**Proteome-wide screen for RNA-dependent proteins**

**Warum ist es eigentlich interessant, was wir hier auswerten?**

RNA ist nicht gleich RNA. Bild von Transkription und auch Translation hat jeder vor Augen, und man erinnert sich an einige daran beteiligte Proteine, die sich zum Beispiel an die mRNA anlagern. Aber die Interaktionsfelder von Proteinen an RNA gehen noch deutlich weiter, eine RNA ohne die Beteiligung von Proteinen wäre nutzlos. Daher ist es von Vorteil zu wissen, welche Proteine RNA-bindend sind, und auch zu verstehen, anhand welcher Strukturen solch eine Bindung erfolgt.

Es gibt solche Proteine mit RNA-binding-domains = RBDs (z.B. RNA recognition motif: RRM, K-homolgy: KH, etc.), ein großer Teil enthält aber keine RBDs, sondern unorganized regions.

Diese Proteine ohne RBDs haben häufig vielerlei zelluläre Funktion ohne Verbindung zum RNA-Metabolismus. Sie sind Kandidaten für Proteine, die nicht RNA regulieren, aber von RNA reguliert werden. Dies nennt man auch Riboregulation. Riboregulation ist das Komplementäre, auch RNA beeinflusst nämlich die Aktivität, Konformation und Positionierung von Proteinen. Man sieht also: Eng verflochtenes Netzwerk, das Schicksal jeder RNA hängt wesentlich von Proteinen ab. RBPs interagieren mit RNA über bestimmte Sequenzen und Strukturen, allerdings gibt es auch welche, die keine speziellen Motive haben.

Diese Interaktionsfelder, in der Proteine die RNA beeinflussen sind bspw: Splicing, Transport, Modifikation, Stabilität, Translation, Abbau. Betrifft übrigens nicht nur mRNA, sondern auch non-coding RNA. Auch die RBPs untereinander regulieren sich (Kooperation vs. Antagonismus). RNAs bilden zusammen mit RBPs über starke Interaktionen (RBDs) und schwache Interaktionen (intrinsically disorganized regions) sogenannte membrane-less-Organellen, die sich dynamisch verändern.

Kleiner Ausblick: RBPs haben weitreichende Funktionen, ihre Missregulation ist häufig mit Krankheiten assoziiert, neurologische Krankheiten oder, wenn die Beeinflussung von lncRNA gestört wird, Krebs. Auch interessanter Hebel für die sehr schnelle Steuerung der Genexpression, wenn man die Aktivität der RBPs moduliert.

**Welche Zellen verwenden wir, und warum sind die so gut geeignet?**

Wir verwenden nicht-synchronisierte HeLa-Zellen. Diese sind gut für allgemeine Zellbiologieexperimente, da sich die Zellen in allen möglichen Stadien des Zellzyklus befinden, und somit repräsentativer für biologische Prozesse sind. Für unsere Untersuchungen spielt es noch eine weitere Rolle, denn wir wollen RNA-dependent Proteine finden, die Zusammensetzung der RNPs ist aber dynamisch, auch in den Phasen des Zellzyklus. Nicht-synchronisierte Zellen zu nehmen, bietet uns also ein breiteres Bild und erfasst mehr in Frage kommende Proteine.

**Alternative eingesetzte Methoden zur Identifizierung von RBPs?**

CLIP: Crosslinking RNA und Protein -> Immunoprezipitation -> RNase digestion -> Sequencing des RNA-Protein gebundenen Stücks -> Herausfinden des genauen Bindemotives

**Das R-DeeP-Verfahren – Besonderheiten**

Bisheriger Stand der Forschung hat RNA-bindende Proteine betrachtet, die sich also wirklich direkt an die RNA anlagern. Dies birgt diverse Nachteile: Bias von Aufreinigungsschritten und dafür notwendiges Finden von Unterschieden zwischen RBP und non-RBP, eventuell Crosslinking von Proteinen und RNA.

Hier ist das R-DeeP-Verfahren eine Neuerung. Es betrachtet nämlich alle RNA-dependent Proteine. Ein Protein ist RNA-dependent, wenn sein Interaktom von RNA abhängig ist, also wenn es entweder direkt oder indirekt mir RNA in Kontakt tritt. Hierbei ist wichtig zu erwähnen, dass diese RNPs sehr komplexe, große und dynamische Komplexe sind, Bisherige Methoden haben quasi nur den Kern ergründet auf Basis direkter Interaktion, R-DeeP erweitert die Suche auf funktionelle Abhängigkeit, denn wenn das Protein-Interaktom abhängig ist von RNA, dann wahrscheinlich auch seine Funktion, auch wenn es nicht direkt an die RNA bindet.

