

**Titel des Vorhabens: Eine Metagenomik-Softwareplattform für die klinische Infektionsdiagnostik**

**Projektnummer/Aktenzeichen: SAW-2016-HPI-5**

## Executive Summary

Zielsetzung des Antrages war es, mit einem auf Hochdurchsatzsequenzierung (next generation sequencing, NGS) beruhenden Ansatz eine integrierte, skalierbare Software Lösung zum hypothesenfreien Erregernachweis in diagnostischen Laboren zu entwickeln. Obgleich Publikationen zumeist einzelne Fachberichte über die Anwendung von NGS in der Erregerdiagnostik zugenommen haben, so ist diese Methode bis heute aufgrund unzureichender Standards insbesondere zur Validierung, Interpretation und Befundung nicht in die Routine- Erregerdiagnostik integriert. Darüber hinaus fehlt es an Machbarkeit- und Validierungsstudien, die einen klaren klinischen Mehrwert belegen.

Das Vorhaben zielte daher auf die Entwicklung einer integrierten und skalierbaren, klinisch orientierten Softwareplattform ab, die eine schnelle und zuverlässige Analyse von NGS-Daten in diagnostischen Routinelaboratorien ermöglichen soll. Die einfache Bedienung sowie intuitive Darstellung der Analyseergebnisse in einer Form, welche die Interpretation durch klinisches Personal erlaubt, waren dabei wesentliche technischen Herausforderungen des Projektes. Diesem Antrag vorangegangen war bereits die Entwicklung einer bioinformatischen Pipeline für die NGS-basierte, metagenomische Identifizierung von Infektionserregern in der Forschung. Diese sollte speziell für den Einsatz in mikrobiologischen Laboratorien angepasst werden und richtet sich daher insbesondere an in der Klinik oder Forschung tätiges Personal ohne spezielle Vorkenntnisse in der Bioinformatik. Das zentrale Analyse- und Datenspeichermodul der Software besteht aus einem Backend, welches auf spezieller Hardware unter einem Linux-Betriebssystem oder als vorkonfigurierte virtuelle Maschine läuft. Die Hardwareanforderungen für die zentrale Analyseeinheit erlauben eine Datenverarbeitung in einem Zeitrahmen, der den Anforderungen der klinischen Diagnostik entspricht (je nach Komplexität der Proben zwischen 3 und 24 Stunden). Das Backend verwendet Open-Source-Software zur Durchführung der Primäranalysen und steht der Forschungsgemeinschaft kostenlos zur Verfügung. Es kann vollständig über die Kommandozeile bedient werden. Das vollständige Softwarepaket verfügt über ein Frontend mit grafischer Benutzeroberfläche (GUI), das auf Standard-Desktop-Computern ausgeführt werden kann und eine benutzerfreundliche Konfiguration von Analysen, Datenmanagement und Visualisierung von Ergebnissen erlaubt. Das hier anvisierte Softwarepaket dient der rationalisierten Analyse von metagenomischen/metatranskriptomischen NGS Daten für diagnostische Forschungszwecke. Die Arbeitsabläufe umfassen keine personalintensiven oder kontaminationsanfällige Anreicherungsstrategien (z. B. Virusanreicherung oder Capture Sequenzierungen) und sind daher mit den routinemäßigen/automatisierten Extraktionsprotokollen in diagnostischen kompatibel. Das vollautomatische Backend umfasst alle primären Analysen (insbesondere Contig-Assemblierung und taxonomische Klassifizierung) sowie eine Datenbank zur Speicherung und Abruf der Primär- und Analysedaten. Eine Besonderheit des Backends ist die Implementation von *on demand* Funktionen, wie z.B. die Unterstützung von Kohortenanalysen. Dieser Quervergleich einer großen Anzahl von diagnostischen Proben ist bedeutend für die Interpretation von NGS Analysen wie z. B. phylogenetische Analysen in der Untersuchung von Ausbrüchen und Infektionsketten oder auch Identifizierung mutmaßlicher Krankheitserreger, die in einer Probenkohorte vorhanden sind, aber in einer Kontrollkohorte weniger häufig oder gar nicht vorkommen.

