sogar neu entwickelt werden müssen. Der potenzielle Nutzen solch eines menschlichen »Pangenoms« ist jedoch unumstritten (Liao et al. 2023).

Zuletzt sei angemerkt, dass auch ein Pangenom immer unvollständig bleibt, weil es nur die bekannte Variation enthalten kann. Mit der Geburt eines Kindes gelangt jedoch neue Variation in die Bevölkerung. Die Anzahl neuer Mutationen pro Kind schätzen wir mit durchschnittlich etwa 120 zwar sehr gering (Jonsson et al. 2017, Kristmundsdottir et al. 2023), sie sind jedoch besonders für die Diagnose seltener Krankheiten von großer Bedeutung. Und auch die Veränderungen der Genomkopien in all unseren Zellen wird ein Pangenom nicht abbilden können. Bei jeder Zellteilung können Kopierfehler im Genom passieren, die sich über das Leben eines Menschen anreichern und nicht an Kinder vererbt werden. Im schlimmsten

Fall führen sie zu unkontrollierter Zellteilung und Krebs. Aber auch gesunde Zellen verändern sich mit der Zeit, angefangen mit stark expandierenden Immunzellen bis hin zu langlebigen Nervenzellen. Die unterschiedlichen Genome des Menschen weiter zu erforschen und somit Grundlagen für die Medizin zu schaffen, bleibt die Aufgabe der Bioinformatik in der Zukunft.

Referenzen

National Human Genome Research Institute (NHGRI). **Fact Sheet: Human Genomic Variation**. Retrieved October 6, 2023, from <https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-variation>

Audano PA, Sulovari A, Graves-Lindsay TA, ..., Eichler EE. **Characterizing the major structural variant alleles of the human genome**. *Cell* 2019; 176(3):663-675.e19

Krannich T, White WTJ, Niehus S, ..., Kehr B. **Population-scale detection of non-reference sequence variants using colored de Bruijn graphs**. *Bioinformatics* 2022; 38(3):604-611

Niehus S, Jonsson H, Schönberger J, ..., Kehr B. **PopDel identifies medium-size deletions simultaneously in tens of thousands of genomes**. *Nature Communications* 2021; 12:730

Nurk S, Koren S, Rhie A, ..., Phillippy AM. **The complete sequence of a human genome**. *Science* 2022; 376(6588):44-53

Liao WW, Asri M, Ebler J, ..., Paten B. **A draft human pangenome reference**. *Nature* 2023; 617: 312-324

Jonsson H, Sulem P, Kehr B, ..., Stefansson K. **Parental influence on human germline de novo mutations in 1,548 trios from Iceland**. *Nature* 2017; 549(7673):519-522

Kristmundsdottir S, Jonsson H, Harderson MT, ..., Stefansson K. **Sequence variants affecting the genome-wide rate of germline microsatellite mutations**. *Nature Communications* 2023; 14:3855

DREI FRAGEN ...

Was möchten Sie in den nächsten Jahren entwickeln?

Bei einer Immunreaktion teilen sich bestimmte Immunzellen sehr häufig und bilden sogenannte »Klone«. Ich möchte neue Ansätze entwickeln, die es erlauben, die Rolle von genomischen Veränderungen in Immunzellklonen zu untersuchen.

Die Lösung von welchem fachlichen Thema haben Sie sich leichter vorgestellt?

Bisher alle. Jede neue Aufgabe in der Forschung birgt unerwartete Wendungen. Es scheint eher die Ausnahme, dass die Lösung eines Themas so ausfällt, wie man sie sich zu Beginn vorstellt. Zeitpläne multipliziere ich daher prinzipiell mal drei.

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

Ein persönlicher Wunsch: Die Variation im Kringle-Repeat des LPA-Gens positionsgenau auflösen, dessen direkte Rolle für das genetisch bedingte Risiko für Herz-Kreislauf-Erkrankungen ungeklärt ist.



Foto © Universität Regensburg

Prof. Dr. Birte Kehr

leitet seit 2020 die Forschungsgruppe für Algorithmische Bioinformatik am Leibniz-Institut für Immuntherapie und hat eine W2-Professur an der FIDS inne. Zuvor baute sie eine Nachwuchsgruppe am Berlin Institute of Health und der Charité-Universitätsmedizin Berlin auf, nachdem sie drei Jahre lang bei deCODE generics/Amgen Inc. mit Sitz in Reykjavik, Island, gearbeitet hatte. Sie promovierte am Fachbereich Mathematik und Informatik der Freien Universität Berlin und erhielt ihr Diplom in Bioinformatik von der Friedrich-Schiller-Universität Jena.