

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Abschlussbericht

Zuwendungsempfänger: Technische Universität München

Projektleiter: Prof. Dr. Julien Gagneur

Projekttitle: Personalized Mitochondrial Medicine (PerMiM): Optimizing diagnostics and treatment for patients with mitochondrial diseases

Personalisierte Mitochondrienmedizin (PerMiM): Optimierung der Diagnose und Behandlung von Patienten mit mitochondrialen Erkrankungen

Förderkennzeichen: **01KU2016B**

Laufzeit des Projektes: 36 Monate + 12 Monate Verlängerung

Berichtszeitraum: 01.07.2020 - 30.06.2024

Kontaktperson: Prof. Dr. Julien Gagneur
Technische Universität München, Institut für Informatik
Boltzmannstraße 3
85748 Garching
gagneur@in.tum.de
+49 89 289 19411

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Berichtsblatt

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel Personalisierte Mitochondrienmedizin (PerMiM): Optimierung der Diagnose und Behandlung von Patienten mit mitochondrialen Erkrankungen	
4. Autor(en) [Name(n), Vorname(n)] Gagneur, Julien Yépez, Vicente Wagner, Nils Scheller, Ines Hözlwimmer, Florian R. Londhe, Shubhankar Bruns, Yanik Hingerl, Johannes	5. Abschlussdatum des Vorhabens 30.06.2024
	6. Veröffentlichungsdatum
	7. Form der Publikation
8. Durchführende Institution(en) (Name, Adresse) Technische Universität München School of Computation, Information and Technology Boltzmannstraße 3 85748 Garching	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 01KU2016B
	11. Seitenzahl
12. Fördernde Institution (Name, Adresse) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben
	14. Tabellen
	15. Abbildungen
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum)	
19. Schlagwörter Regulatorische Genomik, Variantenannotation, Transkriptomik	
20. Verlag	21. Preis

I. Kurze Darstellung

1. Aufgabenstellung und Stand der Wissenschaft und Technik

Mitochondriale Erkrankungen (MDs) sind erbliche Stoffwechselstörungen, die sich durch eine große klinische und genetische Heterogenität auszeichnen. Sie betreffen mindestens 1 von 5.000 Personen und führen häufig zu einer fortschreitenden körperlichen Beeinträchtigung und vorzeitigem Tod. Eine Heilung ist bislang nicht möglich. Allerdings sind etwa 40 bekannte Subtypen mit personalisierten, meist kofaktorbasierten Therapien behandelbar.

Die Einführung von Genom- bzw. Exomsequenzierung, hat die genetische Diagnoseraten erheblich verbessert – dennoch bleiben etwa 50 % der Betroffenen nach der Testung weiterhin undiagnostiziert. Dies ist teilweise auf Einschränkungen bei der Erkennung von nicht-kodierenden oder regulatorischen Varianten zurückzuführen. Dieses Projekt zielt darauf ab, diese Lücke zu schließen, indem Multi-Omics-Technologien (wie Ganzgenomsequenzierung [WGS], RNA-Sequenzierung und Proteomik) integriert werden, um die Diagnostik zu verbessern und eine präzisere Medizin zu ermöglichen.

Bereits vor Beginn dieses Projekts hatten die beteiligten Institutionen Pionierarbeit bei der funktionellen Validierung genetischer Befunde bei MDs geleistet. Das Projekt baut auf der E-Rare-Initiative GENOMIT auf und umfasst ein globales Netzwerk von Expertenzentren aus Deutschland, Österreich, Italien, Frankreich und China.

2. Ablauf des Vorhabens

Das Projekt „Personalisierte mitochondriale Medizin (PerMiM)“ war in drei zentrale Arbeitspakete (APs) gegliedert (Abb. 1):

- **AP1** konzentrierte sich auf die Entwicklung einer Multi-Omics-Diagnostikpipeline über mehrere Zentren hinweg. Klinische Daten sowie Daten auf Genom-, RNA- und Proteinebene wurden integriert, um krankheitsverursachende Varianten zu identifizieren und die diagnostische Genauigkeit zu erhöhen. Neben neuen Algorithmen wurde eine sichere, institutionsübergreifende Analyseplattform entwickelt.
- **AP2** testete therapeutische Strategien anhand kultivierter Patienten-Zelllinien. Die Effekte von Kofaktor- und Multivitamin-Supplementierung wurden analysiert und zeigten funktionelle Verbesserungen in einem Teil der Zelllinien. RNA-Daten wiesen sowohl auf positive metabolische Veränderungen als auch auf nachteilige zelluläre Effekte hin.
- **AP3** verglich diagnostische und therapeutische Strategien an den Partnerstandorten. Ein einheitlicher diagnostischer Algorithmus für Verdachtsfälle von MD wurde etabliert. Studien zur psychosozialen Auswirkung der Diagnose in Österreich führten zu verbesserten Unterstützungsangeboten für Patient:innen. Trotz Einschränkungen bei multizentrischen Studien durch die COVID-19-Pandemie wurden alternative Therapien wie ketogene Diäten und Aminosäure-Supplementierungen untersucht.

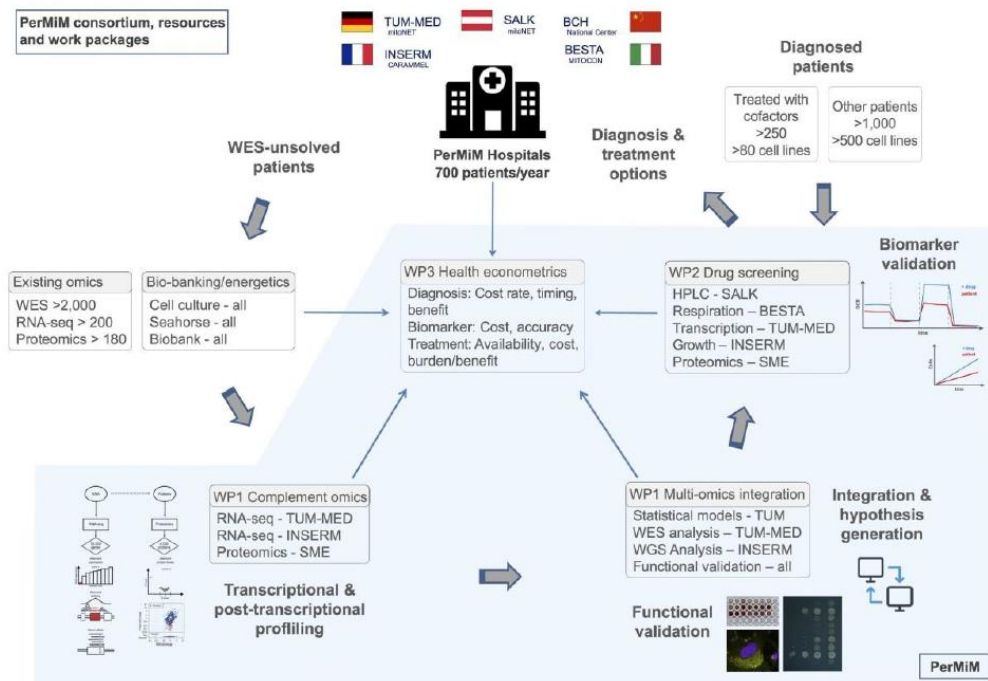


Abbildung 1 | Überblick des PerMiM Projekts

3. Wesentliche Ergebnisse und Zusammenarbeit

Im Arbeitspaket 1 (AP1) unterstützten wir genetische Diagnosen großer Patientenkohorten (Lee et al., 2022; Yépez et al., 2022) und identifizierten neue Krankheitsgene, etwa *LETM1* bei ZNS-Störungen (Kaiyrzhanov et al., 2022), *ATP5F1B* bei Dystonie (Nasca et al., 2023) sowie *PEX1*-Mutationen bei Netzhauterkrankungen (Muñoz-Pujol et al., 2022a). Zudem entwickelten wir bioinformatische Methoden zur Erkennung gestörter Genfunktionen und zur Integration von Multi-Omics-Daten (Wagner et al., 2023; Yépez et al., 2023).

In AP2 testete das Konsortium Nahrungsergänzungen in Zelllinien. Idebenon zeigte keine Wirkung bei LHON-Zellen (Baglivo et al., 2023), ein häufiger *NQO1*-Polymorphismus beeinflusst jedoch die Behandlung (Aleo et al., 2024). 12 % der Patienten-Zellen profitierten von einem Multivitamin-Cocktail (MC); eine zehnfache Dosis wirkte sich negativ aus. Im AP2 war unser Team lediglich in begrenztem Umfang beteiligt.

