

# dem posttranskriptionalen regulatorischen code auf der spur

## doRiNA: die BIMS-B-Datenbank für posttranskriptionale Kontrollelemente

von Markus Schüler und Christoph Dieterich

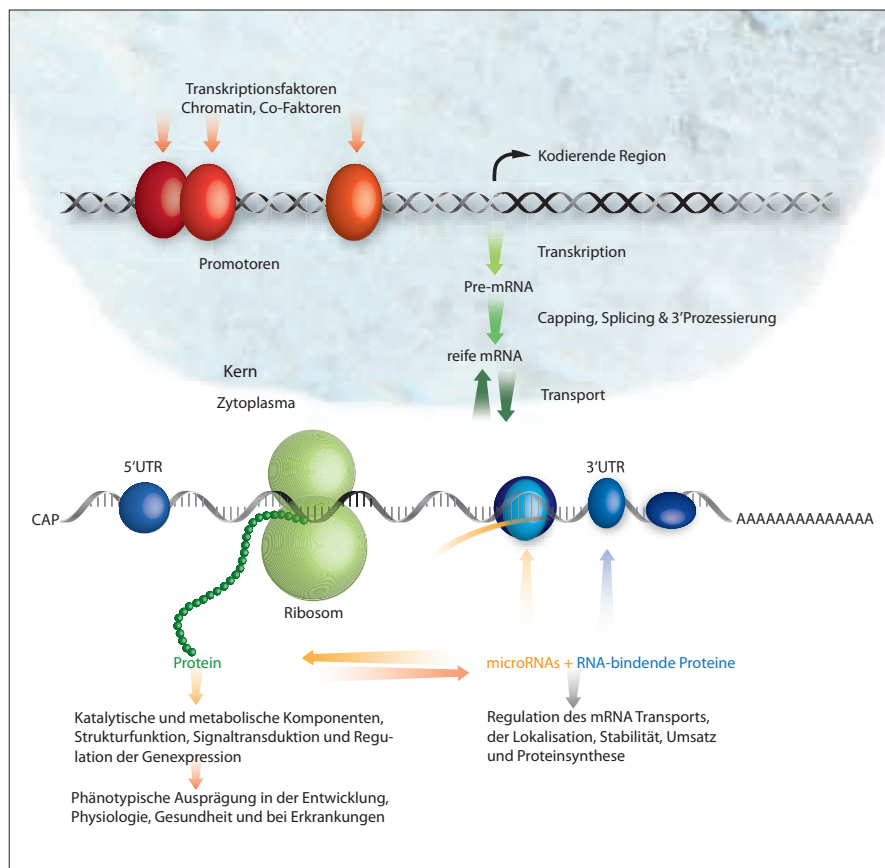
Das Genom, der Bauplan des Lebens, muss immer wieder neu ausgelesen und interpretiert werden, um unseren Organismus an die Herausforderungen einer sich wandelnden Umwelt anzupassen. Die genetische Information wird im Zellkern gespeichert und abgelesen, prozessiert und als Botenmoleküle (mRNAs) in das Cytoplasma exportiert. Dort dienen die mRNAs als Vorlage für die Synthese von Proteinen. mRNAs sind auf ihrem Weg zu den Protein-fabriken vielen Kontroll- und Verarbeitungsschritten unterworfen. Die „doRiNA“-Datenbank (Anders *et al.*,

2012) des Berlin Institute for Medical Systems Biology (BIMSB) ermöglicht nun erstmals eine umfassende Untersuchung dieser Vorgänge, die unter dem Namen „Posttranskriptionale Genregulation“ zusammengefasst werden.

