

Sequenzierung

Moderne Sequenzierungstechnologien und ihre Anwendungen

JOCHEN DOBNER, ANDREA ROSSI

IUF - LEIBNIZ-INSTITUT FÜR UMWELTMEDIZINISCHE FORSCHUNG, DÜSSELDORF

CRISPR-mediated genome editing enables to study disease-relevant mutations. Next Generation Sequencing offers high throughput and accuracy but requires substantial investment. Nanopore sequencing on the other hand provides cheap entry. Here, we describe our work using sequencing techniques to identify edited cells, analyze mitochondrial DNA, and gene expression. Finally, we describe how our software tools CRISPRnano and Duesselpore aid to democratize the use of modern sequencing technologies.

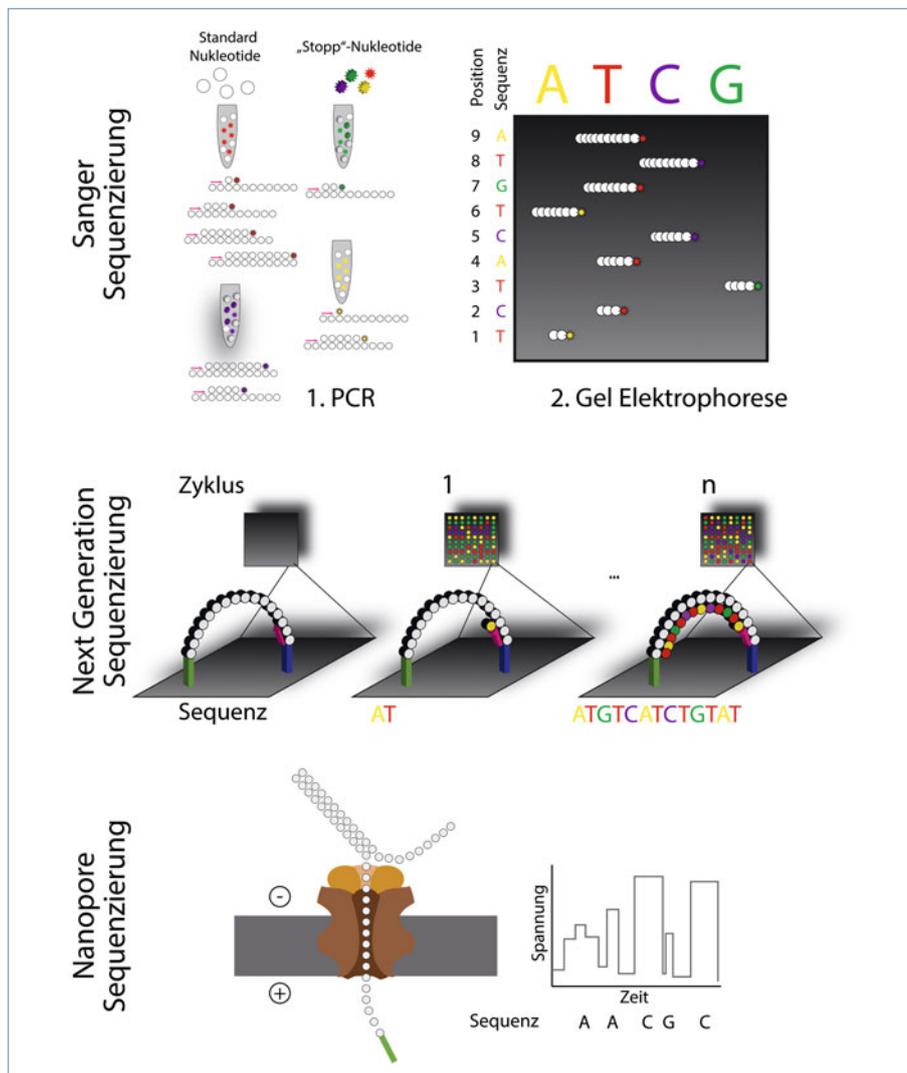
DOI: 10.1007/s12268-023-1989-5

© Die Autoren 2023

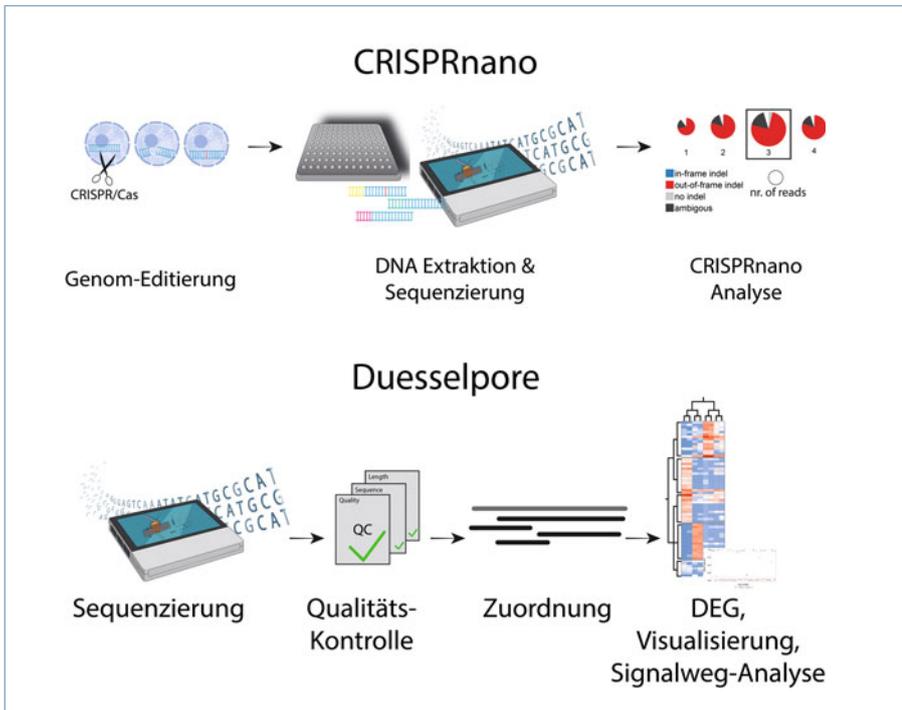
■ Mit der routinemäßigen Nutzung der Gen-schere CRISPR [1] ist die Generierung genomeditierter Zelllinien heute einfach wie nie. Die Anwendungen reichen über Generierung von Knockout-Modellen bis hin zur komplexen Einführung gerichteter Mutationen durch Knock-In, Base- oder Prime-Editing [2]. Unabhängig davon ob eine resultierende Zelllinie für die Forschung oder für den Einsatz in Patienten verwendet werden soll, ist es essenziell, Mutationen eindeutig nachzuweisen. Klassisch wird hierfür die Sequenzierung nach Sanger benutzt, die auf der Synthese eines DNA-Strangs anhand einer Vorlage basiert. Fluoreszenzmarkierte Didesoxyribonukleosidtriphosphate führen zu stochastisch verteilten Abbrüchen der Reaktion und ermöglichen es so, die Zielsequenz abzulesen (**Abb. 1**). Die Sanger-Sequenzierung bietet jedoch zwei entscheidende Nachteile, nämlich geringen Durchsatz und geringe Sensitivität zum Aufspüren seltener Mutationen. Beide Nachteile werden durch *next generation sequencing* (NGS) [3] überwunden.