AP3 führte zur Entwicklung eines diagnostischen Algorithmus für mitochondriale Erkrankungen (Wortmann et al., 2022), zur Evaluation psychosozialer Aspekte (Heath et al., 2024) und zur Definition neuer Biomarker für Diagnostik und Therapie (Aleo et al., 2024). Ergänzende Therapien wie ketogene Diät oder Aminosäuregaben wurden geprüft (Bölsterli et al., 2022; Oswald et al., 2023). Unser Team hat AP3 in den Bereichen statistische Auswertung, Datenanalyse und Visualisierung fachlich unterstützt.

Internationale Zusammenarbeit war von zentraler Bedeutung: Kliniker:innen, Bioinformatiker:innen und Wissenschaftler:innen aus verschiedenen Ländern trugen mit ihrer Expertise in Diagnostik, Softwareentwicklung und Zellbiologie zum Projekt bei. Weitere Kooperationen bestanden mit Francesco Paolo Casale (Helmholtz Zentrum München), Oliver

Stegle (Deutsches Krebsforschungszentrum, DKFZ), Frederic Tort und Antonia Ribes (Hospital Clínic de Barcelona), Brian Chung (University of Hong Kong), Kei Murayama (Chiba Children's Hospital), Christina Ludwig (Technische Universität München).

II. Eingehende Darstellung

1. Ergebnisse

1a. Arbeitspaket 1

Detektion von aberrantem Spleißen

Die Erkennung abweichend gespleißter Gene ist ein wichtiger Schritt in der RNA-Seq-basierten Diagnostik seltener Krankheiten. FRASER ist eine von uns zuvor entwickelte Methode, die auf einem Denoising-Autoencoder basiert und alternative Methoden zur Erkennung von abweichendem Spleißen übertrifft (Mertes et al., 2021). Im Rahmen von PerMIM haben wir FRASER in mehreren Aspekten verbessert. Erstens: Da die drei Spleiß-Metriken von FRASER teilweise redundant sind und dazu neigen, empfindlich auf die Sequenzierungstiefe zu reagieren, führen wir hier eine robustere Intron-Exzisions-Metrik ein, den Intron Jaccard Index, der das alternative Donor-, das alternative Akzeptor- und das Intron-Retentions-Signal zu einem einzigen Wert kombiniert. Zweitens haben wir die Modellparameter und Filter-Schwellenwerte optimiert, indem wir seltene Varianten, die potenziell Spleiß-Störungen verursachen, als unabhängige Evidenz herangezogen haben. Auf 16.213 GTEX-Proben identifizierte unser verbesserter Algorithmus, FRASER 2.0, typischerweise zehnmal weniger Spleiß-Ausreißer und erhöhte dabei den Anteil an seltenen, Spleiß-störenden Varianten um das Zehnfache. Gleichzeitig wurde der Einfluss der Sequenzierertiefe auf die Anzahl der detektierten Ausreißer deutlich reduziert (Abb. 2). Um den Aufwand für die Korrektur von Mehrfachtests zu verringern, führen wir eine Option ein, mit der die zu testenden Gene für jede Probe anstelle eines transkriptomweiten Ansatzes ausgewählt werden können. Diese Option kann besonders nützlich sein, wenn Vorinformationen wie potenzielle Varianten oder Gene vorliegen. Die Anwendung auf 303 Proben von seltenen Krankheitsfällen bestätigte die relative Reduktion der Anzahl an Ausreißer-Detektionen bei einem minimalen Sensitivitätsverlust; FRASER 2.0 identifizierte 22 von 26 zuvor identifizierten pathogenen Spleiß-Fällen mit Standard-Schwellenwerten und 24, wenn die Korrektur durch Mehrfachtests auf OMIM-Gene mit seltenen Varianten beschränkt wurde. Insgesamt tragen diese methodischen Verbesserungen zu einer effektiveren RNA-Seq-basierten Diagnostik seltener Krankheiten bei, indem die Anzahl der Spleiß-Ausreißer pro Probe drastisch reduziert wird, ohne signifikante Einbußen bei der Sensitivität (Scheller et al., 2023).

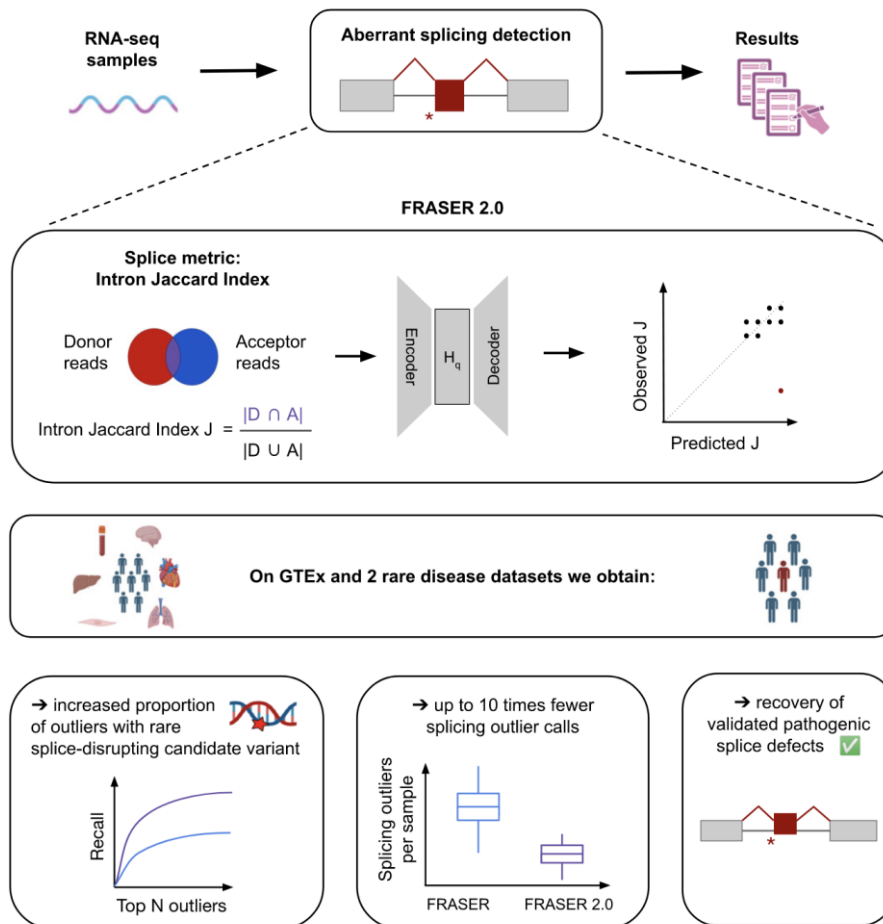


Abbildung 2 | Überblick über FRASER2.0. Wir haben eine neue Spleiß-Metrik eingeführt: den Intron Jaccard Index. Wir haben das gleiche Autoencoder-Prinzip wie in FRASER angewandt, um ihn zu modellieren. Die Methode wurde mit dem GTEx-Datensatz und zwei Kohorten seltener Krankheiten evaluiert. Insgesamt bietet die neue Methode, FRASER2.0, eine höhere Anreicherung von spleißstörenden Varianten und weniger Spleiß-Ausreißer als andere Methoden.

Multi-Omics-Ansatz zur Steigerung der Diagnoseraten im Vergleich zur DNA-Sequenzierung

Die RNA-Häufigkeit und das Spleissen sind nur eine Ebene der Genregulation. Auch die Häufigkeit von Proteinen spielt eine Rolle. In einer bahnbrechenden Studie haben wir die Wirksamkeit der Kombination von Genomik, Transkriptomik und - zum ersten Mal - Proteomik und phänotypischen Deskriptoren in einem systematischen diagnostischen Ansatz bewertet, um die genetische Ursache mitochondrialer Erkrankungen zu ermitteln (Abbildung 3, (Kopajtich et al., 2021)). An Fibroblasten-Zelllinien von 145 Personen wurden durch Tandem-Mass-Tag markierte Proteomik etwa 8.000 Proteine pro Probe nachgewiesen und über 50% aller mit Mendelschen Krankheiten assoziierten Gene erfasst. Die Analyse der differentiellen Proteinexpression ermöglichte die Validierung von Varianten, die zur Destabilisierung von Proteinen führen, und lieferte darüber hinaus unabhängige, ergänzende funktionelle Beweise für Varianten, die zu einer abweichenden RNA-Expression führen. Insgesamt führte unser integrativer rechnergestützter Arbeitsablauf zur genetischen Auflösung von 22% der 121

genetisch ungelösten negativen Fälle mit Gesamtexom oder Gesamtgenom und zur Entdeckung von zwei neuen Krankheitsgenen. Mit der zunehmenden Demokratisierung von Hochdurchsatz-omics-Assays bieten unser Ansatz und unser Code eine Blaupause für die Implementierung von Multi-omics-basierter Diagnostik von Mendelschen Krankheiten in der klinischen Routinepraxis.