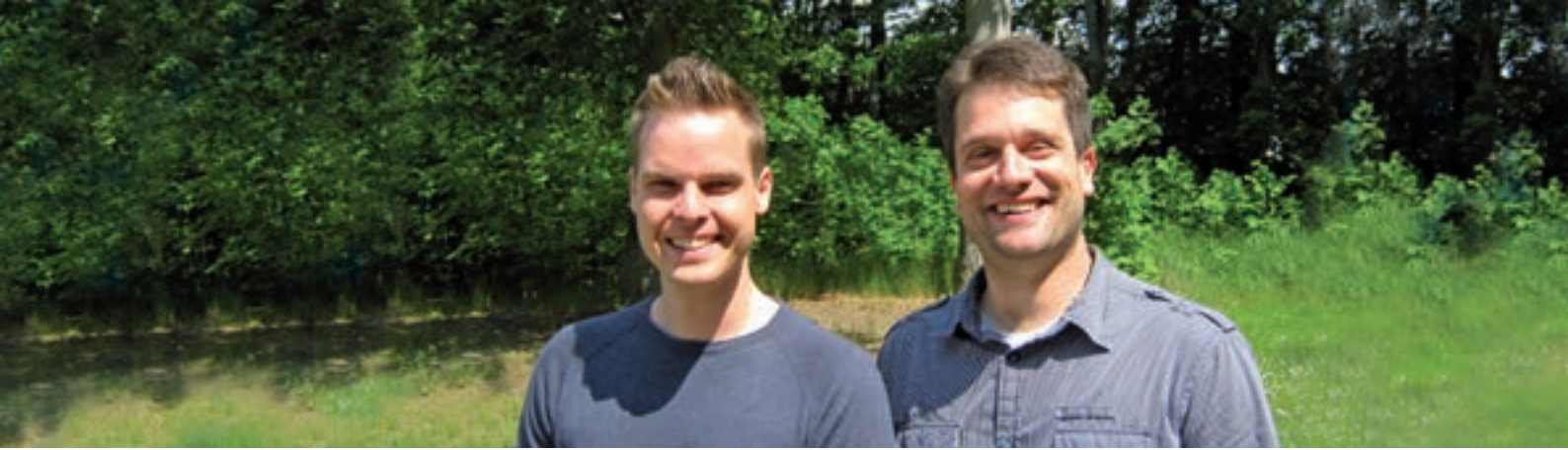
### Posttranskriptionale Genregulation: Ein Code aus Buchstaben, Molekülen und Interaktionen

Die Aktivität unserer Gene wird an spezifischen Stellen gesteuert und kontrolliert. Getreu dem zentralen Dogma der Molekularbiologie werden Gene kontextabhängig ausgelesen

Abbildung 1: Lebenslauf einer mRNA



Nach Transkription der genetischen Information in RNA durchläuft diese mehrere Modifizierungsschritte zur finalen mRNA. Diese wird wiederum von Ribosomen in Proteine translatiert. All diese Schritte werden durch das Zusammenspiel von RNA-bindenden Proteinen und microRNAs reguliert (© BIMSB/MDC).



Markus Schüler (links) und Christoph Dieterich (rechts) aus der BIMSB Bioinformatics/Mathematical Modelling-Gruppe am MDC Berlin (Foto: Martin Siegert)

(Transkription), d. h. von der DNA- auf die RNA-Ebene überführt. Diese RNA-Abschriften werden prozessiert und nach einem „Reifungsprozess“ als mRNA aus dem Zellkern exportiert. Die diverse Population von mRNA-Molekülen kann nun in den Proteinfabriken der Zelle, den Ribosomen, in Proteine überführt werden (Translation), die dann wiederum beispielsweise als Enzyme stoffwechselrelevante Funktionen wahrnehmen. Das Forschungsfeld der posttranskriptionalen Genregulation untersucht all jene Regulationsschritte, die zwischen der Transkription des Gens und der Translation der mRNA in ein Protein stattfinden. Praktisch alle Vielzeller im Tier- und Pflanzenreich verfügen über umfangreiche Kontrollmechanismen auf posttranskriptionaler Ebene. Diese steuern die Verfügbarkeit von mRNA für die Translation im Raum (also die Lokalisierung der mRNA in der Zelle), als auch in der Zeit (also wann eine mRNA in ein Protein übersetzt wird). Einen Gesamtüberblick über den Lebenslauf einer mRNA gibt die Abbildung 1. Die Lebensdauer, der Transport sowie die Bevorratung und Translation von mRNAs wird durch ihre Wechselwirkung mit RNA-bindenden Proteinen (RBPs) und anderen RNAs bestimmt. Zu diesen RNAs gehören vor allem die microRNAs, kleine nicht kodierende RNAs mit einer Länge von durchschnittlich 22 Nukleotiden, die eine entscheidende Rolle bei der posttranskriptionalen Genregulation spielen, indem sie zum Abbau bestimmter mRNAs oder der Unterbindung der Translation führen.

