

Umwelt-Sequenzierproben

Informationsschatz aus Umweltproben

EMANUEL WYLER, MARKUS LANDTHALER

MAX-DELBRÜCK-CENTRUM FÜR MOLEKULARE MEDIZIN, BERLINER INSTITUT FÜR MEDIZINISCHE SYSTEMBIOLOGIE, BERLIN

Our environment is a treasure trove of genetic information. High-throughput sequencing enables a spectrum of investigations, from observing human pathogens to assessing the constitution of an ecosystem. Here, we summarize our work on longitudinal total RNA/DNA sequencing from wastewater samples combined with in-depth data analysis. We explore the wealth of data from known virus detection to discovery of novel biotechnological enzyme sequences, showcasing the potential of such systematic approaches.

DOI: 10.1007/s12268-025-2401-4
© The Author(s) 2025

■ Unsere Umwelt ist prall gefüllt mit genetischer Information. Manche Bakterien und Pilze findet man überall, andere sind auf ihre Nischen spezialisiert. Im Wald findet man auch noch die DNA von dort lebenden Tieren, in Gewässern im Jahresverlauf und je nach Nährstoffangebot verschiedene Algen, im Abwasser menschliche Krankheitserreger oder Darmbakterien. Die Hochdurchsatzse-

quenzierung von DNA- und RNA-Molekülen kann diesen Informationsschatz erschließen – zu laufend niedrigeren Preisen bei steigender Qualität. In der Wissenschaft können die unterschiedlichsten Fragestellungen damit angegangen werden. Wie verändert sich die Biodiversität entlang eines Flusslaufes? Wie geht die Evolution eines Virus voran, wie verbreitet er sich weltweit? Wie bewegen

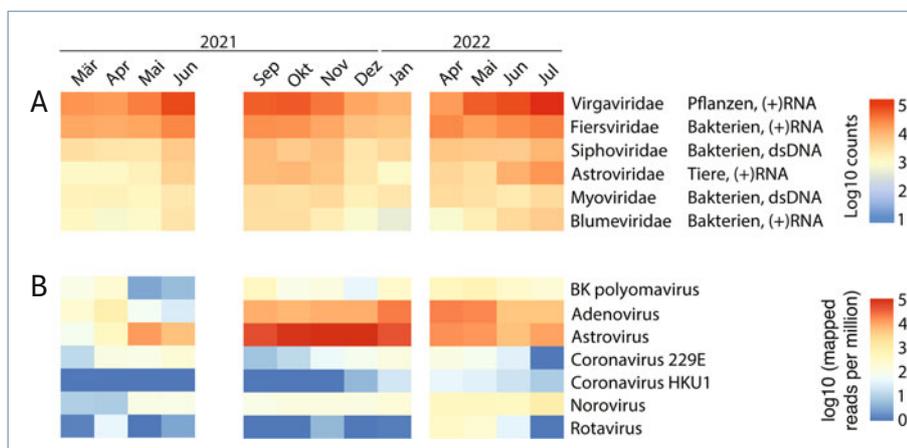
sich Antibiotikaresistenzen durch die Landschaft? Wie lassen sich negative Entwicklungen, wie etwa ein Fischsterben, frühzeitig prognostizieren? Gleichzeitig weisen diese Fragen auch in Richtung möglicher Anwendungen für die Gesundheit von Menschen, Tieren, Ökosystemen, für die Trinkwasserversorgung, für den Umgang mit Industrieabwässern und -unfällen usw.