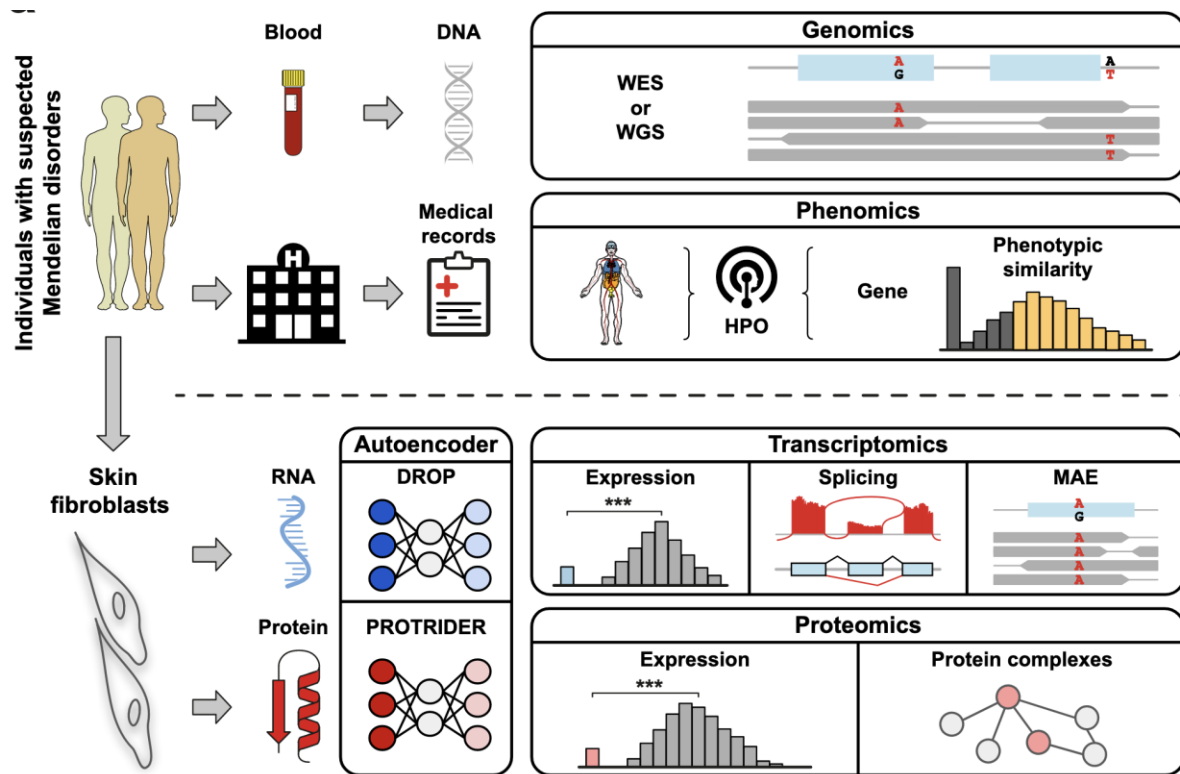


Abbildung 3 | Überblick über einen Multi-omics-Ansatz, der DNA, RNA, Proteomics und Phänotyp integriert, um die Diagnoserate von Patienten mit mitochondrialen Erkrankungen zu erhöhen.

DROP: eine integrative computergestützte Pipeline für die RNA-seq-basierte Diagnostik seltener Krankheiten

Vor PerMIM hatten wir den Prototyp eines integrativen Arbeitsablaufs namens DROP entwickelt, der abweichende Expressionsniveaus, abweichendes Spleißen und monoallelische Expression in RNA-seq-Daten mit speziellen statistischen Methoden kombiniert. DROP ist ein modularer rechnergestützter Arbeitsablauf, der alle Analyseschritte integriert, parallele Berechnungsinfrastrukturen nutzen kann und durchsuchbare Webseitenberichte erzeugt. Im Rahmen von PerMIM haben wir die erste Version von DROP fertiggestellt und veröffentlicht und pflegen und verbessern sie weiter (Yépez et al., 2021). Wir haben ein Begleitprotokoll veröffentlicht, das beschreibt, wie man Diagramme zur Qualitätskontrolle erstellt und bewertet und die Analyseergebnisse interpretiert (Abb. 4). DROP wurde von mehreren Gruppen weltweit

angewandt, wie aus der Anzahl der Zitate (~100) und der Aktivität des Git-Repositorys <https://github.com/gagneurlab/drop> hervorgeht.

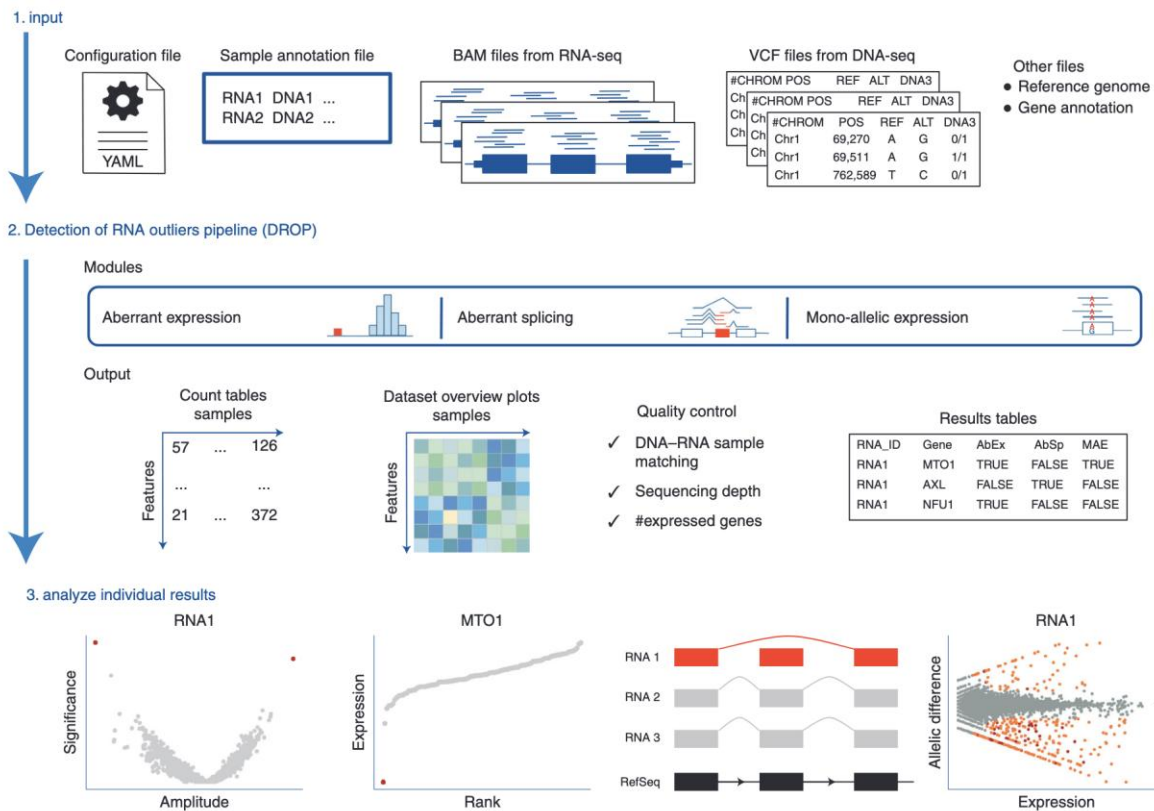


Abbildung 4 | Pipeline zur Detektion von RNA-Ausreißern - DROP. Die Eingabe besteht aus Sequenzierungsrohdaten und Metadaten. DROP kann auf modulare Weise aberrante Expression, Spleißen und monoallelische Expression erkennen. Die Rohdaten werden vorverarbeitet und die Ergebnisse werden für jedes Modul in Form von Tabellen und Diagrammen erstellt.

Klinische Implementierung von RNA-Sequenzierung

Nachdem wir DROP in einem öffentlich zugänglichen Datensatz etabliert hatten, wollten wir es auf eine große PerMIM-Kohorte anwenden. Wir implementierten ein automatisiertes RNA-seq-Protokoll und einen rechnergestützten Arbeitsablauf, mit dem wir Hautfibroblasten von 303 Personen mit Verdacht auf eine mitochondriale Erkrankung analysierten, die zuvor einer Ganz-Exom-Sequenzierung unterzogen worden waren (Abbildung 5, (Yépez et al., 2022)). Außerdem haben wir anhand von Simulationen untersucht, wie abweichende Expression und monoallelische Expressionstests von der RNA-seq-Abdeckung abhängen. Wir konnten durchschnittlich 12.500 Gene pro Probe nachweisen, darunter etwa 60% aller Krankheitsgene - eine wesentlich höhere Abdeckung als bei Vollblut, was die Verwendung von Hautbiopsien unterstützt. Wir priorisierten Gene, die eine abweichende Expression, ein abweichendes Spleißen oder eine monoallelische Expression aufwiesen. Die Pipeline benötigte weniger als eine Woche von der Probenvorbereitung bis zur Ergebnismeldung und lieferte im Median acht krankheitsassoziierte Gene pro Patient zur Untersuchung. Bei 16% der 205 WES-inklusiven Fälle wurde eine genetische Diagnose gestellt. Der Nachweis einer aberranten Expression trug wesentlich zur

Diagnose bei, einschließlich der Fälle, in denen die Expression um 50% reduziert war, was zusammen mit der monoallelischen Expression die Diagnose dominanter, durch Haploinsuffizienz verursachter Störungen ermöglichte. Darüber hinaus ermöglichte das Aufrufen von aberrantem Spleißen und Varianten aus RNA-seq-Daten den Nachweis und die Validierung von spleißstörenden Varianten, von denen die meisten außerhalb der von WES erfassten Regionen lagen. Zusammengenommen zeigen diese Ergebnisse, dass rationalisierte experimentelle und computergestützte Prozesse die Implementierung von RNA-seq in der Routinediagnostik beschleunigen können.