Als RBPs werden all jene Proteine bezeichnet, die RNA binden können. Diese Bindung kann ursächlich zu einer Reaktion des Proteins mit der RNA führen oder aber Teil einer zuvor beschriebenen Kontrollfunktion sein. Vielfach untersuchte Beispiele sind die Interaktion zwischen microRNAs und Argonautenproteinen, evolutionär stark konservierte Proteine, die eine wichtige Rolle in der Prozessierung von mRNAs spielen. Die Kombination aus Argonautproteinen, microRNAs und gegebenenfalls weiterer Kofaktoren bindet spezifisch an vorhandenen

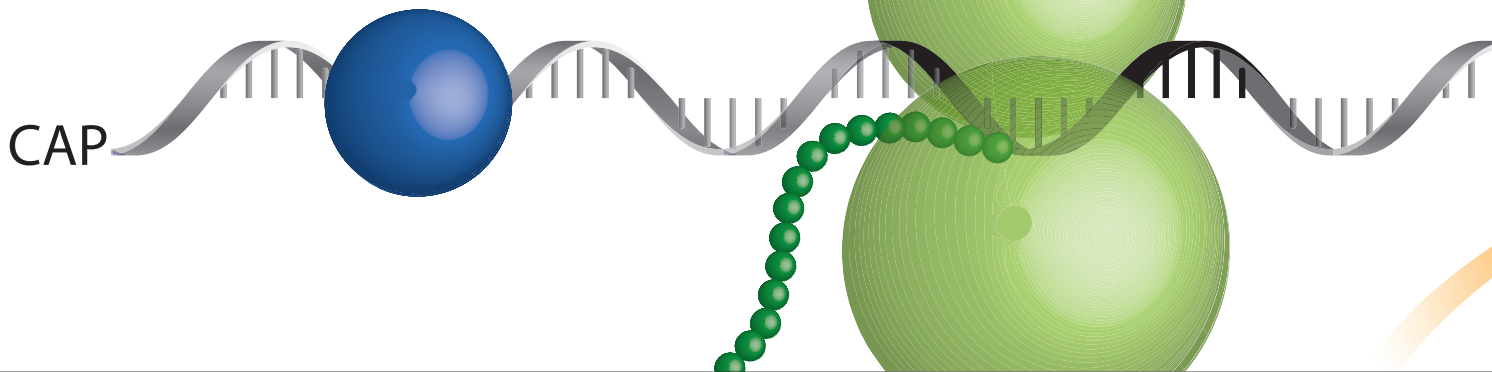
mRNAs und führt zu deren Abbau oder zu einer Unterbindung der Translation (Bartel, 2009). Viele Wissenschaftler gehen heute von einem generellen Zusammenspiel der Bindung von RBPs und microRNAs aus, um die richtige Menge Protein zur richtigen Zeit und am richtigen Ort zu synthetisieren. Dieses Zusammenspiel, oft auch als der „posttranskriptionale regulatorische Code“ (Keene 2007) bezeichnet, steht im Fokus der Forschung vieler Arbeitsgruppen weltweit.

Die Arbeitsgruppe Dieterich am Berlin Institute for Medical Systems Biology (BIMSB) hat nun vor kurzem in Kooperation mit den Arbeitsgruppen Landthaler und Rajewsky eine umfassende Webapplikations- und Datenbanklösung präsentiert, die die Ergebnisse aus diesen Forschungsanstrengungen bündelt und so einen umfassenderen Blick bei der Suche nach dem regulatorischen Code ermöglicht: Die „database of RNA interactions in post-transcriptional regulation“, kurz doRiNA (<http://dorina.mdc-berlin.de>). Die doRiNA-Datenbank führt Informationen über spezifische miRNA- und RBP-Bindungsstellen aus verschiedenen experimentellen und computergestützten Ansätzen zusammen. Momentan stehen Daten aus der Zellkultur (Maus und Mensch) sowie Modellorganismen wie dem Fadenwurm (*Caenorhabditis elegans*) zur Verfügung. Weitere Organismen und Datensätze werden folgen. Über die Zeit entsteht so eine immer dichter werdende Karte des posttranskriptionalen Netzwerks.

### doRiNA, die BIMSB-Datenbank für posttranskriptionale regulatorische Elemente

Das Ziel bei der Entwicklung von doRiNA war größtmögliche Flexibilität bei komfortabler Benutzbarkeit und gleichzeitig hoher Geschwindigkeit. Zu diesem Zweck wurde doRiNA aus zwei Komponenten aufgebaut, die eng miteinander verknüpft sind.