R-Deep ist ein proteome-wide, unbiased und enrichment-free screen. Es ist in zwei Punkten einzigartig: Es braucht kein Enrichment-procedure, und ist damit unabhängig von Unterschieden der Affinität oder physikochemischen Eigenschaften der zellulären Komponenten. Es ist auch unabhängig von RNA und Protein-Sequenzen der Interaktionsstellen. Außerdem liefert es quantitative Aussagen über den Anteil jedes RNA-dependent Proteins (!!!), der tatsächlich an die RNA-angelagert ist (der RNA-dependent Anteil). Man hat >500 humane Proteine als RNA-dependent identifiziert, zu denen davor keine Verbindung zu RNA bekannt war.

Limitationen von R-DeeP: Identifiziert nicht die binding site von RNA und Protein, unterscheidet auch nicht zwischen direkter und indirekter Interaktion. Bei Zelllyse könnten durch den Detergenz-beinhaltenden Puffer außerdem Interaktionen geschaffen werden, die in der Zelle natürlich nicht vorkommen.

Kategorien für die Klassifizierung von Proteinen: RNA-dependent, partially-RNA-dependent, RNA-independent (siehe mehr bei Shifts)

**Weitere mögliche Anwendungen des R-DeeP-Verfahrens?**

Anstatt das Gesamt-Zell-Lysat zu nehmen, kann man auch nur die Lysate von speziellen Kompartimenten testen.

Vor Protein-Lysate-Vorbereitung kann man der Zelle Inhibitoren zusetzen, und deren Einfluss auf Protein-RNA-Interaktionen betrachten.

Anstatt RNAse, kann man siRNA dazugeben, um so nur ganz spezielle RNAs zu betrachten

Anstatt RNA-dependency anzugucken, kann man auch DNA-dependency betrachten.

Mitose/Interphase-spezifische Untersuchungen der RNA-dependent proteins in bestimmten Zellzyklusstadien, und damit, wie sich RNPs dynamisch verändern, sodass es RNA-dependent Proteine gibt, die nur in ganz bestimmten Phasen identifiziert wurden, zum Beispiel wohl, weil sie für die mitotische Spindel eine Rolle spielten. RNA-dependent Proteine sind in der mitotischen Spindel vertreten, RNA und RBPs haben wohl eine Bedeutung für diesen Teil der Zellteilung, RNase-Behandlung kann die Spindel nämlich stören. RNA hat bedeutende Rolle in der Vermittlung von RNA-abhängigen Protein-Protein Interaktionen.

**Wie genau funktioniert das Verfahren von R-DeeP, und was genau sehen wir in unserem Datensatz?**

Ein Bild, das Text, Diagramm, Screenshot, Plan enthält.

KI-generierte Inhalte können fehlerhaft sein.

1. Bestimmte Zellzahl entnehmen

2. vorsichtige Zell-Lyse

3. BCA-Assay, um die Protein-Konzentration des Lysates zu bestimmen

4. Sucrose-Gradient vorbereiten, mit absteigend 5-50 % Sucrose-Lösung

5. Die beiden Proben oben auftragen. Probe 1: Kontrolle, unverändert, Probe 2: RNase-Probe, mit 5 verschiedenen RNasen behandelt.

Dies ist der wichtige experimentelle Schritt, hier soll die RNA abgebaut werden, und dadurch sollen sich die RNA-dependent Komplexe auflösen, die Protein subunits aus ihren Verbänden herausgelöst werden, und – anstatt als Komplex in einer Schicht zu landen – in unterschiedliche Schichten des Sucrose-Gradienten wandern -> Dies sind erzeugte Shifts, die dann erkannt werden können

6. Ultrazentrifugation, die einzelnen Komponenten der Probe werden in unterschiedliche Fraktionen des Gradienten wandern, abhängig von Größe, Form, und Dichte des Komplexes

7. Fraktionen 1-25 entnehmen, jeweils als Triplicates

8. Single -Protein-Analyse: Western Blot, klassisches Verfahren, man schaut sich an, 1: In welcher Fraktion ist das Protein gelandet, 2: Wo ist das Signal am häufigsten, 3: Hat es einen Signifikanten Shift gegeben zwischen Control und RNase?