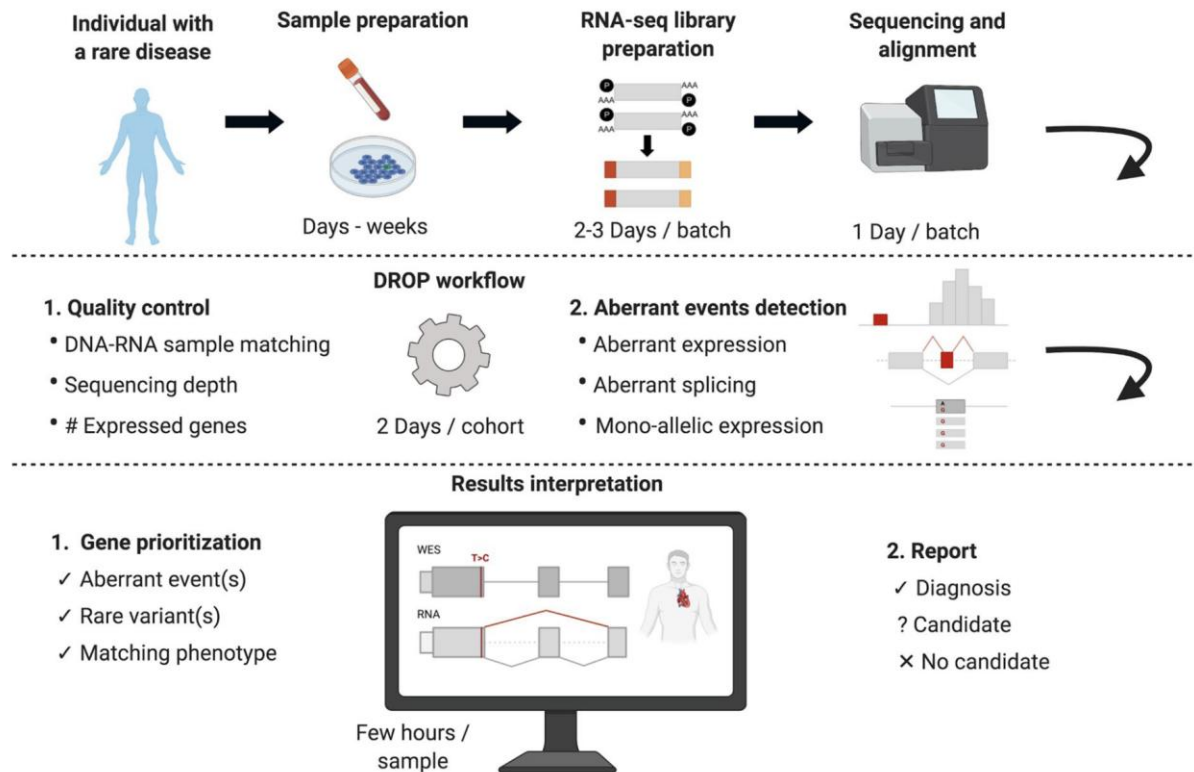


Abbildung 5 | Klinische Umsetzung von RNA-seq-Daten mit DROP. RNA wird von Personen, die an einer seltenen Krankheit leiden, extrahiert und sequenziert. DROP wird ausgeführt, einschließlich relevanter Schritte zur Qualitätskontrolle, gefolgt von der Erkennung von Expressionsausreißern. Die RNA-Ergebnisse werden dann mit der DNA und dem Phänotyp integriert, um Kandidatengene zu erhalten und die Personen zu diagnostizieren.

Vorhersage von aberranter Geneexpression und aberrantem Spleißen - Anwendung auf PerMIM Kohorte mitochondrialer Erkrankungen

In Kombination mit einer ergänzenden Finanzierung aus einem anderen BMBF-Projekt (MERGE - 031L0174A) haben wir verschiedene Methoden zur Vorhersage von Varianteneffekten und Auswirkungen auf den Phänotyp entwickelt. Mit der Förderung durch PerMiM haben wir diese in Datensätzen zu seltenen Krankheiten evaluiert (z. B. mitochondriale Erkrankungen, ALS, UDN und Solve-RD). Zusammengefasst haben wir FRASER, den Detektor von aberrantem Spleißen, auf den GTEx-Datensatz angewendet, um einen einzigartigen Benchmark-Datensatz zu

erstellen, der über 8,8 Millionen seltene Varianten in 49 menschlichen Geweben umfasst (Abb. 6). Wir konnten zeigen, dass moderne sequenzbasierte Vorhersagemodelle derzeit nur eine begrenzte Leistung aufweisen, mit einer Präzision von nicht mehr als 12% bei einer Sensitivität von 20% (Abb. 6). Durch die Erfassung und Quantifizierung der gewebespezifischen Nutzung von Spleißstellen im gesamten Transkriptom sowie die Modellierung der Konkurrenz zwischen alternativen Transkript-Isoformen konnten wir die Präzision bei gleicher Sensitivität verdreifachen. Die Integration von RNA-Sequenzierungsdaten klinisch zugänglicher Gewebe steigerte die Präzision auf 60%. Im Rahmen von PerMiM haben wir dieses Modell, AbSplice, auf einen Datensatz zu mitochondrialen Erkrankungen angewandt und damit seine Robustheit und breite Anwendbarkeit demonstriert (Wagner et al., 2023).

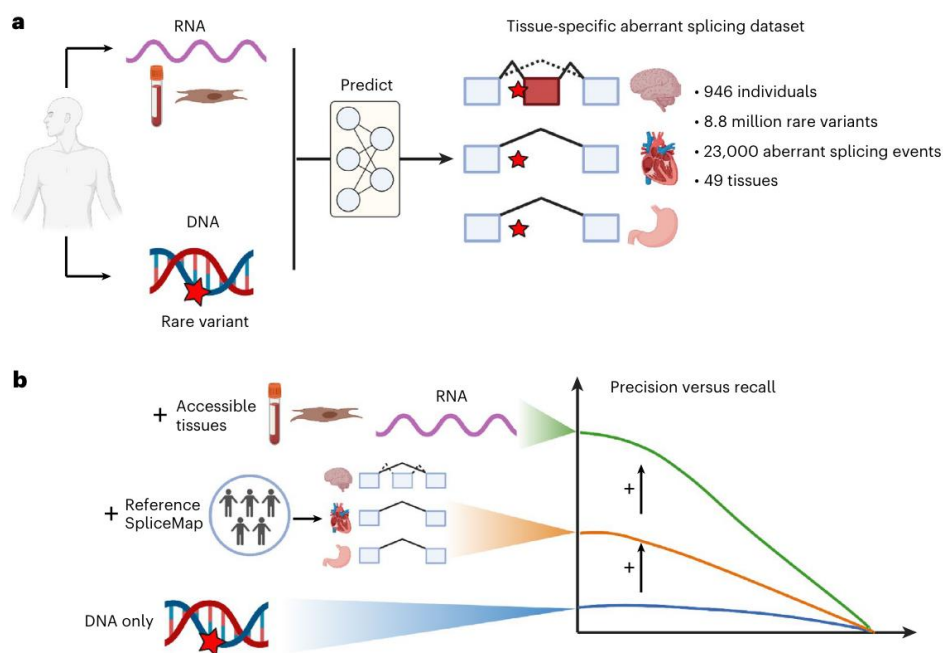


Abbildung 6 | Studiendesign und Take-homes von AbSplice.