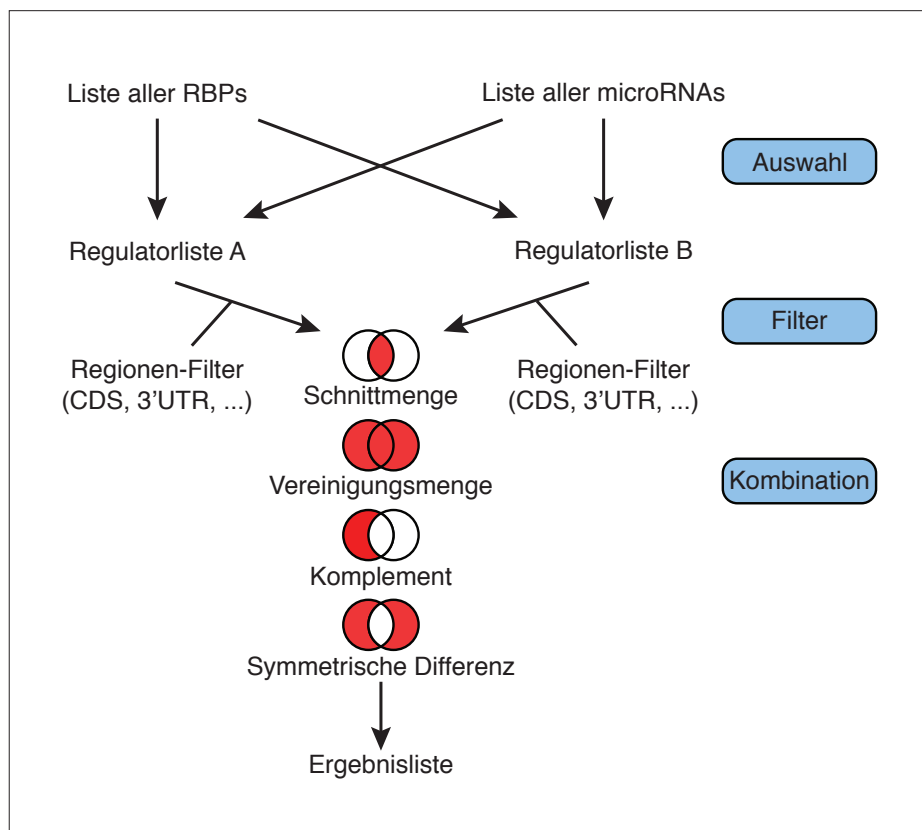
Auf der einen Seite steht die Webapplikation, also der Teil, den man im Browser sehen kann. Dieser ermöglicht es dem



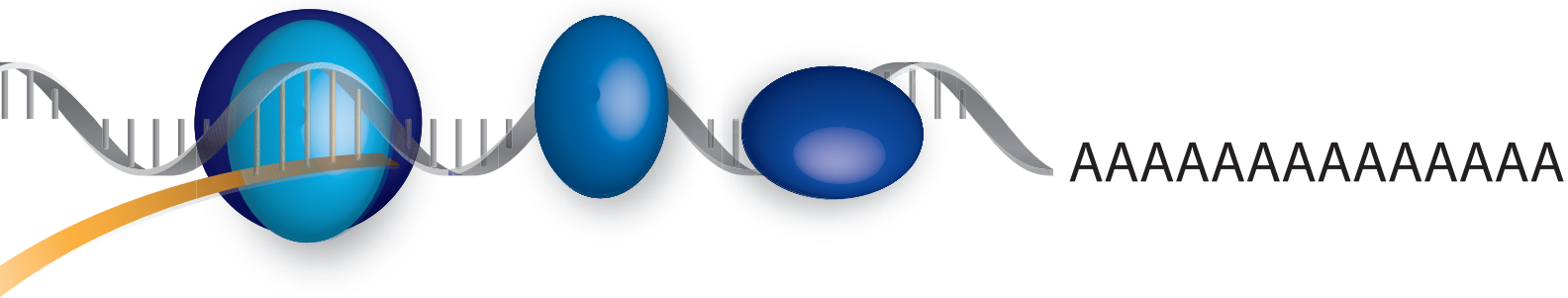
Benutzer, seine Suche auf jenen Teil des regulatorischen Netzwerkes einzuschränken, der ihn interessiert. Hierzu gibt es die Möglichkeit, Anfragen aus der Perspektive der mRNA zu stellen („Welche microRNAs/RBPs binden an eine spezifische mRNA?“) sowie aus der Sicht der Regulatoren („Welche mRNAs werden von einer bestimmten microRNA oder einem bestimmten RBP reguliert?“). Um dies so einfach wie möglich zu gestalten und trotzdem auch komplexere Abfragen zu ermöglichen, wurde die Oberfläche in einen vereinfachten „Simple search“-Bereich mit eingeschränkter Wahlmöglichkeit sowie einen erweiterten „Combinatorial search“-Bereich mit vollständiger Funktionalität unterteilt. Ersterer nimmt dem Benutzer die meisten Entschei-