9. Whole-proteome-Analyse: MS  
Zuerst TCA-Präzipitation als Reinigung und Vorbereitung auf die MS, TCA fällt die Proteine, löst sie vom störenden Rest und von der Sucrose-Lösung, und denaturiert sie auch schonmal,

Trypsin-Verdau erzeugt kürzere Fragmente, TM6-Label markiert diese Fragmente, Tandem-Mass-Tag ist individuell für die 6 Proben einer Fraktion

Alles zusammenkippen, und Tandem-Massenspektrometrie durchführen, MS1 und MS2 Spektrum geben Aufschluss über Sequenz und Vorkommen in den ursprünglichen 6 Gefäßen, die individuellen TMT6-Reporterionen zeigen an, ob das Peptid und Control, RNase, oder in beidem enthalten war.

Die TMT-Reporter-Ionen-INTENSITÄT gibt auch Info über die Quantität der Peptide in den einzelnen Proben.

Es handelt sich hier um eine relative Quantifizierung mit direktem Vergleich der Proben 1-6 untereinander, keine absolute Bestimmung

**Was genau sind diese Shifts, und wie sind sie zu interpretieren?**

Apparent molecular weight hat sich geändert.

Left shift (Shift in niedrigere Sucrose-Dichte-Fraktionen) deutet auf einen Verlust an Bindungspartner hin

Right shift (Shift in höhere Sucrose-Dichte-Fraktionen) deutet auf einen Zugewinn an Bindungspartnern hin

Right Shifts sind deutlich seltener, dies ist auch zu erwarten, weil bei RNA-Abbau eher Interaktionen verloren gehen sollte, als neue geknüpft.

Es gibt außerdem noch die Präzipitations-Fraktion, hier landen Komplexe und Proteine, die sich nicht voneinander lösen und bis in die letzten Fraktionen des Sucrose Gradienten vordringen

Ein Protein kann auch in mehreren Fraktionen sein, mit mehreren Peaks, diese Proteine sind Teil vieler Komplexe. Anhand der Stärke des Shifts lässt sich auch die Größe des Komplexes schätzen.

Für den Vortrag Kriterien entwickeln, wie ein solcher Shift als signifikant oder auch nicht signifikant zu bewerten wäre, vergleiche hierzu die 5 Kriterien von Herger Seite 4 der ersten Quelle

Shifts sind absolut nicht eindimensional, es gibt viele Proteine, die nur partially RNA-dependent sind, und daher zum großen Teil nicht shiften, während ein kleiner Peak doch shiftet, die Quelle bringt hier das Konzept von Shifting Koeffizienten rein, sieht interessant aus, kann man sich mal angucken

Shifting Proteine haben allgemein einen höheren isoelektrischen Punkt hindeuten auf besondere Art der Wechselwirkungen

**Was genau zeigt unser Datensatz jetzt?**

Quantitative Ergebnisse pro Protein pro Fraktion

Wie genau werden wir unseren Datensatz jetzt weiter bearbeiten, welche Zwischenziele wollen wir erreichen, woran werden wir uns orientieren

* Die 7-Schritte Pipeline aus dem Review sieht hierzu sehr gut aus, es gibt auch eine Maske, die man herunterladen kann zur Orientierung
* Als Milestones bieten sich diese Zwischenziele am Ende der einzelnen Schritte an

**Was sind die antizipierten Ergebnisse?**

**Reproducibility?**

Gut, man hat auch schon Ergebnisse für andere Zelllinien liefern können

Man muss aber bedenken, dass Control und RNase in getrennten Tubes laufen, das ist immer kritisch für die direkte Vergleichbarkeit

Man bekommt zwischen den Replicates häufig hohe Pearson-Korrelationswerte in Bezug auf die Peak-Positionen, spricht für gute Wiederholbarkeit der Experimente.

Nur auf Nachfrage:

**Beispiel für eine ehemalige Methode zur Identifizierung von RBPs durch UV-Crosslinking und Oligo(dT)-Affinity Purification:**

Zur Identifikation des mRNA-gebundenen Proteoms werden Zellen zunächst mit den photoreaktiven Nukleotiden (z.B. 4SU und 6SG) metabolisch markiert und anschließend mit UV-Licht (365 nm) bestrahlt, um kovalente Bindungen zwischen mRNA und assoziierten Proteinen herzustellen. Danach werden mRNA-Protein-Komplexe mithilfe von Oligo(dT)-beads isoliert, die gezielt an die Poly(A)-Schwänze der mRNA binden. Die RNA wird durch RNase abgebaut, sodass nur die direkt gebundenen Proteine übrig bleiben. Schließlich erfolgt die Identifikation dieser Proteine durch (SILAC-gestützte) Massenspektrometrie (im Vergleich zu nicht-crosslinkten Kontrollen).