Sequenzierung durch Synthese

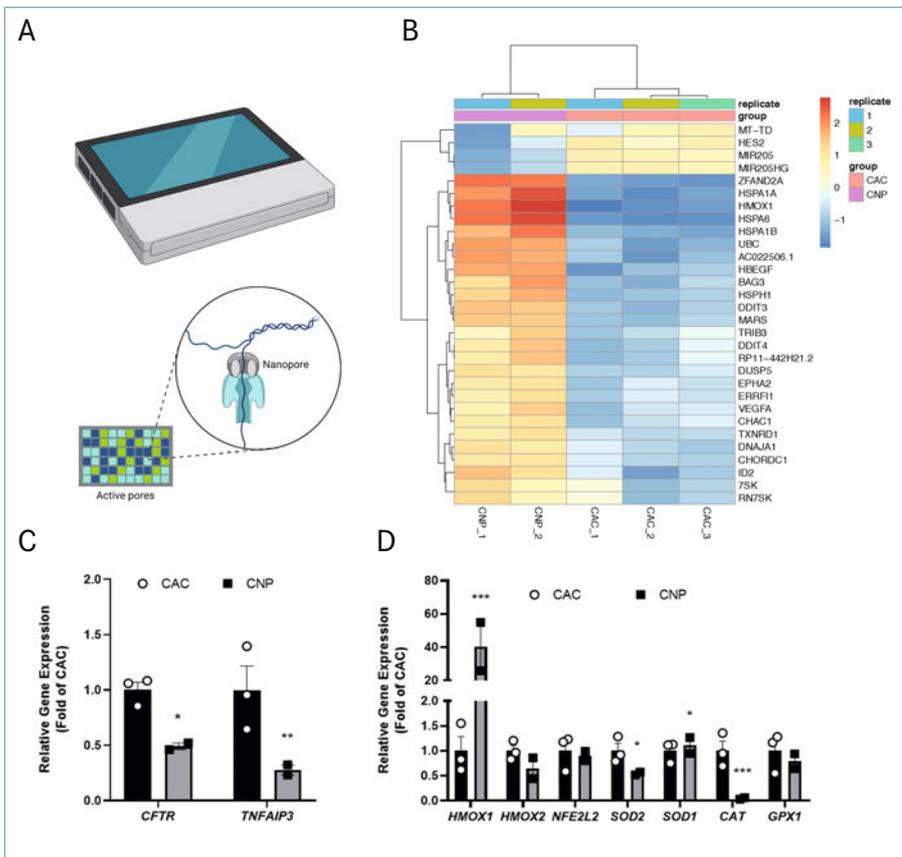
Die am weitesten verbreitete NGS-Methode der Firma Illumina basiert ebenfalls auf der Sequenzierung durch Synthese unter Ver-



◀ **Abb. 1:** Übersicht über die wichtigsten Sequenzierungs-Techniken. Bei der Sanger-Sequenzierung (Kettenabbruch-Methode) wird anhand einer Vorlage ein Komplementärstrang synthetisiert. Durch Zugabe fluoreszenzmarkierter Stopp-Nukleotide bricht die Synthese immer dann ab, wenn ein solches eingebaut wird. Mittels Gelelektrophorese werden die Fragmente nach ihrer Größe aufgetrennt und das jeweils zuletzt eingebaute kann so einer Position in der Sequenz zugeordnet werden. Beim *next generation sequencing* durch Synthese werden fluoreszenzmarkierte Nukleotide eingebaut, die über ein hochsensitives Mikroskop detektiert und so einer Sequenz zugeordnet werden können. Die Nanopore-Sequenzierung hingegen basiert auf der Spannungsänderung beim Durchschleusen von DNA durch eine Nanopore und ermöglicht die Analyse von Sequenzen nahezu beliebiger Länge.



▲ **Abb. 2:** Tools zur Sequenz-Analyse. In unserem Labor entwickelte Tools ermöglichen es, ohne bioinformatische Kenntnisse die Stärke von Sequenzierungen zu nutzen. CRISPRnano dient der Identifikation genomeditierter Zellen, indem es niedrigqualitative Nanopore-Sequenzdaten direkt unmittelbar vom Sequenziergerät mit einer Referenz abgleicht und so Mutationen ermitteln kann. Duesselpore ermöglicht fortgeschrittene Transkriptom-Analysen, z. B. differenzielle Genexpressions-Analysen, durchzuführen. Diese können ebenfalls unmittelbar nach einem Sequenzier-Lauf durchgeführt werden und ermöglichen dadurch eine nahtlose nutzerfreundliche Analyse.