dungen ab und erlaubt damit einen unkomplizierten Einstieg in die Analyse, während die komplexe Suche weitere Möglichkeiten bietet, den Suchraum noch genauer zu bestimmen. Hierzu gehört die Beschränkung auf Bindungsstellen in spezifischen Bereichen der mRNA (z. B. die kodierende Sequenz, Introns, 3'UTRs und 5'UTRs) sowie die Verknüpfung zweier unabhängiger Suchanfragen mittels verschiedener Mengenoperationen. Abbildung 2 zeigt eine Veranschaulichung der vielfachen Anfragemöglichkeiten, die mit der komplexen Suche möglich sind. Die Formulierung einzelner Anfragen wird zusätzlich unterstützt durch fortgeschrittene Web2.0-Merkmale wie der automatische Vervollständigung bei der Eingabe, der Prüfung der korrekten Schreibwei-

Abbildung 2: Diagramm der Einstellungsmöglichkeiten der komplexen Suche



Ausgehend von der kompletten Liste an RBPs und microRNAs definiert der Benutzer zwei Submengen von Regulatoren. Diese können durch zusätzliches Filtern auf Bindungsstellen in bestimmten Bereichen der mRNA reduziert werden. Als letzten Schritt werden die beiden Mengen mittels vom Benutzer gewählter Mengenoperation miteinander verrechnet und das Ergebnis ausgegeben (Bild: Aus Anders *et al.*, 2012, Nucleic Acids Res., Oxford University Press).



se, der eingabespezifischen Aktivierung von Schaltflächen sowie einer umfangreichen Hilfefunktion. Nur Anfragen, die als korrekt erkannt worden sind, werden schließlich auch ausgeführt.

Die zweite an die Webapplikation angekoppelte Komponente von doRiNA ist die zugrundeliegende Datenbank und der Webserver. Geschwindigkeit wird hier vor allem durch die Integration schneller Computersysteme erreicht, jedoch auch durch ein speziell auf die Daten zugeschnittenes Abfragesystem, welches einen optimierten Zugriff auf die Daten erlaubt, um auch lange Ergebnislisten und komplexe Anfragen schnell bearbeiten zu können. Die Ausgabe, die für den Benutzer generiert wird, ist schließlich eine Liste der gebundenen mRNAs bzw. der Regulatoren einer oder mehrerer mRNAs. Diese Listen beinhalten eine Vielzahl von Informationen, wie den Namen der zu den mRNAs zugehörigen Genen, deren Regulatoren, die genaue Positionen der Bindungsstellen, usw. Der Benutzer kann die Tabelle nach all diesen Kriterien sortieren, um so die für ihn wichtigen Ergebnisse sofort angezeigt zu bekommen. Zur weiteren Analyse der Ergebnisse ist doRiNA vollständig mit einer lokalen Kopie des UCSC-Genome-Browsers (<http://genome.ucsc.edu>) verknüpft. Der UCSC-Genome-Browser ermöglicht eine hochauflösende Darstellung einzelner Bindungsstellen und ihres genomischen Kontexts (Abb. 3). Der UCSC-Genome-Browser ist unter Biologen auf der ganzen Welt aufgrund seiner einfachen Bedienbarkeit hoch geschätzt.