Ergänzend zum abweichenden Spleißen ist die abweichende Expression kritischer Gene eine weitere wichtige Ursache für seltene Krankheiten. Varianten, die ein abweichendes Spleißen verursachen, stimmen nicht unbedingt mit Varianten überein, die eine abweichende Expression verursachen. Daher werden ergänzende Werkzeuge zu AbSplice benötigt, um Varianten zu identifizieren, die eine abweichende Expression verursachen. Um diesem Bedarf gerecht zu werden, haben wir AbExp entwickelt, ein Modell, das die abweichende Expression für 49 erwachsene menschliche Gewebe auf der Grundlage des Genotyps und, falls verfügbar, der RNA-seq-Daten eines verfügbaren Gewebes vorhersagt. Ähnlich wie bei AbSplice haben wir einen Benchmark für die Vorhersage aberranter Expression zusammengestellt, der 8,2 Millionen seltene Varianten von 633 Personen in 48 Geweben umfasst (Abb. 7). Bestehende Tools zeigten eine geringe Leistung bei der Vorhersage gewebespezifischer abweichender Expression (1-1,5%

durchschnittliche Präzision) für den Schädlichkeitsscore CADD (Rentzsch et al., 2019) und den Loss-of-Function-Predictor LOFTEE (Karczewski et al., 2020). Anschließend erstellten wir ein Modell, das eine durchschnittliche Präzision von 20% erreicht, indem wir diese und andere Varianten-Annotationsscores sowie die abweichende Expression aus klinisch zugänglichen Geweben integrierten. Im Rahmen von PerMIM replizierten wir diese Ergebnisse, indem wir sie auf eine Kohorte mitochondrialer Erkrankungen anwandten. Die Studie ist als Preprint verfügbar (Hözlzimmer et al., 2023).

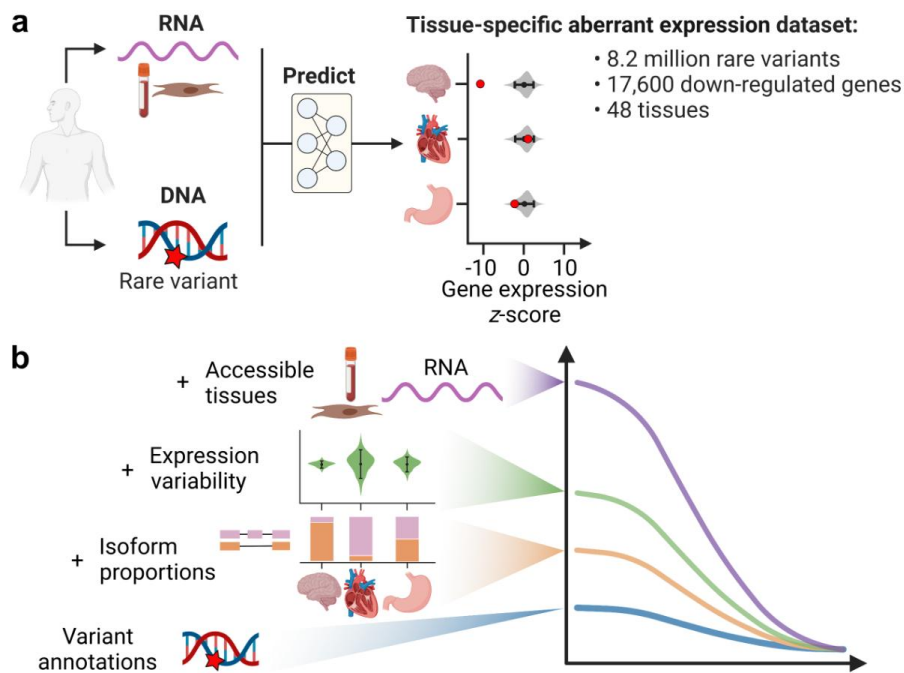


Abbildung 7 | Studiendesign und Take-homes von AbExp.

Mit Unterstützung eines anderen Projekts (MERGE) haben wir das Potenzial alternativer funktionaler Gen-Repräsentationen für Anwendungen des maschinellen Lernens in der Humangenetik untersucht. Beim maschinellen Lernen sind Repräsentationen Vektordarstellungen von Objekten, mit denen Algorithmen arbeiten können. Wir haben herausgefunden, dass funktionale Gen-Repräsentationen, die aus dem Textmining wissenschaftlicher Literatur und aus Studien mit wenigen oder einzelnen Genen abgeleitet wurden, unter dem „streetlight bias“ leiden., wodurch sie genetische Signale aus selten untersuchten Genen in GWAS schlechter erfassen. Dies steht im Gegensatz zu funktionalen Gen-Repräsentationen, die auf systematischen genomweiten Experimenten, wie Transkriptom-Profiling, beruhen (Abbildung 8, (Brechtmann et al., 2023)).

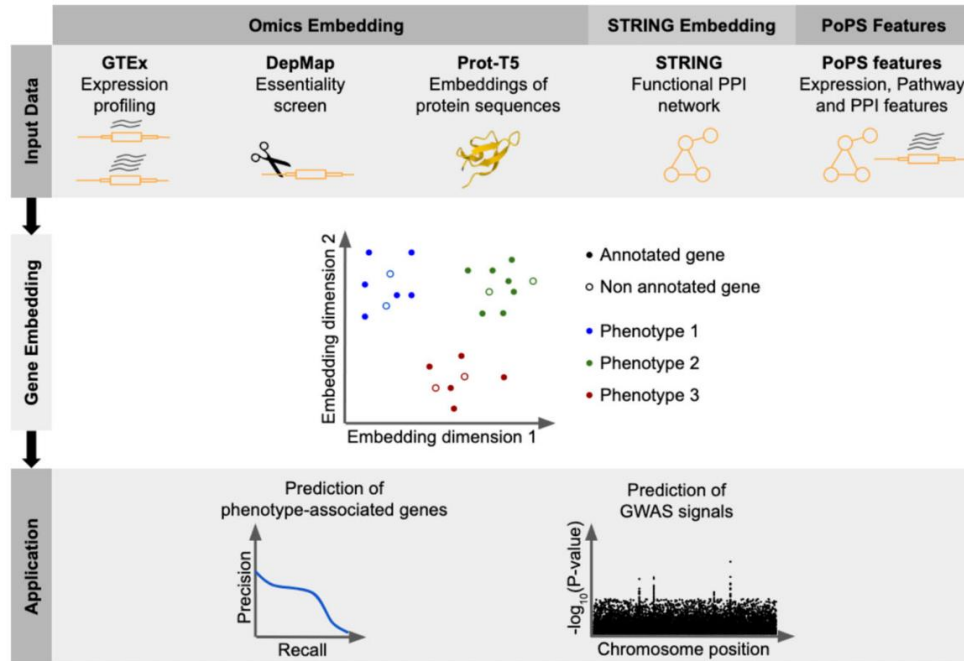


Abbildung 8 | Übersicht über die Studie zu funktionalen Genrepräsentationen.

Seltene genetische Varianten können starke effekte auf Phänotypen haben, was bei der Entdeckung von Effektorgenen und potenziellen Arzneimittel-targets helfen kann. Allerdings stehen Assoziationstests für seltene Varianten vor Herausforderungen durch Datenknappheit und die notwendige Korrektur für multiples testen. Aufbauend auf unseren funktionalen Gen-Repräsentationen und DeepRVAT (Clarke et al., 2024) einer Methode für Assoziationstests für seltene Varianten, die wir in Zusammenarbeit mit der Gruppe von Oliver Stegle (Deutsches Krebszentrum, Heidelberg) entwickelt haben, adressieren wir diese Herausforderungen mit einer neuen Methode, FuncRVP. FuncRVP ist ein Bayessches neuronales Netzwerk, das Informationen über die Funktion von Genen nutzt, um Phänotypen aus seltenen Varianten vorherzusagen (Abb. 9). FuncRVAT lernt einen Prior über die Auswirkungen von Genbeeinträchtigungen anhand von mehrdimensionalen Gen-Repräsentationen aus genomweiten Tests. Unter Verwendung von DeepRVAT trainierten wir FuncRVP auf 41 quantitativen Merkmalen basierend auf 161.850 Exomen der UK Biobank. FuncRVP erreichte in einem unabhängigen Testdatensatz eine vergleichbare oder bessere Leistung als bestehende Methoden zur Phänotypvorhersage und übertraf bei 11 Merkmalen traditionelle Algorithmen zur Phänotypvorhersage signifikant. Darüber hinaus kann FuncRVP durch die Analyse der posterioren Verteilung der Effektgrößen gezielt Gene identifizieren, die mit einem Merkmal assoziiert sind. Um die Genauigkeit der geschätzten Gen-Effekte zu bewerten, nutzten wir als Referenz die Effekte, die mithilfe klassischer Regressionsmethoden in 209.492 weiteren, nicht verwandten Individuen berechnet wurden. Bemerkenswerterweise korrelierten die mit FuncRVP in der Entdeckungs-Kohorte geschätzten Gen-Effekte besser mit dieser Referenz als die Effekte, die direkt durch Regression ermittelt wurden – selbst bei einer Einschränkung der Analyse auf Gene, die in klassischen Assoziationstests signifikant waren. Wir stellten fest, dass FuncRVP auch moderate bis kleine Effektgrößen erfasst, die von traditionellen Assoziationstests häufig übersehen werden. Die mit FuncRVP neu identifizierten Gene zeigen zudem eine signifikant

geringere Toleranz gegenüber Inaktivierung im Vergleich zu Genen, die durch klassische Assoziationstests entdeckt wurden. Während FuncRVP bekannte Assoziationen erfolgreich reproduziert, identifiziert es zusätzlich neue Gen-Assoziationen mit kardiovaskulären Merkmalen, die in früheren Studien zu seltenen Varianten unentdeckt blieben, jedoch experimentell validiert sind, wie etwa die Assoziation zwischen *ARHGAP15* und dem Lymphozytenanteil (Londhe et al., 2024).