wendung fluoreszenzmarkierter Nucleotide. DNA von ca. 300–500 bp Länge wird hierbei parallel in Millionen von Clustern anhand eines DNA-Vorlagestrangs synthetisiert. Nach jedem Zyklus detektiert ein hochempfindliches Mikroskop das jeweils eingebaute Nucleotid (**Abb. 1**). Anschließend wird mittels bioinformatischer Methoden die erhaltene Sequenz mit einer Vorlage verglichen. Durch die extreme Genauigkeit (lediglich 0,1 bis 1 % Fehlerrate) können so geringe Mutationsraten zuverlässig aufgedeckt werden. Dies ist besonders bei geringen Editierungseffizienzen ausschlaggebend. Entscheidende Nachteile sind die hohen Geräteanschaffungskosten, die Notwendigkeit bioinformatischer Kenntnisse, und ein entsprechend leistungsstarker Computer für die Analyse. Weitere Methoden zur Sequenzierung durch Synthese sind Ion-Torrent-NGS und die von PacBio angewandte *single molecule real time* (SMRT)-Methode. Während bei Ion-Torrent-NGS eingebaute Nucleotide über eine Veränderung des pH-Werts ausgelesen werden, können mittels SMRT lange DNA-Fragmente (im Schnitt bis zu 15.000 bp) über den Einbau fluoreszenzmarkierter Nucleotide identifiziert werden. Da beide Technologien allerdings nicht die Relevanz von Illumina-NGS besitzen, seien sie hier nur der Vollständigkeit halber erwähnt. Seit kurzem gibt es zudem einen weiteren Akteur, der für gehörige Bewegung im NGS-Bereich sorgt: Die Nanopore-basierte Sequenzierung, auf die im folgenden Abschnitt genauer eingegangen wird.

Nanopore-Sequenzierung

Bei der Oxford-Nanopore-Technologies (ONT)-basierten Sequenzierung können DNA-Fragmente (theoretisch) beliebiger Länge sequenziert werden. Außerdem ist diese Technologie nicht auf die Synthese angewiesen, was simultan die Analyse von Nucleotid-Modifikationen, etwa epigenetischer Natur, ermög-

◀ **Abb. 3:** Duesselpore ermöglicht das Auslesen von Sequenzdaten ohne bioinformatische Datenprozessierung. **A–B,** Im dargestellten Beispiel wurden Nanopore-Sequenzdaten ausgewertet. **C–D,** Diese zeigten, dass kohlenstoffhaltige Partikel (CNP) aus der Umgebungsluft gegenüber einer Kontrolle (CAC) das zystische Fibrose assoziierte CFTR-Gen herunterregulieren (C) und oxidative Stress-Markergene modulieren (D). Abbildung aus [7]. Reproduzierbarkeit gewährleistet unter einer Creative Commons Attribution 4.0 International License.

licht. Die ONT-Methode basiert auf der Translokation von Nukleinsäuren mittels eines Motorproteins durch eine Nanopore und gleichzeitiger Messung elektrischer Signale, die charakteristisch für einzelne die Pore durchlaufende Nukleotide sind (**Abb. 1C**). Das Übersetzen der elektrischen Signale in Nukleotide erfolgt in Echtzeit und ermöglicht eine dynamische Versuchsanordnung. Als weitere Vorteile sind zudem eine geringe Grundinvestition (ab ~ 1000 €) und die Mobilität der Geräte zu nennen. So weist das kleinste eigenständig betriebsfähige Gerät etwa die Größe einer Brieftasche auf. Auch die zu Beginn recht hohe Fehlerquote ($\geq 10\%$) wurde stetig verbessert und ermöglicht inzwischen teils nahezu identische Genauigkeitsraten zu den zuvor besprochenen Systemen.

Identifizierung genomeditierter Zellen

Die Sequenzierung kurzer DNA-Fragmente mittels Synthese stellt den Goldstandard zur Identifizierung genomeditierter Zellen dar, wenn hunderte von Zellklonen simultan analysiert werden. Daher existiert eine große Zahl an freien Softwaretools zur Auswertung der generierten Sequenzdaten. Allerdings ist die Geräte-Grundanschaffung mit mehreren Zehntausend Euro hoch. Die geringen Grundanschaffungskosten von ONT-Geräten (s. o.) wurden jedoch gerade zu Beginn durch ihre hohe Fehleranfälligkeit erkauft. Im Zusammenspiel mit ihrer relativen Neuheit besteht für ONT-Sequenzierung somit Nachholbedarf, um die einfache Identifizierung genomeditierter Zellen zu ermöglichen. Aus diesem Grund entwickelten wir einen Workflow mit zugehöriger Software, der es erstmals ermöglicht, kostengünstige ONT-Sequenzierung zur Identifizierung genomeditierter Zellen einzusetzen. CRISPRnano (**Abb. 2A**, [4]) benutzt einen Smith-Waterman-Algorithmus zur Sequenz-Zuordnung, der es ermöglicht, niedrigqualitative ONT-Sequenzdaten, etwa durch erlaubte Lücken an den Enden eines PCR-Produkts, schnell, zuverlässig und effizient zu identifizieren, quantifizieren und visualisieren. CRISPRnano benötigt keine Internet-Verbindung oder hochleistungsfähige Computer-Hardware und unterstützt somit die breite Nutzung einfacher, kostengünstiger Sequenzierung durch jedes Labor.