Der Vorteil, der durch die Vernetzung doRiNAs mit dem UCSC-Genome-Browser erreicht wird, ist vor allem, dass Biologen sich nicht an ein neues System gewöhnen müssen, was die Akzeptanz erhöht. Gleichzeitig können die Daten zur posttranskriptionalen Regulation aus doRiNA in einen größeren Kontext gestellt werden, da der UCSC-Genome-Browser bereits viele öffentlich zugängliche Datensätze enthält (z. B. Einzelnukleotid-Polymorphismen oder SNPs, Bindemotive, Expressionsdaten usw.). Für Benutzer, die ihre Ergebnisse darüber hinaus selbstständig weitergehend analysieren wollen, besteht zudem die Möglichkeit, diese als Tabellen auf den eigenen Rechner zu laden.

### Keine Datenbank ohne Daten

Eine wichtige Frage, die bisher noch nicht angesprochen wurde, ist die nach dem Ursprung der in doRiNA integrierten Bindungsstellen. Hierbei muss zwischen RBP- und microRNA- Bindungsstellen unterschieden werden.

Die in doRiNA integrierten RBP-Bindungsstellen basieren alle auf „Cross-Linking-Immunopräzipitations“- oder kurz „CLIP“-Experimenten. Bei dieser Technik wird das RNA-bindende Protein chemisch an die mRNA gebunden. Danach werden die gebundenen mRNA-Fragmente in cDNA überführt und sequenziert. Über den Sequenzierungsschritt können die Bindungsstellen von RBPs kartiert werden. Eine neue Variante von CLIP-Experimenten, die sog. „Photoaktivierbaren Ribonukleosid-verbesserten CLIPs“ (PAR-CLIPs, Hafner *et al.*, 2010), ermöglichen dabei eine noch genauere Bestimmung von einzelnen Bindungsstellen. Dies wird durch den Einsatz von „photoaktivierbarer Ribonukleoside“ erreicht. Diese werden in die mRNA integriert und führen dort zu charakteristischen Mutationen bei der Sequenzierung. Die Einlagerung des Moleküls „4-Thiouridin“ führt überall dort zum Austausch von T- zu C-Basen, wo das untersuchte RBP bindet. Durch diese charakteristische Mutation kann die Sequenz einer Bindungsstelle leicht von Sequenzen anderen Ursprungs unterschieden werden. Die zurzeit in doRiNA integrierten Datensätze basieren vor allem auf am BIMSMB generierten PAR-CLIP Experimente. Publierte Datensätze werden sukzessive eingepflegt.

Im Gegensatz zu RBPs gibt es für microRNAs bisher noch keine effiziente *in vivo*-Methode zur experimentellen Bestimmung von Bindungsstellen. Allerdings lassen sich einige wesentliche Eigenschaften von microRNAs zur computergestützten Vorhersage von Bindungsstellen nutzen. In doRiNA wurde das von der Arbeitsgruppe Rajewsky entwickelte Programm „PicTar 2.0“ (Lall *et al.*, 2006) integriert. Das Programm nutzt beispielsweise die mRNA-Sequenzen vom Menschen und vergleicht diese mit anderen nah und weiter entfernt verwandten Organismen, um falsch-positive Vorhersagen herauszufiltern. Damit der Nutzer selber entscheiden kann, wie sehr er sich auf diesen Vergleich

verlassen möchte, stellt doRiNA die Möglichkeit bereit, die maximale evolutionäre Distanz zu wählen (Vergleiche innerhalb der Säugetiergenome, bis zum Hühnergenom oder sogar bis zu den Fischen). Das Ergebnis ist wiederum eine Liste von vorhergesagten Bindungsstellen, die der Benutzer durchsuchen oder im UCSC-Genome-Browser betrachten kann.

DoRiNA ist als Komplettlösung für die Untersuchung posttranskriptionaler Experimente im Kontext eines regulatorischen Codes gedacht. Schon jetzt fungiert doRiNA als Knotenpunkt zur Integration der Ergebnisse der verschiedenen BIMS-Arbeitsgruppen. Um die Aktualität der Daten zu gewährleisten, werden zukünftig eine Vielzahl weiterer Experimente sowohl von BIMS Forschern als auch externer Forschungsgruppen in die Datenbank aufgenommen werden. Ziel ist mit den so bereitgestellten Daten sowie der Entwicklung weiterer Verbesserungen, z.B. einer direkten Visualisierung der regulatorischen Netzwerke, unser Wissen über das Zusammenspiel von RNA-bindenden Proteinen und microRNAs weiter auszubauen und die Grundlage für neue Erkenntnisse im Bereich der posttranskriptionalen Genregulation zu schaffen.