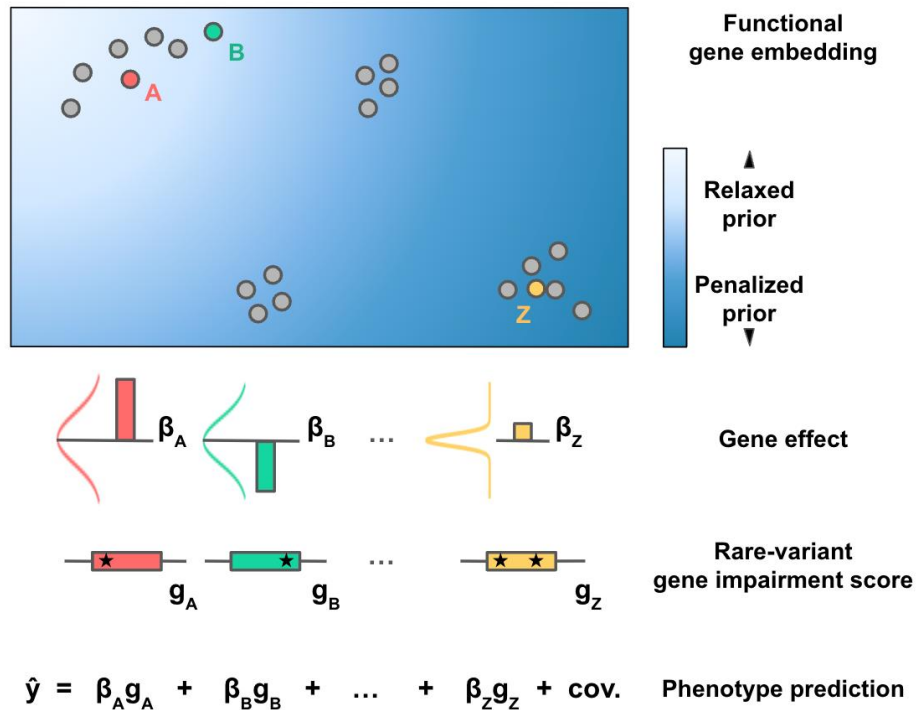


Abbildung 9 | FuncRVP-Modellübersicht. FuncRVP ist ein Bayessches Modell, das Phänotypen basierend auf seltenen Varianten vorhersagt. Der Phänotyp eines Individuums wird als gewichtete Summe der aus seltenen Varianten resultierenden Gen Scores über alle Gene modelliert (unten). Eine aus den Daten gelernte Prior-Verteilung begrenzt die Stärke der Gen-Effekte in Abhängigkeit von ihrer Position in einer funktionalen Gen-Repräsentation. Dadurch wird die Regularisierung in Bereichen gelockert, die dicht mit merkmalsassoziierten Genen besiedelt sind (Gene A-B, oben).

Priorisierungswert für Krankheitsgene

Mit unseren Modellen zur Detektion abweichender Genexpression und Priorisierung von Varianten nahmen wir an der 6. Runde der CAGI "SickKids Clinical Genomes and Transcriptomes"-Challenge teil. Ziel der Challenge war es, das diagnostische Potenzial von Multi-Omics-Ansätzen bei der Identifikation und Klärung unerkannter genetischer Erkrankungen anhand klinischer Daten von 79 Kindern mit vermuteten Mendel'schen Störungen zu evaluieren. Wir setzten ein maschinelles Lernmodell ein, das auf einer Kohorte von 93 gelösten Fällen mitochondrialer Erkrankungen trainiert wurde, um Kandidatengene zu priorisieren (Abb. 10). In unserer Analyse der SickKids-Kohorte konnten wir die kausalen Gene in 2 von 3 diagnostizierten Fällen, die Auffälligkeiten auf RNA-Seq-Ebene zeigten, erfolgreich priorisieren sowie in 6 von 12

Fällen, in denen keine RNA-Auswirkungen erkennbar waren. Damit gehörte unsere Lösung zu den Gewinnern der Challenge. Die Challenge und unser Ansatz unterstreichen den unschätzbaren Wert einer integrativen Analyse genetischer, transkriptomischer und klinischer Daten, um krankheitsverursachende Gene zu identifizieren. Die Challenge wurde anhand von drei zuvor diagnostizierten Fällen bewertet, in denen RNA-Seq-Daten für die Diagnostik hilfreich waren, sowie zwölf weiteren Fällen, die ausschließlich durch DNA-Analysen diagnostiziert wurden. Einige dieser Fälle wurden später von Deshwar et al. gemeldet. Unser Modell konnte 2 der 3 RNA-Seq-unterstützten Fälle unter den Top-3-Rängen priorisieren und erzielte eine Recall-Rate von über 50% innerhalb der Top-100-Gene über alle 15 Fälle hinweg (Yépez et al., 2023).

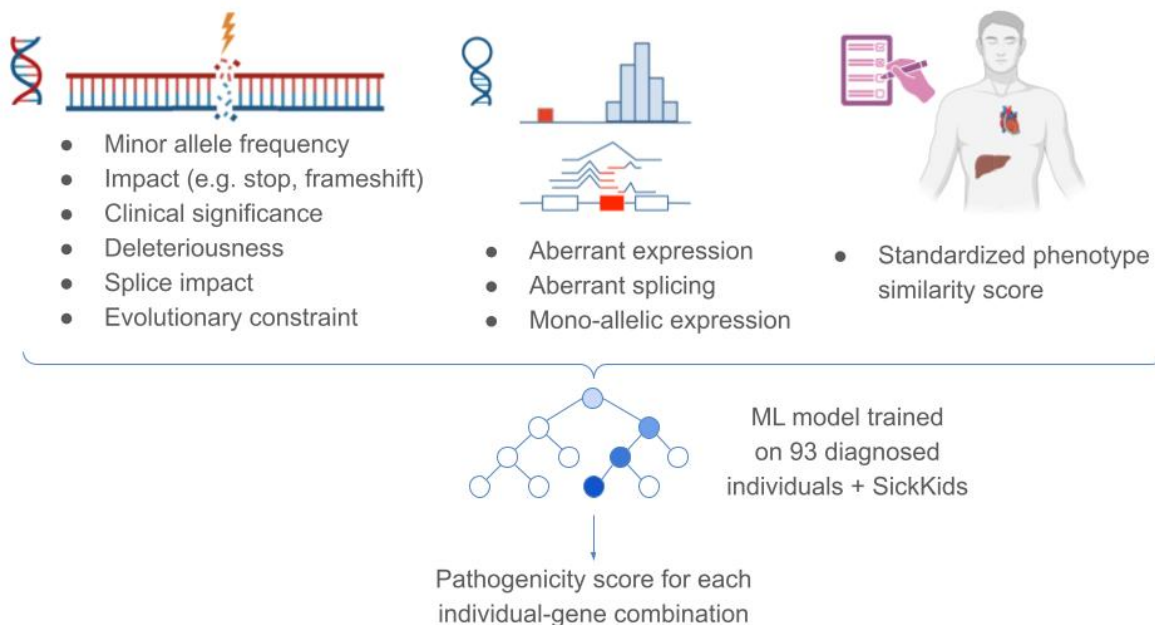


Abbildung 10 | Integratives Vorhersagemodell, das Variantenannotationen, abweichende Genexpression und standardisierte Phänotyp-Scores kombiniert, um einen einzigen Score für jede Kombination aus Individuum und Gen zu erzeugen.

RNA-seq zur Verbesserung der Diagnosequoten bei verschiedenen Krankheiten

Wir begannen eine enge Zusammenarbeit mit der Gruppe von Frederic Tort am Hospital Clínic de Barcelona, die zur Veröffentlichung von drei Einzel-Gen-Studien führte:

- Eine intronische Variante im ATP6AP1-Gen bei einem betroffenen Individuum und eine intronische Deletion bei einem anderen, nicht verwandten, betroffenen Individuum, die beide zur Bildung eines neuen Exons führten (Morales-Romero et al., 2024).
- Eine intronische Variante im PEX1-Gen, die bei einem betroffenen Individuum zur Bildung eines neuen Exons führte, sowie eine 1-bp-Deletionsmutation, die bei einem anderen, nicht verwandten, betroffenen Individuum zu einer abnormen Expression führte (Muñoz-Pujol et al., 2022a).