Abwassermonitoring auf Humanpathogene

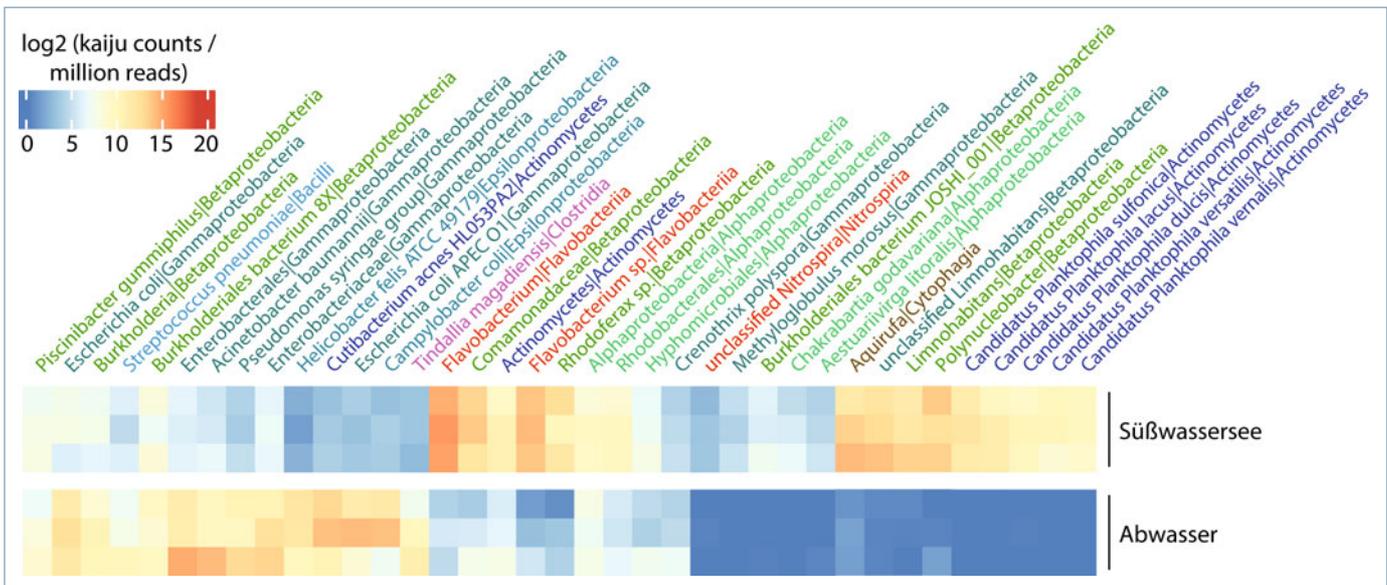
Humanpathogene werden seit Jahrzehnten in einer besonderen Art von Umweltproben untersucht, in Abwasser. Seit 1939 ist bekannt, dass sich Polioviren im Abwasser nachweisen lassen (was kürzlich auch wieder in Deutschland geschehen ist [1]), mit der Corona-Pandemie ist diese Art von Monitoring richtig durchgestartet [2]. In den letzten Jahren wurde mehrfach dargelegt, dass das Abwasser nicht nur zeigt – mit einigen Tagen Vorlauf gegenüber klinischer Testung –, wieviel SARS-CoV-2 in der Bevölkerung zirkuliert, sondern dank Sequenzierungen auch welche Varianten [3].

Schon seit mehreren Jahren war klar, wie vielfältig die Mikroben im Abwasser sind; eigentlich wenig überraschend für eine Matrix, die sich aus so unterschiedlichen Quellen speist. Was in unseren Augen aber fehlte, war eine Beobachtung der gesamten genetischen Information im Abwasser über einen längeren Zeitraum und über verschiedene Kontinente hinweg, kombiniert mit einer eingehenden Datenanalyse.

Wir extrahierten daher Nukleinsäuren (RNA und DNA getrennt) aus Abwasserproben eines Berliner Klärwerks über einen Zeitraum von insgesamt 17 Monaten, und sequenzierten die gesamte RNA bzw. DNA darin zu einer Tiefe von jeweils 20–40 Mio. Sequenzbereichen [4]. Aus einer einzigen Probe – 60 ml Rohzufluss, gefiltert durch 2 und 0,2 µm – kann genetische Information extrahiert werden, die ungefähr 10.000–20.000 verschiedenen Mikroben zugeordnet werden kann. Die meisten davon sind Bakterien, daneben insbesondere Phagen (Viren,



▲ **Abb. 1:** Bestimmung verschiedener Viren im Berliner Abwasser. **A,** Intensität von verschiedenen Virusfamilien über den Untersuchungszeitraum, zusammengefasst nach Monat. Farben entsprechen \log_{10} -transformierten Counts aus der Metagenomik-Pipeline kaiju. Neben den Virusfamilien sind die Wirte sowie die Art des Virusgenoms aufgeführt. **B,** Anzahl der Sequenzier-Reads für verschiedene humane Viren nach Anreicherung. Die Farbgebung entspricht der Menge zugeordneter Reads (*mapped reads per million*), zusammengefasst nach Monat.