### Steckbrief Forschungsprojekt:

DoRiNA, die Datenbank für posttranskriptionale regulatorische Elemente, wird entwickelt am Berlin Institute of Medical Systems Biology (BIMS) durch die Gruppen „Bioinformatics in Quantitative Biology“ (Gruppenleiter: Christoph Dieterich) in Zusammenarbeit mit Markus Landthaler (PAR-CLIP-Experimente) und Nikolaus Rajewsky (PAR-CLIP-Analyseprogramme, PicTar). Das Projekt wird durch Zuwendungen vom MDC Systems Biology Network (MSBN) als Teil der Helmholtz-Allianz für

Systembiologie, dem Bundesministerium für Bildung und Forschung und dem Senat von Berlin gefördert (0315362A).

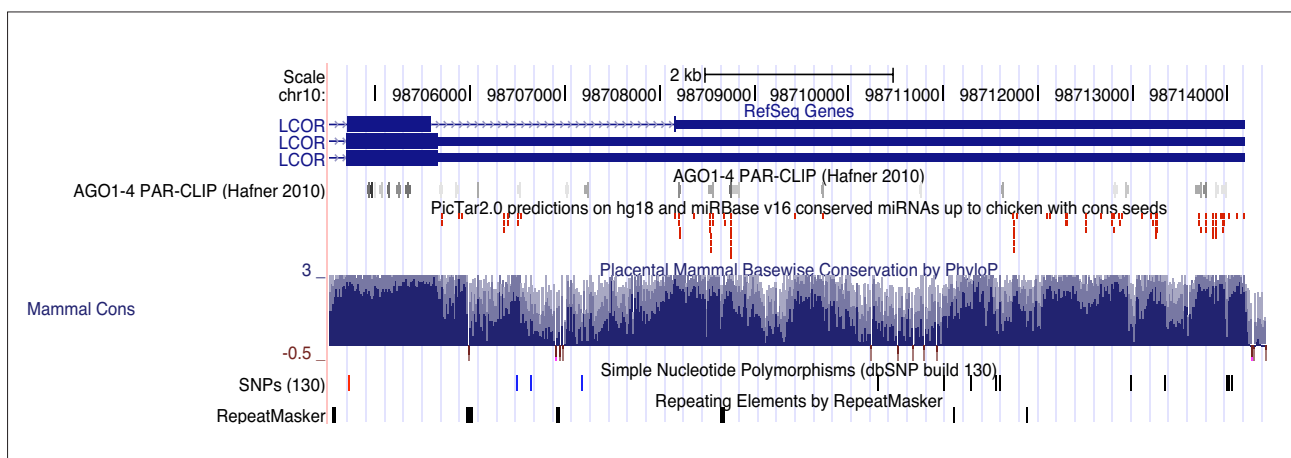
### Referenzen:

- Anders, G., Mackowiak, S.D., Jens, M., Maaskola, J., Kuntzagk, A., Rajewsky, N., Landthaler, M., Dieterich, C. (2012). doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* 40, D180-D186
- Bartel, DP. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell* 136, 215–233.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.* 8, 533–543.
- Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y.-L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., Macmenamin, P., *et al.* (2006). A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* 16, 460–471.

### Kontakt:

**Dr. Markus Schöler und Dr. Christoph Dieterich**  
 Bioinformatics in Quantitative Biology / Berlin Institute for Medical Systems Biology (BIMS)  
 Max-Delbrück-Centrum für Molekulare Medizin (MDC)  
 Berlin-Buch  
 christoph.dieterich@mdc-berlin.de

Abbildung 3: Ausgabe einer doRiNA-Abfrage mit dem UCSC-Genome-Browser



Das Bild zeigt den 3'UTR-Bereich des Gens „LCOR“ sowie die experimentell verifizierten Bindungsstellen der Argonautenproteine 1-4 („AGO 1-4“) und bioinformatische Vorhersagen von microRNA-Bindungsstellen („PicTar2.0“). Zusätzlich sind weitere Annotationen gezeigt, wie das Auftreten von SNPs, DNA-Konservierung („Mammal Cons“) und repetitiven Regionen („RepeatMasker“) (Bild: www.genome.ucsc.edu).