

## FiND

Entwicklung bioinformatischer Tools, um Muster in Sequenzierungsdaten zu finden und die Biomarkerentwicklung zu stützen.

|                                 |                                                                 |                        |               |
|---------------------------------|-----------------------------------------------------------------|------------------------|---------------|
| <b>Programm / Ausschreibung</b> | FORPA, Forschungspartnerschaften NATS/Ö-Fonds, FORPA OEF2020    | <b>Status</b>          | abgeschlossen |
| <b>Projektstart</b>             | 01.04.2022                                                      | <b>Projektende</b>     | 31.03.2025    |
| <b>Zeitraum</b>                 | 2022 - 2025                                                     | <b>Projektlaufzeit</b> | 36 Monate     |
| <b>Keywords</b>                 | Biomarker, Molekulare Diagnostik, Bioinformatik, Programmierung |                        |               |

## Projektbeschreibung

Ausgangssituation: Kurze RNA Moleküle, sogenannte mikroRNAs, haben vier Merkmale die sie besonders für die Verwendung als Biomarker auszeichnen: als Regulatoren der Genexpression tragen sie wesentlich zur Kontrolle biologischer Prozesse und Krankheiten bei. Prozesse zur Freisetzung aus Zellen und hohe Stabilität ermöglichen nicht-invasive Detektion in Blut, Urin, oder Speichel. Hohe Gewebespezifität ermöglicht es systemische Veränderungen im (Blut) mit Vorgängen in bestimmten Zelltypen zu verknüpfen. Die PCR Analyse zur Messung einzelner mikroRNAs ist eine kosteneffiziente Methode mit hoher klinischer Relevanz (siehe SARS-CoV-2 PCR).

Auf Basis dieser Vorteile wurden mehrere klinische Anwendungen von mikroRNA Biomarkern etabliert, zum Beispiel für die Überwachung toxischer Nebeneffekte von Arzneimitteln in klinischen Studien (siehe Empfehlungen von EMA und FDA).

Die Labormethode der Wahl für die Identifikation von neuartiger mikroRNA Biomarkern ist die Hochdurchsatzsequenzierung (NGS). Diese Analyse zeichnet sich gegenüber der PCR-Analyse einerseits durch extrem hohen Informationsgehalt („Big Data“) aus, der für die frühe Phase der Biomarkerforschung entscheidend ist, andererseits aber auch durch höhere Kosten und Ressourceneinsatz, wodurch die Anzahl der zu messenden Proben limitiert ist. Aus diesen Limitationen ergibt sich die Motivation bioinformatische Methoden zu entwickeln um 1) den hohen Informationsgehalt („Big Data“) bestmöglich zu analysieren, und 2) Ressourcen zu schonen indem der Zugriff auf bereits vorhandene Daten vereinfacht und beschleunigt wird.

Ziele: Dieses Projekt verfolgt drei wesentliche Ziele. 1) Die Entwicklung bioinformatischer Analyseverfahren („Pipeline“) von NGS Daten die durch unterschiedliche Module den hohen Informationsgehalt vollständig nutzbar machen (Quantifikation, Isoformen, Sequenzmotive), automatisierbar sind, und die Ergebnisse strukturiert und grafisch aufbereiten. 2) Die Entwicklung einer Datenbank für geordnete Speicherung von NGS Daten und auf die aus der in 1) entwickelten Pipeline zugegriffen werden kann um für prospektive Datensätze Information zur Verfügung zu stellen. 3) Die Überprüfung anhand von experimentellen Datensätzen ob die Verwendung der Pipeline die Effizienz und die Qualität der Biomarkerforschung erhöhen kann. Auf Basis dieser innovativen Methoden erwarten wir uns deutlich tiefere Einblicke in die mikroRNA Muster von gesunden von erkrankten Personen und somit die Qualität und Effizienz der Biomarkerforschung maximal steigern können.

Angestrebte Erkenntnisse: Wir erwarten uns durch die Entwicklung und Anwendung der Methoden und Tools sehr konkrete Antworten auf zwei Forschungsfragen. 1) welche Informationen sind neben der quantitativen Auswertung von NGS Daten noch hilfreich um Biomarker zu identifizieren? 2) kann (analog zu „Biobanken“) die systematische Erfassung und Speicherung historischer NGS Daten („Biodatenbanken“) die Effizienz und Qualität der Biomarkerforschung steigern?

### **Projektpartner**

- TAmiRNA GmbH