- Eine Spleißstellen-Variante im PTC3-Gen bei einem betroffenen Individuum, die zum Überspringen eines Exons und einer abnormen Expression führte (Muñoz-Pujol et al., 2022b).

Zusätzlich starteten wir eine Zusammenarbeit mit der Gruppe von Brian H. Y. Chung von der Universität Hongkong. Dort wurde RNA-seq an Zellen aus Fruchtwasser (FW) durchgeführt, um die Anwendbarkeit in der pränatalen Diagnose zu testen. Die Anzahl der ausreichend exprimierten Gene in FW-Zellen war mit der in Fibroblasten vergleichbar und deutlich höher als in Blutproben über verschiedene Krankheitskategorien hinweg. Wir fanden, dass RNA-seq in FW-Zellen in der pränatalen Diagnose (n = 4) machbar und vorteilhaft ist, da die transcriptomischen Daten die molekularen Konsequenzen aufzeigten, die zur Pathogenitätsaufwertung von Varianten in den Genen CHD7 und COL1A2 führten und die in silico Vorhersage einer Variante im MYRF-Gen überarbeiteten. RNA-seq in FW-Zellen könnte eine vernünftige Wahl für postnatale Patienten werden, da es im Vergleich zu Fibroblasten und Blut Vorteile bietet, da es invasive Verfahren vermeidet. FW-Zellen war mit der in Fibroblasten vergleichbar und deutlich höher als in Blutproben über verschiedene Krankheitskategorien hinweg. Wir fanden, dass RNA-seq in FW-Zellen in der pränatalen Diagnose (n = 4) machbar und vorteilhaft ist, da die transcriptomischen Daten die molekularen Konsequenzen aufzeigten, die zur Pathogenitätsaufwertung von Varianten in den Genen CHD7 und COL1A2 führten und die in silico Vorhersage einer Variante im MYRF-Gen überarbeiteten. RNA-seq in FW-Zellen könnte auch eine sinnvolle Alternative für postnatale Diagnostik sein, da es im Vergleich zu Fibroblasten und Blut Vorteile bietet, invasive Verfahren vermeidet (Lee et al., 2022).

Multizentrische und krankheitsübergreifende Analyse in Solve-RD

Da der Ansatz des föderierten Lernens nicht umgesetzt werden konnte, konzentrierten wir uns auf eine robuste integrative Software und setzten sie in Proben aus mehreren Zentren ein. Wir wurden Teil des europäischen Konsortiums Solve-RD, wo wir den RNA-seq-Datensatz analysierten. Der Solve-RD RNA-seq-Datensatz besteht aus mehr als 450 Proben, die aus 4 Geweben von Personen gesammelt wurden, die in mehr als 30 Kliniken aus 6 europäischen Referenznetzwerken untersucht wurden. Für die überwiegende Mehrheit der Proben liegen entsprechende DNA-Sequenzierungsdaten vor, häufig auch für die Eltern. Wir haben neue Richtlinien, Algorithmen und Methoden entwickelt, um mit dem heterogenen Datensatz umgehen zu können (Abb. 11). Die Proben wurden mit dem Arbeitsablauf DROP analysiert. Zwei neue RNA-seq-Qualitätskontrollmodule wurden zu DROP hinzugefügt, um die Annotationen von Geschlecht und Gewebe zu bestätigen. Beide halfen, eine Handvoll falsch markierter Proben zu erkennen. Um die Überprüfung der Ergebnisse zu standardisieren, haben wir einen konsortiumweiten Annotationsrahmen entwickelt, der die Bewertung eines aberranten Ereignisses auf vier Ebenen integriert: i) Rohdatenunterstützung der Ausreißer-Calls, ii) Genfunktion, iii) Variante(n) mit einem plausiblen Regulationsmechanismus (z. B. Spleißstelle oder NMD) und iv) Segregation der Variante(n). Die automatisch erstellten Berichte wurden so konzipiert, dass sie alle für die Annotatoren erforderlichen Daten übersichtlich darstellen. Um die Konsistenz der Annotationen weiter zu erhöhen, haben wir Präsenzworkshops, den Solvathon, durchgeführt, in denen Schulungen und gemeinsame Dateninspektionen kombiniert wurden, und wöchentliche

Beratungssitzungen angeboten. Die Analyse hat zu 20 gelösten Fällen und >40 starken Kandidaten geführt, darunter ein neues Krankheitsgen. Insgesamt tragen unsere neuen Algorithmen und Prozesse zu verbesserten RNA-seq-basierten Diagnosemethoden für seltene Krankheiten bei.

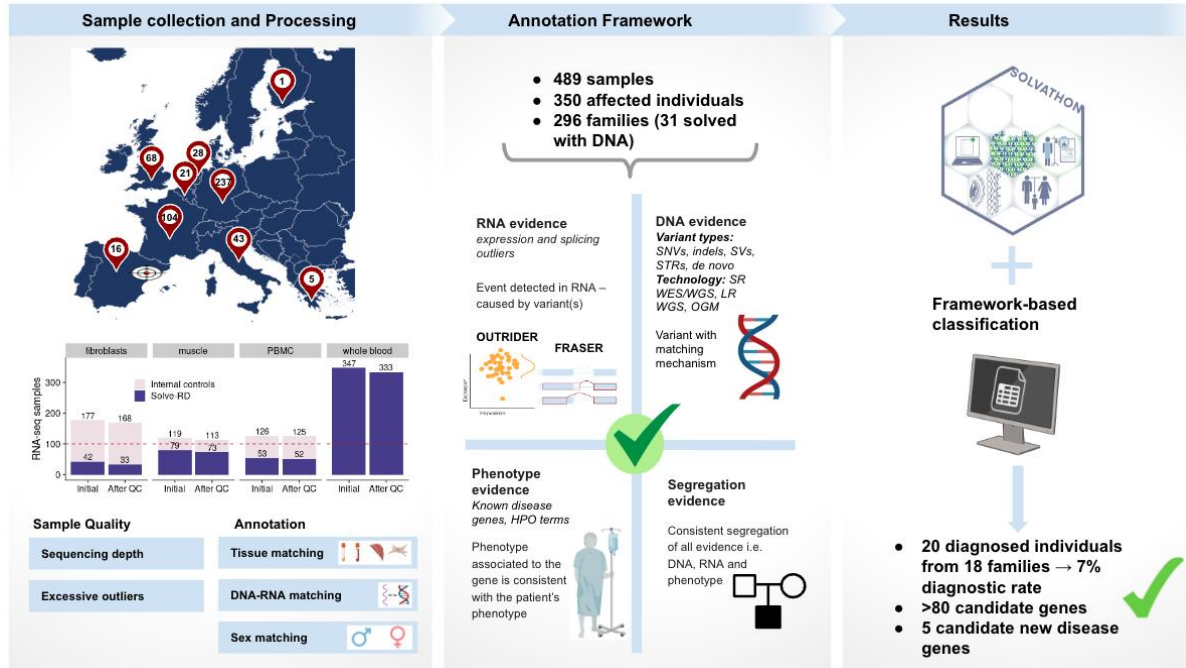


Abbildung 11 | Überblick über das SolveRD RNA-Seq-Projekt, das in 3 Säulen unterteilt ist: 1. Probensammlung, 2. Annotationsrahmen und 3. Analyseergebnisse.

1b. Arbeitspaket 3

Unser Team hat AP3 durch Beratung in der Versuchsplanung sowie bei der Analyse genetischer und multi-omischer Daten unterstützt. Diese Unterstützung erfolgte insbesondere durch Prof. Julien Gagneur und Vicente Yépez im Rahmen regelmäßiger projektweiter Treffen und Retreats.

Im Ergebnis trug diese Zusammenarbeit zu zwei Studien bei:

- In einer Arbeit wurde eine genetische Variante im NQO1-Gen identifiziert, die die Wirksamkeit der Idebenon-Behandlung beeinflusst (Aleo et al., 2024).
- In einer weiteren Studie mit primären Fibroblasten-Zelllinien zeigte sich, dass eine Supplementierung mit Idebenon nicht zwischen Respondern und Non-Respondern unterscheidet und somit keine prädiktive Aussage über die klinische Wirksamkeit der Behandlung zulässt (Baglivo et al., 2023).