▲ **Abb. 2:** Abundanzen einer Auswahl von Familien/Gattungen/Arten von Bakterien aus drei Abwasser- und drei Süßwasserseeproben. Die Bezeichnungen der Taxa sind eingefärbt nach Klasse, diese sind jeweils nach dem senkrechten Strich angefügt. Die Farben der Felder entsprechen \log_{10} -transformierten Counts aus der Metagenomik-Pipeline kaiju. Unterscheidbar sind bspw. ubiquitäre Bakterien und solche, die zu unterschiedlichem Maße vor allem in Abwasser oder Süßwasser vorkommen.

die Bakterien befallen), aber auch ein großes Spektrum von humanpathogenen Viren (**Abb. 1**). Wie viel genetische Informationen wir im Abwasser von einem Virus finden, das Menschen infiziert, hängt natürlich von der Anzahl Menschen mit Ansteckung im Einzugsgebiet ab, aber auch, wie sehr virales Erbgut in die Kanalisation kommt, und wie stabil es ist. Während wir gesehen haben, dass die RNA von SARS-CoV-2 wahrscheinlich in fragmentierter Form vorliegt, kommen Enteroviren infektiös, d. h. wohl als Ganzes intakt, in der Kläranlage an. Das ist auch der Grund, warum die zu den Enteroviren gehörenden Polioviren aus Abwasser auch in kleinsten Mengen nachgewiesen werden können, wie eben vom Robert-Koch-Institut auch in Deutschland demonstriert [1]: Die Viren werden nicht durch Sequenzierung gefunden, sondern durch Anzucht in Zellkultur.

Um wenig abundante virale Genome bzw. Fragmente zu finden, verwendeten wir ein breites Anreicherungspanel mit über 100 humanpathogenen Viren. Damit kommt dann bspw. auch das Respiratorische Syncytialvirus zum Vorschein, das in der Sequenzierung der Gesamt-RNA nicht identifiziert wird. Wie auch von vielen anderen Forschungsgruppen gezeigt, ist damit grundsätzlich eine Vielzahl von humanen Pathogenen nachweisbar [5, 6].

Neue Viren, neue Enzyme aus dem Abwasser

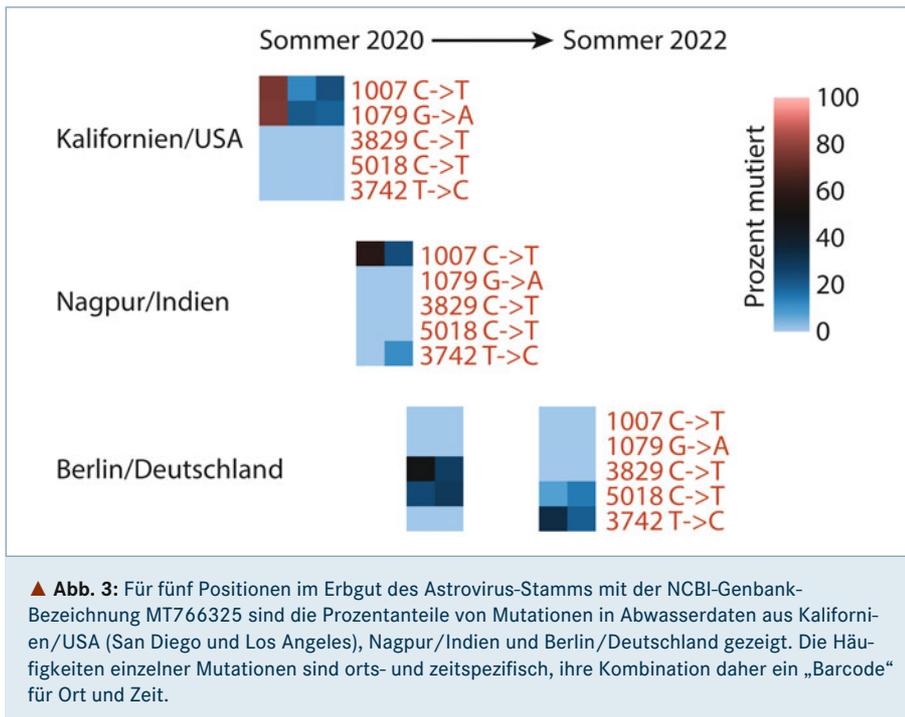
Mit der so großen Menge an genetischer Information aus dem Abwasser lässt sich aber nicht nur nachweisen, was schon bekannt ist. In den letzten Jahren haben computergestützte Analysen aus gigantischen Sequenzier-Datenmengen im Petabyte-Bereich (10^6 bis 10^7) gezeigt, wie viele Virusarten wir noch gar nicht kennen [7, 8] – weniger solche, die Menschen infizieren, als vielmehr das ganze Spektrum von Arten auf diesem Planeten. Für diese Definition des „planetaren Viroms“ scheint nun das Abwasser eine ganz besonders ergiebige Quelle zu sein, fanden wir doch über 10.000 „neue“ Viren – wobei deren Artengrenzen sehr fluide sein können, und nicht immer klar ist, wo eine Art endet und die nächste anfängt. Mit solchen Erkenntnissen erschließen sich neue Forschungsfelder im Bereich ökosystemweiter Betrachtungen infektiöser Elemente.