Weitere NGS-Anwendungen

Zusätzlich zur Identifizierung genomeditierter Zellen bietet NGS vielfältige weitere

Anwendungen. Dazu gehört etwa die Identifizierung genomischer Mutationen, die besonders in der personalisierten Medizin und Diagnostik von unschätzbarem Wert ist. Weiterhin ist die Identifizierung von Varianten pathogener Keime, z. B. SARS-COV2 zu nennen, die es beispielsweise ermöglicht nachzuvollziehen, welche Varianten zu einem bestimmten Zeitpunkt besonders stark in der Bevölkerung kursieren. Studien zur epigenetischen Gen-Regulierung, wie die Analyse von DNA- oder RNA-Methylierung, sind ein weiteres wichtiges Anwendungsfeld für Sequenzierungs-Technologien. Hier bietet ONT-Sequenzierung einen inhärenten Vorteil, da Modifikationen ohne weitere Manipulation der Nukleinsäuren analysiert werden können. Ein viertes Beispiel stellt das Monitoring der Integrität mitochondrialer DNA (mtDNA) dar, deren Analyse eine spezielle Herangehensweise voraussetzt [5].

Integritätsanalyse mitochondrialer DNA

Im Gegensatz zur Analyse der nukleären genomischen DNA mittels Genom-Sequenzierung verlangt die Analyse der zirkulären, multiploiden mtDNA einen speziellen Ansatz. So gibt es eine erhebliche Anzahl nukleärer mitochondrialer Sequenzen, die Sequenzzuordnungen erschweren. Dies erfordert die rigorose Daten-Präprozessierung. Da bereits ein geringer Anteil an Mutationen in mtDNA, eine Heteroplasmie, ausreichen kann, ein Krankheitsbild zu manifestieren, ist es von enormer Wichtigkeit, ihre Integrität in Patienten nachzuweisen. Auch hier existiert ein Mangel an einfach durchzuführenden

Analysen, weshalb wir aktuell eine kombinierte Lösung aus Arbeitsablauf und Software entwickeln, die sich diesem Problem widmet.

Genexpressionsanalysen

Ein weiterer Hauptanwendungsbereich für NGS sind Transkriptom-Analysen. Hierbei wird ausgehend von der mRNA, die zum Entnahmezeitpunkt X die Menge aller exprimierten Gene einer Probe darstellt, eine Transkript-Bibliothek erstellt, die anschließend sequenziert wird. Bioinformatisch werden dann die erhaltenen Sequenzen identifiziert und ausgezählt. Dadurch können unter anderem aktivierte Signalkaskaden als Reaktion auf eine Behandlung oder negative Umwelteinflüsse identifiziert werden. Die Entdeckung neuartiger Marker-Gene für Krankheiten oder physiologische Prozesse ist ein weiteres Anwendungsgebiet. Bis vor wenigen Jahren war NGS jedoch nicht umfassend zugänglich und für die Analyse die Hilfe eines Bioinformatikers nötig. Hier

Hier steht
eine Anzeige.