Die Software wurde durch Kooperationspartner aus der mikrobiologischen Diagnostik eingehend getestet und auf Grundlage regelmäßiger Feedbacks (monatlicher Jour Fix) zu Leistung, Benutzerfreundlichkeit und Implementierung zusätzlicher Funktionen fortlaufend

optimiert. Weiterhin ist die Software ein wesentliches Werkzeug innerhalb des ENNGS (*ESCV network on next generation sequencing*), eines Netzwerkes von Institutionen aus Mikrobiologie, Virologie und Infektiologie zur Erarbeitung von Richtlinien der Anwendung von NGS in der mikrobiologischen Diagnostik. Das Backend Softwarepaket ist quelloffen frei verfügbar. Ein Frontend zur Visualisierung der Daten (ursprünglich als lizensierbares Produkt geplant) steht Kooperationspartnern zur Verfügung.

## 1. Zielerreichung und Umsetzung der Meilensteine

Im Folgenden sind die ursprünglichen Milestones aufgeführt (blaue Schrift) mit Kommentaren zur Umsetzung in schwarzer Schrift.

- Milestone 1: M1-M8 Entwicklung der Backend-Analysesoftware (basierend auf der bestehenden Plattform) und Implementierung zusätzlicher Analysefunktionen.
- Milestone 2: M9-M12 Interne Tests und Optimierung des Backends unter Verwendung vorhandener Datensätze
- Milestone 3: M15. Veröffentlichung eines Manuskripts, in dem das Backend beschrieben wird; das Backend ist als voll funktionsfähige Open-Source-Software über eine spezielle Website verfügbar.

Milestones 1-3 wurden ohne Einschränkungen innerhalb des angegebenen Zeitrahmens erreicht und sind in mehreren Publikationen dokumentiert, allen voran die Publikation des quelloffen frei verfügbaren Backends, Alawi et al., Sci Rep. 2019.

Alawi M, Burkhardt L, Indenbirken D, Reumann K, Christopeit M, Kröger N, Lütgehetmann M, Aepfelbacher M, Fischer N, Grundhoff A (2019). DAMIAN: an open source bioinformatics tool for fast, systematic and cohort based analysis of microorganisms in diagnostic samples. Sci Rep. 2019 Nov 14;9(1):16841. doi: 10.1038/s41598-019-52881-4.

Das quelloffe Software Paket kann über die folgende Website heruntergeladen werden:  
<https://sourceforge.net/projects/damian-pd/>

Ein WIKI mit Informationen zu Installation der Software und Analyse der Datent steht unter <https://sourceforge.net/p/damian-pd/wiki/Home/> zur Verfügung.

- Milestone 4: M6-M18 Entwicklung und Test des Frontends/der grafischen Benutzeroberfläche.

Die graphische Oberfläche wurde entsprechend der Vorhabenbeschreibung entwickelt und den Kooperationspartnern zur Verfügung gestellt.

- Milestone 5: M19-M30 Implementierung/Evaluierung der Software an den Standorten der Kooperationspartner. Optimierung und Test der Software wird am LIV fortgesetzt
- Milestone 6: M31-M36 Abschließende Optimierung der Software entsprechend dem Feedback der Partner.

Im Rahmen der Milestones 5 und 6 war eine Implementierung und Evaluierung der Software an zwei Standorten anvisiert: 1. Institut für Medizinische Mikrobiologie, Virologie und Hygiene, Universitätsklinikum Hamburg-Eppendorf und 2. Institut für Virologie, Stiftung Tierärztliche Hochschule Hannover. Für die genaue Umsetzung der kollaborativen Tätigkeiten verweisen wir auf Kapitel 5. Die hier genannten Meilensteine wurden wie geplant durchgeführt.

- Milestone 7: M36 Fertiggestelltes Softwareprodukt für den Vertrieb

Milestone 7 wurde aufgrund einer Neubewertung der kommerziellen Erfolgsaussichten nicht umgesetzt (siehe genauere Erläuterung in Kapitel 2) .