Anhand der genetischen Informationen können aber nicht nur neue Viren gefunden werden. In einem zweiten Ansatz suchten wir nach möglicherweise biotechnologisch nutzbaren Enzymsequenzen. Als Beispiel nahmen wir Transposase B (TnpB). Wie die berühmt gewordenen Cas9-Enzyme können sie mithilfe von guide-RNAs bestimmte DNA-Segmente spezifisch erkennen und verändern [9]. Sie sind aber deutlich kleiner, und deswegen einfach handhabbarer als Cas9-Enzyme. Wie bei den Viren fanden wir reich-

haltig neue mögliche TnpB-Sequenzen. Wahrscheinlich sind nicht alle davon aktiv, geschweige denn irgendwie direkt nutzbar. Warum ist dann diese Sequenz-Sammlung trotzdem nützlich? TnpBs bestehen aus knapp 400 Aminosäuren. Die aktiven Stellen mit ungefähr 20–30 Aminosäuren sind eindeutig definiert, der Rest nicht – dieser Rest entscheidet aber über die Stabilität, Aktivität und Spezifität des Enzyms. Auch wenn mittlerweile computergestützte Methoden schon sehr gut in der Vorhersage stabiler 3D-Strukturen sind – die vielfältigen biochemischen Charakteristika sind nur experimentell bestimmbar. Wenn nun eine spezifische Aminosäuresequenz im Abwasser auffindbar ist, bedeutet das, dass es irgendwo Bakterien gibt, die genau dieses TnpB-Gen in sich tragen. Das wiederum heißt, dass die entsprechende Sequenz biochemisch funktioniert. Das Abwasser hilft also, den gigantischen „search space“ für Aminosäuresequenzen für interessante Enzyme auszuloten und einzugrenzen.

Metagenomik als Methode in der umfassenden Umweltbeobachtung

Die immer günstiger und einfacher werdende Hochdurchsatzsequenzierung hat nicht nur die Abwasserforschung verändert. Metagenomik von Umweltproben, sei es aus Süßwasser, Ozeanen, Sedimenten, Boden, Wolken, Abstrichen von Pflanzenblättern im Urwald ist allgegenwärtig – alles kann