 Springer

bietet sich wiederum eine Nische für ONT-Sequenzierung, da eine extrem niedrige Fehlerquote für die Zuordnung langer DNA-Sequenzen zu Transkripten im Normalfall nicht essenziell ist. Um solche Sequenzdaten auszuwerten, sind aber auch hier bioinformatische Kenntnisse unabdingbar. Um diese Lücke zu schließen und Transkriptom-Analysen massentauglich zu machen, entwickelten wir Duesselpore (**Abb. 2**). Duesselpore ist eine lokal ohne Internetverbindung laufende Software, die es ermöglicht, ohne jegliche bioinformatische Kenntnisse ONT-Transkriptom-Sequenzdaten auszuwerten und grafisch darzustellen [6]. Duesselpore bietet automatische differenzielle Genexpressions- und Signalweg-Analysen, um Unterschiede zwischen experimentell behandelten Proben aufzudecken und ihre Bedeutung einzuordnen (**Abb. 3**, [7]). Auf diese Weise ermöglicht es Duesselpore, die Portabilität von ONT-Geräten voll auszunutzen und Transkriptom-Analysen ortsungebunden durchzuführen.

Fazit

Zusammenfassend lässt sich sagen, dass NGS einen essenziellen Bestandteil der Analyse-Batterie in der wissenschaftlichen Forschung darstellt. Mit unserer Arbeit tragen wir dazu bei, dass eine breite Nutzerschaft von den vielfältigen Möglichkeiten profitieren kann. Denn nicht jedes Labor kann es sich leisten, teures Equipment anzuschaffen und einen Bioinformatiker zu beschäftigen. Auf diese Weise unterstützen wir den wissenschaftlichen Fortschritt, denn seit jeher trägt eine Demokratisierung der Forschungs-

landschaft maßgeblich zur entdeckenden Kraft der wissenschaftlichen Gemeinschaft bei.

Literatur

- [1] Pakari K, Wittbrodt J, Thumberger, T (2023) CRISPR-Fortschritte – Schnitt für Schnitt zu neuen Möglichkeiten. *BIOspektrum* 29: 25–28
- [2] Dobner J, Ramachandran H, Rossi A (2022) Genome Editing in Translational Medicine: An Inventory. *Front Bioscience-Landmark* 27: 241
- [3] Aigrain L (2021) Beginner's guide to next-generation sequencing. *Biochem (Lond)* 43: 58–64
- [4] Nguyen T, Ramachandran H, Martins S et al. (2022) Identification of genome edited cells using CRISPRnano. *Nucleic Acids Res*, DOI: 10.1093/nar/gkac440
- [5] Rossi A, Lickfett S, Martins S, Prigione A (2022) A call for consensus guidelines on monitoring the integrity of nuclear and mitochondrial genomes in human pluripotent stem cells. *Stem Cell Rep* 17: 707–710
- [6] Vogeley C, Nguyen T, Woeste S et al. (2022) Rapid and simple analysis of short and long sequencing reads using DuesselporeTM. *Front Genet* 13: 931996
- [7] Stermann T, Nguyen T, Stahlmecke B et al. (2022) Carbon nanoparticles adversely affect CFTR expression and toxicologically relevant pathways. *Sci Rep* 12: 14255

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. Jochen Dobner
 Genome Engineering and Model Development
 IUF – Leibniz-Institut für umweltmedizinische
 Forschung
 Auf'm Hennekamp 50
 D-40225 Düsseldorf
 Jochen.Dobner@IUF-Duesseldorf.de

AUTOREN



Jochen Dobner

2010–2016 Studium der Molekularen Zell- und Entwicklungsbiologie an der Universität Innsbruck, Österreich. 2021 Promotion an der Universität Düsseldorf. 2021–2022 PostDoc bei Prof. Dr. D. Willbold, Universität Düsseldorf. Seit 2022 PostDoc bei Dr. A. Rossis Lab, IUF – Leibniz-Institut für umweltmedizinische Forschung.



Andrea Rossi

1999–2005 Studium der Pharmazie, Universität Bari, Italien. 2009 Promotion in Biochemie und Molekularbiologie an der Universität Bari. 2010–2012 PostDoc in Alan Verkman's Labor an der University of California, San Francisco, USA. 2012–2018 PostDoc bei Prof. Dr. D. Stainiers, Max-Planck-Institut für Herz- und Lungenforschung, Bad Nauheim. Seit 2019 Arbeitsgruppenleiter des Genome Engineering and Model Development Labors, IUF – Leibniz-Institut für umweltmedizinische Forschung.