Entsprechend des Finanzplans wurden Projektmittel zur Finanzierung eines mit der Entwicklung der Software betrauten Bioinformatikers, sowie zur Anschaffung zweier Analyse-Workstations verwendet, welche den klinischen Partnern für die Dauer der Evaluationsphase leihweise zur Verfügung gestellt wurden. Im Laufe des Verfahrens zeigte sich, dass die Validierung der entwickelten Methoden die Erhebung von Primärdaten aus der laufenden klinischen Diagnostik erforderte. Die ursprünglich zur Freistellung eines klinischen Mitarbeiters des UKE vorgesehenen Personalmittel wurden daher in Sachkosten zur Sequenzierung umgewandelt. Personalressourcen zur Evaluation der Software am Institut für Mikrobiologie, Virologie und Hygiene wurden durch den Institutsleiter Herrn Prof. Aepfelbacher zur Verfügung gestellt (siehe Abschnitt 7)

## 2. Aktivitäten und Hindernisse

Zu einzelnen Aktivitäten innerhalb des Vorhabens verweisen wir auf die Erläuterung zu den vorgesehenen Milestones in Kapitel 1. Wir möchten innerhalb dieses Kapitels ausschließlich auf Entwicklungen im Forschungsgebiet und sich daraus ergebende Abweichungen von der ursprünglich vorgesehenen Planung eingehen. Im Wesentlichen ergaben sich diese aus der Tatsache, dass mehrere kommerzielle Unternehmen die Methodik der Shotgun-Metagenomik übernommen und kommerzielle Tests für verschiedene Körperquellen entwickelt haben. Dazu gehört z.B. der Karius Blut/Plasma-Test. Dieser Test, der als mikrobieller zellfreier DNA-Test oder als nicht-invasive Flüssigbiopsie bezeichnet wird, wurde 2016 erstmals auf dem Markt für Infektionsdiagnostik eingeführt. Der Test ist nicht von der FDA zugelassen, wird aber als ein akkreditierter Test von Karius zu einem Preis von 2,500 US Dollar und einer 1 wöchigen Bearbeitungszeit pro Probe durchgeführt. Ein weiterer kommerziell entwickelter Shotgun-Metagenomik-Test ist Explify™ Respiratory, der in erster Linie mit BAL-Proben durchgeführt wird und 2017 von IDbyDNA und ARUP Laboratories eingeführt wurde und seit 2022 von Illumina vertrieben wird. Der Nachweis von Mikroorganismen basiert auf Nachweisschwellen, die durch Qualitätskontrollstudien festgelegt wurden, und die Ergebnisse werden semi-quantitativ angegeben. Die Ergebnisse werden auch stratifiziert und als potenzielle Krankheitserreger oder zusätzliche Mikroorganismen auf der Grundlage bekannter Pathogenität angegeben. Auch dieser Tests ist kein zugelassener diagnostischer Test und auch hier bedarf es an klinischen Studien. Insbesondere die Bedeutung polymikrobieller Ergebnisse und die Frage, ob die hohen Kosten der Tests durch die klinischen Ergebnisse aufgewogen werden, wird derzeit diskutiert. Dennoch hat die Anzahl von Firmen mit Lösungen zu Metagenomik in diagnostischen Proben während der Projektlaufzeit erheblich zugenommen, und erfährt derzeit u.a. durch die infolge der Pandemie gestiegene Wahrnehmung des zusätzlichen Erkenntnisgewinns durch Pathogen-Genomik eine weitere Zunahme. Dabei investieren die kommerziellen Anbieter erhebliche Ressourcen in entsprechende Serviceangebote, und betreiben zudem einen hohen Aufwand um Aufträge von klinischen/diagnostischen Laboren einzuwerben. Aufgrund der in diesem Umfang nicht vorhergesehenen Konkurrenzlagen wurde entschieden, den in Milestone 7 vorgesehenen Vertrieb des Frontends im Rahmen eines kommerziellen Lizenzmodelles vorerst nicht umzusetzen. Dies ergibt sich insbesondere aus den erheblichen Ressourcen, die für den Support eines solchen Produktes fortlaufend zur Verfügung gestellt werden müssen, die aber in Anbetracht der veränderten Konkurrenzlage jedoch kaum durch die voraussichtlichen Einnahmen gedeckt werden können. Für die Zukunft planen wir stattdessen, Analysen auf Grundlage der entwickelten Software als Service anzubieten. Die seit Mai 2022 in Kraft getretenen neuen Vorschriften der Europäischen Union zu In-vitro-Diagnostika (EU IVDR) stellen dabei eine wesentliche Herausforderung dar. Derzeit engagieren wir uns im Rahmen mehrerer Verbünde (siehe Abschnitt 3), um Standards für im Sinne der IVDR umsetzbare metagenomische Abläufe zu etablieren.