sequenziert werden [10]. Eine einfache Analyse ist schlicht die Bestimmung und der Vergleich der Biodiversität in verschiedenen Proben, wie wir es bspw. aus Abwasser- und Süßwasserseedaten machen können (Abb. 2). Longitudinale Datenerhebungen und eine gemeinsame Analyse mit physikalischen und chemischen Daten wie Salz- und Nährstoffgehalt, Temperatur, Wasserstand und Flussraten usw. erlauben in einem ersten Schritt ein umfassendes Verständnis der biologischen Prozesse in einem Ökosystem, wie bspw. einem Flusslauf. Daraus können Schlüsse gezogen werden, etwa was die Auswirkungen steigender Temperaturen im Zuge der Klimaerwärmung angeht, oder wie die mikrobielle Zusammensetzung im Gewässer mit der Trinkwasserqualität aus den daraus gespeisten Aufbereitungsanlagen zusammenhängt. Besonders interessant sind Prognosemöglichkeiten. Nach dem Motto „viel hilft viel“ können Vorhersagen umso robuster werden, je mehr Daten vorhanden sind. Für die Analyse dieser gigantischen Datensammlungen sind Algorithmen der „künstlichen Intelligenz“ ideal, da es darum geht, im Rauschen der vielen Sequenzen relevante Muster zu erkennen und die mit anderen Messwerten zu verknüpfen.

Der Fluss genetischer Elemente in Raum und Zeit

In unserer Publikation haben wir auch deswegen auf Viren fokussiert, da sie sich

außerhalb eines Wirts nicht vermehren können – im Gegensatz zu Bakterien: Wenn genetische Information von Bakterien in einer Abwasserprobe von einem Klärwerk gefunden werden, kann ein Bakterium aus einem Menschen stammen, vielleicht hat es sich aber auch in einem Biofilm in der Kanalisation vermehrt. Nun kann man die Bewegung von Viren in einer zeitlichen Dimension verfolgen: Wann sind Infektionswellen, wie verändert sich das Virus über die Zeit? Genauso interessant ist die räumliche Dimension: Auf welchen Wegen und wie schnell verbreitet sich ein bestimmtes Virus? Für einen ersten Ansatz in diese Richtung haben wir die Stämme und Genomsequenzen von Astroviren aus den Berliner Abwasser-Sequenzierdaten mit solchen aus Kalifornien und der indischen Stadt Nagpur verglichen. Beide waren vorgängig von zwei unterschiedlichen Forschungsgruppen frei verfügbar gemacht worden. (Die dazugehörigen Publikationen gingen nicht auf Astroviren ein; das Beispiel zeigt also auch, wie wichtig es ist, den Zugang von Hochdurchsatzdaten zu ermöglichen. Der Reichtum solcher Daten ist durch eine Forschungsarbeit allein oft gar nicht zu erschließen.)

Das Erbgut von Astroviren, die in Bevölkerungen mit ausreichender Ernährung und grundsätzlich guter Gesundheit kaum merkbare Magen-Darm-Erkrankungen auslösen, ist im Abwasser in relativ großen Mengen

vorhanden. Dadurch lassen sich nicht nur Virusstämme zuordnen, sondern das einzelne Veränderungen im Erbgut über Ort und Zeit verfolgen. Der Vergleich der Daten über Ort und Zeit zeigt nun, wie sich bestimmte Erbgutvarianten verbreiten (Abb. 3). Eine Ausweitung solcher Analysen könnte darstellen, wie schnell und auf welchen Wegen sich unterschiedliche Viren über den Erdball verbreiten und wie häufig dabei ihr Erbgut mutiert. Praktisch relevant wäre dieses Wissen, wenn es die Verbreitung eines neuen Virus – wie aktuell H5N1-Influenza oder mpox – vorauszusagen gelte.

Grundsätzlich könnte diese Art von Analyse mit jedem genetischen Element durchgeführt werden. Interessant wären Antibiotika-Resistenzgene, die unter Bakterien weitergegeben werden. Sie tauchen eigentlich nur auf, wenn sie in bei einem Menschen behandelt werden müssen. Wo sie aber „zwischen-durch“ sind, ist weniger klar. Inwieweit bewegen sich die Gene bspw. zwischen der Umwelt, von Wild- zu Zuchttieren und zurück und dann zum Menschen? Diese Forschung steht erst am Anfang [11]. Systematische Probennahmen kombiniert mit Sequenzierungen können aber auch hier helfen, zu verstehen, wie sich Antibiotika-Resistenzgene im gesamten Ökosystem verbreiten und entwickeln und wie sie gefährlich werden können.