2. Positionen

Die Mittel bei der AG Gagneur wurden wie geplant eingesetzt:

Position	Verwendung
Reise	Das Reisebudget wurde für Konferenzen ausgegeben: Genomics of Rare Disease 2022, ASHG 2021 & 2022, ESHG 2022, Solve-RD Tagung 2023, PerMiM Tagungen 2022 & 2024, ISMB 2022 & CAGI6 Tagung 2022. Darüber hinaus wurde das Reisebudget genutzt, um zum Start der Zusammenarbeit den Solve-RD-Partner CNAG in Barcelona zu besuchen.
Personal	Für die Entwicklung und Implementierung der statistischen und bioinformatischen Methoden wurden Doktoranden (ausgebildete Bioinformatiker, Informatiker, Physiker) angestellt. Außerdem haben wir HiWis mit der Optimierung der Software beauftragt. Die Mittel sind bis auf die in Laufzeitverlängerung dargestellten Ausgaben damit aufgebraucht worden.
Verbrauchsmaterial	Das Budget für die RNA-Sequenzierung wurde wie geplant verwendet.
Anschaffung von Geräten	Der geplante Speicherserver wurde zu Beginn der Förderperiode (2020) angeschafft.
Publikationen	Eine Open-Access Publikation

3. Notwendigkeit

Die im Antrag für das Teilprojekt genannten Ziele und definierten Meilensteine wurden erreicht. Die beantragten Mittel für Reisen, Personal und den Workshop wurden vollständig benötigt und entsprechend eingesetzt.

4. Nutzen

Wirtschaftliche Erfolgsaussichten		Während der Laufzeit des Vorhabens	Im Anschluss an die Laufzeit
Ziel	Erläuterungen	<i>bitte zutreffende Felder ankreuzen</i>	<i>bitte zutreffende Felder ankreuzen</i>

Volkswirtschaftliche Verwertung	Die Sequenzvorhersagewerte können von Unternehmen verwendet werden, die genetische Tests anbieten, wie beispielsweise SOPHiA Genetics.	-	x
---------------------------------	--	---	---

Wissenschaftliche und/oder technische Erfolgsaussichten		Während der Laufzeit des Vorhabens	Im Anschluss an die Laufzeit
Ziel	Erläuterungen	<i>bitte zutreffende Felder ankreuzen</i>	<i>bitte zutreffende Felder ankreuzen</i>
Verbreitung der Erkenntnisse	Veröffentlichung der erzielten Ergebnisse in wissenschaftlichen Fachzeitschriften und für die breite Öffentlichkeit, Meldung bei Studienregistern, Vorträge und Poster bei Fachkongressen oder anderen Veranstaltungen	x	x
Aus-, Weiter-, Fortbildung	Vorhaben dient der Heranbildung des wissenschaftlichen Nachwuchses, durch die Erstellung von Dissertationen, die Ergebnisse finden Eingang in die Lehre	x	x
Forschungsstrukturen	Entwicklung und Instandhaltung robuster Software zur Vorhersage abweichender Transkriptionsereignisse.	x	x

Wissenschaftliche und wirtschaftliche Anschlussfähigkeit		Im Anschluss an das Vorhaben
Ziel	Erläuterungen	<i>Zeithorizont erläutern</i>

Wissenschaftliche, technische, strukturelle und versorgungsbezogene Verwertungsmöglichkeiten	Die in diesem Projekt erzielten Fortschritte bilden die Basis für weitere Forschungsprojekte und Anträge: GHGA (https://www.ghga.de/) zweite Förderperiode, grant, ERDERA (https://www.ejprarediseases.org/erdera/). Darüber hinaus legte die Arbeit an der groß angelegten Sprachmodellierung für mehrere Spezies erfolgreich die Grundlage für eine neue ERC Synergy Finanzierung namens EPIC.	ERDERA (https://erdera.org/): Beginn 09.2024 GHGA (https://www.ghga.de/) Erneuerung: voraussichtlicher Beginn 10.2025
--	---	---

5. Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Es sind von dritter Seite keine Ergebnisse bekannt geworden, die wesentlichen Einfluss auf die Verwertung der Ergebnisse nehmen.

6. Veröffentlichungen des Ergebnisses

Im Rahmen dieses Projekts wurden folgende Publikationen und Preprints veröffentlicht:

1. Brian Clarke*, Eva Holtkamp*, ..., Felix Brechtmann, Florian R. Hözlzimmer, Julien Gagneur@, Oliver Stegle@. Integration of variant annotations using deep set networks boosts rare variant association genetics. *Nature Genetics*, 2024
2. Shubhankar Londhe, Jonas Lindner, Zhifen Chen, Eva Holtkamp, Florian R. Hözlzimmer, Francesco Paolo Casale, Felix Brechtmann, Julien Gagneur. Functional gene embeddings improve rare variant polygenic risk scores. *bioRxiv*, 2024
3. Blai Morales-Romero, Gerard Muñoz-Pujol, Rafael Artuch, ..., Anna Esteve-Codina, Vicente A. Yépez, Julien Gagneur, Saskia B. Wortmann, Holger Prokisch, Antonia Ribes, Judit García-Villoria@, Frederic Tort@. Genome and RNA sequencing were essential to reveal cryptic intronic variants associated to defective ATP6AP1 mRNA processing. *Molecular Genetics and Metabolism*, 2024, PMID: 38878498
4. Felix Brechtmann, Thibault Bechtler, Shubhankar Londhe, Christian Mertes, and Julien Gagneur. Evaluation of input data modality choices on functional gene embeddings. *NAR Genomics and Bioinformatics*, 2023, PMID: 37942285
5. Florian R. Hözlzimmer, Jonas Lindner, Nils Wagner, Vicente A. Yépez, Francesco Paolo Casale, Julien Gagneur. Aberrant expression prediction across human tissues. *bioRxiv*, 2023
6. Vicente A. Yépez@, Nicholas H. Smith, Ines Scheller, Julien Gagneur, Christian Mertes@. Predicting molecular events underlying rare diseases using variant annotation, aberrant gene expression events, and human phenotype ontology. *Research Square*, 2023
7. Ines Scheller, Karoline Lutz, Christian Mertes, Vicente A. Yépez@, Julien Gagneur@. Improved detection of aberrant splicing and the Intron Jaccard Index. *The American Journal of Human Genetics*, 2023, PMID: 38006880
8. Nils Wagner, Muhammed H. Çelik, Florian R. Hözlzimmer, Christian Mertes, Holger Prokisch, Vicente A. Yépez, Julien Gagneur. Aberrant splicing prediction across human tissues. *Nature Genetics*, 2023, PMID: 37142848

9. Vicente A. Yépez*, Mirjana Gusic*, Robert Kopajtich, Christian Mertes, Nicholas H. Smith, ..., Michaela F. Müller, ..., Julien Gagneur@, Holger Prokisch@. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Medicine*, 2022, PMID: 35379322
10. Mianne Lee*, Anna K. Y. Kwong*, ..., Vicente A. Yépez, Julien Gagneur, Anita S. Y. Kan, Brian H. Y. Chung. Diagnostic potential of the amniotic fluid cells transcriptome in deciphering mendelian disease: a proof-of-concept. *npj Genomic Medicine*, 2022, PMID: 36577754
11. Gerard Muñoz-Pujol, ..., Vicente A. Yépez, Mirjana Gusic, Julien Gagneur, Holger Prokisch, Rafael Artuch, Antonia Ribes, Roser Urreizti, Frederic Tort. Leigh syndrome is the main clinical characteristic of PTCD3 deficiency. *Brain Pathology*, 2022, PMID: 36450274
12. Gerard Muñoz-Pujol, ..., Vicente A. Yépez, Julien Gagneur, Mirjana Gusic, Holger Prokisch, ..., Antonia Ribes. Diagnostic Odyssey in an Adult Patient with Ophthalmologic Abnormalities and Hearing Loss: Contribution of RNA-Seq to the Diagnosis of a PEX1 Deficiency. *International Journal of Molecular Sciences*, 2022, PMID: 36293220
13. Vicente A. Yépez, Christian Mertes, Michaela F. Müller, Daniela S. Andrade, Leonhard Wachutka, Laure Frésard, Mirjana Gusic, Ines Scheller, Patricia F. Goldberg, Holger Prokisch, Julien Gagneur. Detection of aberrant gene expression events in RNA-seq data. *Nature Protocols*, 2021, PMID: 33462443
14. Sarah L Stenton, ..., Julien Gagneur, ..., Vicente A. Yépez, ..., Holger Prokisch. Diagnosing pediatric mitochondrial disease: lessons from 2,000 exomes. *medRxiv*, 2021
15. Robert Kopajtich, ..., Ines Scheller, ..., Christian Mertes, ..., Vicente Yépez, ..., Julien Gagneur, Holger Prokisch. Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv*, 2021
16. Serena J Aleo, ..., Holger Prokisch, ..., Daniele Ghezzi. Genetic variants affecting NQO1 protein levels impact the efficacy of idebenone treatment in Leber hereditary optic neuropathy, *Cell Rep Med*, 2024, PMID: 38272025
17. Morko Baglivo, ..., Holger Prokisch, ..., Anna M Ghelli. Evaluation of Mitochondrial Dysfunction and Idebenone Responsiveness in Fibroblasts from Leber's Hereditary Optic Neuropathy (LHON) Subjects. *Int J Mol Sci*, 2023, PMID: 37628761

München, den 16.12.2024

.....

Prof. Dr. Julien Gagneur