### 3. Ergebnisse und Erfolge

Wie bereits in Kapitel 1 aufgeführt hat dieses Vorhaben zu der Veröffentlichung von Publikationen geführt, die mit der Beteiligung des LIV entstanden sind:

- Alawi M, ...Grundhoff A. Sci Rep. 2019 Nov 14;9(1):16841.
- Rubbenstroth D,...Grundhoff A, ...Teske L. Transbound Emerg Dis. 2019 Jan;66(1):552-561.
- Fux R, ...Grundhoff A, ... Sutter G. Viruses. 2019 Jan 28;11(2).
- Coronado L, ... Grundhoff A, ... Ruggli N. Vet Microbiol. 2017 Mar;201:103-112. 39.
- Günther T, ... Grundhoff A, ... Fischer N. Sci Rep. 2017 Jun 16;7(1):3734.
- Postel A, ... Grundhoff A, ..., Becher P. Emerg Infect Dis. 2017 Dec;23(12):2104-2107.
- Baechlein C, Grundhoff A, ... Becher P. Emerg Infect Dis. 2016 Jul;22(7):1312-4.
- Postel A, ... Grundhoff A, ...Becher P. Sci Rep. 2016 Jun 13;6:27735.

Die Software und Beteiligte des Vorhabens sind Teil mehrerer diagnostischer Netzwerke, die sich der Qualitätssicherung und Regularien von metagenomischer Pathogendetektion in diagnostischen Proben widmen:

- DiaMETA, Deutsches Netzwerk für Metagenomdiagnostik in der Infektionsmedizin <https://diameta-net.fli.de/de/home>
- ENNGS, ESCV Network on Next Generation Sequencing <https://escv.eu/enngs/>

Teile des hier beschriebenen Vorhabens wurden auf zahlreichen Kongressen und Workshops vorgestellt, siehe beiliegende Exceltable.

Das hier beschriebene Vorhaben hat zum Abschluss der Dissertationsarbeit zur Erlangung des Dr.rer. nat. von Herrn Malik Alawi beigetragen:

- 2021: Development of a software for the identification of novel and known pathogens in clinical Samples

### 4. Chancengleichheit, Karriereförderung und Internationalisierung

Das LIV unterstützt die Vereinbarung der wissenschaftlichen Karriere und Familiengründung sowie es viele Förderprogramme bietet in den Bereichen Geschlechtergerechtigkeit, Diversität und Vereinbarkeit. <https://www.leibniz-liv.de/de/institut/gleichstellung/foerderprogramme/>

### 5. Strukturen und Kooperationen

Die Software wird am Standort des Kollaborationspartners 1 eingesetzt und ist dort Teil einer klinischen Serviceleistung geworden, die zusätzliche Diagnostik zum Erregernachweis in spezifischen Proben anbietet. Die Proben werden von einem infektiologischen Board ausgesucht und die Leistung zusätzlich zur Routine-Erregerdiagnostik durchgeführt. Die Software wird für den Kollaborationspartner 2 als remote Lösung betrieben, Proben werden am LIV sequenziert und im Rahmen fortgeführter Kollaborationsprojekte eingehend ausgewertet. Über diese beiden Kollaborationspartner hinaus wurde die Software im Vergleich mit 13 anderen verfügbaren Software Lösungen (z.T. kommerzielle Pipelines, frei verfügbar über Web Interface Lösungen oder per Download) innerhalb des ENNGS Netzwerks, *European Society for Clinical Virology NGS network*, getestet. Über diesen Vergleich der bioinformatischen Auswertemöglichkeiten von metagenomischen Daten der Erregerdiagnostik berichten die folgenden Publikationen:

1. López-Labrador FX et al. Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: wet lab procedure. J Clin Virol. 2021 Jan;134:104691.
2. de Vries JJC, et al. ESCV Network on Next-Generation Sequencing (2021). Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting. J Clin Virol. 2021 May;138:104812. doi: 10.1016/j.jcv.2021.104812.
3. de Vries JJC, et al. Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples. J Clin Virol. 2021 Aug;141:104908.