Ausblick

Vor einigen Jahren war Hochdurchsatzsequenzierung noch eine teure Angelegenheit, jede Analyse einer Nukleinsäureprobe musste abgewogen werden. Die Preise für *short read sequencing* sind mittlerweile deutlich gefallen und beim *long read sequencing* gehen simultan die Preise nach unten und die Qualität nach oben. Aufwändig für eine umfassende Vermessung von Ökosystemen mit Metagenomik und -transkriptomik ist die Probensammlung im großen Stil sowie die Methodologie in der Aufarbeitung – und dann natürlich die Datenanalyse und die Verknüpfung mit anderen relevanten Messwerten und Ereignissen. Interdisziplinäre Kollaborationen aus den Bereichen Umwelt- und Gesundheitsforschung sowie Bioinformatik/Künstliche Intelligenz werden daher notwendig, um diese komplexe Herausforderung anzugehen. ■

Literatur

[1] Robert Koch-Institut (2024) Schluckimpfstoff-abgeleitete Polioviren in Abwasserproben an weiteren Orten in Deutschland nachgewiesen. *Epid Bull* 49: 14

- [2] Kilaru P, Hill D, Anderson K et al. (2023) Wastewater Surveillance for Infectious Disease: A Systematic Review. *Am J Epidemiol* 192: 305–322
- [3] Schumann VF, de Castro Cuadrat RR, Wyler E et al. (2022) SARS-CoV-2 infection dynamics revealed by wastewater sequencing analysis and deconvolution. *Sci Total Environ*, DOI: 10.1016/j.scitotenv.2022.158931
- [4] Wyler E, Lauber C, Manukyan A et al. (2024) Pathogen dynamics and discovery of novel viruses and enzymes by deep nucleic acid sequencing of wastewater. *Environ Int*, DOI: 10.1016/j.envint.2024.108875
- [5] Parkins MD, Lee BE, Acosta N et al. (2024) Wastewater-based surveillance as a tool for public health action: SARS-CoV-2 and beyond. *Clin Microbiol Rev* 37: e00103-22
- [6] Singh S, Ahmed AI, Almansoori S et al. (2024) A narrative review of wastewater surveillance: pathogens of concern, applications, detection methods, and challenges. *Front Public Health* 12: 1445961
- [7] Edgar RC, Taylor B, Lin V et al. (2022) Petabase-scale sequencing alignment catalyses viral discovery. *Nature* 602: 142–147
- [8] Hou X, He Y, Fang P et al. (2024) Using artificial intelligence to document the hidden RNA virosphere. *Cell* 187: 6929–6942
- [9] Karvelis T, Druteika G, Bigelyte G et al. (2021) Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 599: 692–696
- [10] Dauphin B, Rellstab C, Wüest O et al. (2023) Re-thinking the environment in landscape genomics. *Trends Ecol Evol* 38: 261–274
- [11] Abramova A, Berendonk TU, Bengtsson-Palme J (2023) A global baseline for qPCR-determined antimicrobial resistance gene prevalence across environments. *Environ Int*, DOI: 10.1016/j.envint.2023.108084

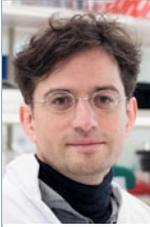
Funding note: Open Access funding enabled and organized by Projekt DEAL.

Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Korrespondenzadresse:

Dr. Emanuel Wyler
 Prof. Dr. Markus Landthaler
 Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtzgemeinschaft
 Berliner Institut für Molekulare Medizin in der Helmholtzgemeinschaft
 Hannoversche Straße 28
 D-10115 Berlin
 emanuel.wyler@mdc-berlin.de
 markus.landthaler@mdc-berlin.de

AUTOREN



Emanuel Wyler

Jahrgang 1981. Biochemiestudium an der Universität Zürich, Schweiz, danach Forschungsaufenthalt am Institut Pasteur in Paris, Frankreich. Doktorarbeit an der ETH Zürich, danach PostDoc und seit 2020 fest angestellter Wissenschaftler am Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft.



Markus Landthaler

Jahrgang 1969. Biologiestudium an der Universität Würzburg. Doktorarbeit an der University of Albany, NY, und Promotion in Würzburg. PostDoc bei Prof. Dr. T. Tuschl an der Rockefeller-Universität, New York. Seit 2009 Forschungsgruppenleiter am Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft. Seit 2017 zudem Professor für RNA-Biologie an der HU Berlin.