Die Anwendung der Software zum Nachweis von Pathogenen in tiermedizinischen Proben in Zusammenarbeit mit dem Kollaborationspartner 2 führte zum erfolgreichen Nachweis von Erregern, die mit der klinischen Symptomatik in Zusammenhang gebracht werden konnten. Der Einsatz der Software ist in den folgenden Publikationen dokumentiert:

1. Rubbenstroth D, et al. Emergence of a novel clade of group A rotaviruses associated with fatal disease in domestic pigeons in Europe. *Transbound Emerg Dis.* 2019 Jan;66(1):552-561.
2. Fux R et al. Piscine Orthoreovirus 3 is not the causative pathogen of proliferative darkening syndrome (PDS) of brown trout (*Salmo trutta fario*). *Viruses.* 2019 Jan 28;11(2).
3. Coronado L et al. Identification of a novel poly-uridine stretch in the 3'UTR and of amino acid substitutions in E2 acquired during attenuation of an endemic CSFV in the field. *Vet Microbiol.* 2017 Mar;201:103-112. 39.
4. Günther T et al. Recovery of the first full-length genome sequence of a parapoxvirus directly from a clinical sample. *Sci Rep.* 2017 Jun 16;7(1):3734.
5. Postel A et al. High Abundance and Genetic Variability of Atypical Porcine Pestivirus in Pigs from Europe and Asia. *Emerg Infect Dis.* 2017 Dec;23(12):2104-2107.
6. Postel A et al. Frequent presence of hepaci und pegiviruses in commercial equine serum pools. *Vet Microbiol.* 2016 Jan 15;182:8-14.
7. Baechlein C et al. Porcine pegivirus infection. *Emerg Infect Dis.* 2016 Jul;22(7):1312-4.
8. Postel A et al. Presence of atypical pestivirus (APPV) genomes in newborn piglets correlates with congenital tremor. *Sci Rep.* 2016 Jun 13;6:27735.

## 6. Qualitätssicherung

Das LIV verfolgt die Einhaltung guter wissenschaftlicher Praxis gemäß der DFG Leitlinien zur Sicherung guter wissenschaftlicher Praxis

[https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf)

Die Software/Backend steht wie unter Kapitel 1 beschrieben der Öffentlichkeit frei zur Verfügung. Die Publikation über das Backend und das Benchmarking derselben wurde in *Scientific Reports* als Open Access Arbeit veröffentlicht. Im Rahmen des Projektes wurden keine Tierversuche durchgeführt.

## 7. Zusätzliche Ressourcen

Im Rahmen des Vorhabens wurden durch das LIV *in kind* Leistungen in Höhe von ca. 92 TEuro erbracht (10 Personenmonate (PM) einer wissenschaftlichen Mitarbeiterin (72 TEuro), sowie 20 TEuro Sachmittel zur Sequenzierung) erbracht. Das Universitätsklinikum Hamburg Eppendorf erbrachte *in kind* Leistungen in Höhe von ca. 80 TEuro (12 Personenmonate einer ärztlichen Mitarbeiterin).

## 8. Ausblick

Das Gebiet der hypothesenfreien diagnostischen Verfahren bzw. Verfahren mit erweitertem Erregerspektrum werden zukünftig einen wesentlichen Teil der mikrobiologischen Diagnostik ausmachen. Das im Rahmen dieses Antrages entwickelte, frei verfügbare Software Paket erlaubt den ressourcenschonenden Einsatz dieser Technologie in klinisch-diagnostischen Laboratorien. Für die Zukunft planen wir, zusammen mit klinischen Partnern auf Grundlage dieser Software einen umfassenden Analyseservice anzubieten, der neben metagenomischer Klassifikation auch eine Bewertung zur klinischen Interpretation der Ergebnisse einschließt.

Anlagen:

- 1-seitiger Erfolgskontrollbericht