

Summary

Whereas the pluripotency state of embryonic stem cells has been thoroughly investigated, we lack understanding about the exit from this state. Thus, this thesis aims to provide better insight into the molecular processes required for leaving the pluripotency state.

The group of Martin Leeb has performed high-throughput transposon mutagenesis screens in haploid embryonic stem cells of rodents. It was essayed, which embryonic stem cells present a differentiation delay phenotype. Cells unable to fully differentiate within 24 hours were subject to sequencing for identification of transposon integration positions.

Since no standard method exists for the analysis of such data, we developed a novel and rigorous statistical framework, which application revealed 230 significant genes. Among the strongest confidence genes we found *Tcf711*, *Trp53*, *Jarid2* and *Fbxw7*, which are all well described for impairing embryonic stem cell differentiation. To elucidate the involvement in signaling pathways and protein complexes, a novel enrichment R-package was created. Interestingly, identified signaling pathways belonged preeminently to the pathway domains of signal transduction and gene expression. Primary targets were related to RAF/MAP signalling e.g. FGFR-signaling, cell-cycle e.g. Notch1-signaling and transcriptional regulation by Tp53 or epigenetic modifications. We obtained in total 88 GO terms for biological processes, 49 signaling pathways and 56 protein complexes.

To ascertain our discoveries, genes were selected by their significance and relation to signaling pathways, gene ontology terms and protein complexes for validation. In order to improve the generation of null-alleles, a novel clustered regularly interspaced short palindromic repeats (CRISPR) workflow was designed and implemented. For a 130 validation genes individual gene knock-outs were performed and for 70 candidates, presenting the strongest phenotype, RNA-seq data were generated.

Analysis of the RNA-seq data allowed to identify global patterns of molecular mechanisms of differentiation delay. The data reveal, that more than a single process is capable of affecting differentiation speed. Functional enrichment of the strongest varying genes showed a variety of terms from developmental pathways, signaling pathways, Tp53-dependent cell cycle arrest, cellular adhesion or energy metabolism to be involved in the decision or facilitation of early embryonic stem cell differentiation.

The nonsense mediated decay (NMD) pathway was one of our strongest enriched pathways and although it has been described to be involved in differentiation arrest, the underlying mechanisms are scarcely understood. Knock-outs of *Smg6* and *Smg7* at two points in time were performed. We in turn identified downstream targets of NMD by extensive computational investigation of upregulated transcripts harboring premature termination codons (PTC). These should be prone to NMD of RNAs in wild-type. Putative targets identified belong to biological functions involved in negative regulation of *Tp53*, cell identity, interaction with *Nanog*, cell growth and adipocyte differentiation.

Kurzzusammenfassung

Die Beschreibung des Pluripotenzstadiums in embryonalen Stammzellen war das Ziel zahlreicher Forschungsvorhaben. Dennoch ist darüber, wie dieses Stadium im Verlauf der Zelldifferenzierung verlassen wird wenig bekannt. Daher ist es das Ziel dieser Studie, zugrunde liegende molekulare Mechanismen aufzuklären und so ein höheres Gesamtverständnis des Vorgangs zu beschreiben.

Die Gruppe von Martin Leeb hat mehrere hoch-durchsatz Transposonmutagenese Experimente in haploiden embryonalen Mausstammzellen durchgeführt.

Es wurde untersucht, welche dieser Stammzellen in ihrer Differenzierungsgeschwindigkeit beeinträchtigt waren. In Zellen, die innerhalb von 24 Stunden nicht einen entsprechenden Differenzierungsphänotypen gezeigt haben, wurden die Transposonintegrationspositionen durch Sequenzierung der DNA ermittelt.

Da es zum aktuellen Zeitpunkt kein Standardverfahren zur statistischen Analyse solcher Daten gibt, haben wir ein entsprechendes Verfahren entwickelt und mit dessen Hilfe 230 signifikante Gene identifiziert. Unter den stärksten Kandidatengenomen waren *Tcf7l1*, *Trp53*, *Jarid2* und *Fbxw7*, welche bereits zur Beeinflussung der Differenzierungsgeschwindigkeit beschrieben worden sind. Zur Aufklärung beteiligter Gen Ontologie Terme, Signaltransduktionswege und Proteinkomplexe wurde ein neues Tool zur Anreicherungsanalyse in R erstellt und angewandt. Dort wurden im Besonderen Funktionen in den Bereichen RAF/MAP signaling z.B. FGFR-signaling, Zell Zyklus z.B. Notch1-signaling and transkriptionelle Regulation durch Tp53 oder epigenetische Modifikationen aufgefunden.

Zur Bestätigung der aufgefundenen Zielgene, wurden diese mit Hilfe von CRISPR ausgeschaltet. Zu diesem Zweck wurde ein optimierter Workflow entwickelt, welcher nach Möglichkeit die Generierung von Null-Allelen beim Knock-out verbessern soll. 130 Gene wurden auf diese Weise inaktiviert und von den stärksten 70 Phänotypen wurden RNA-seq Daten erzeugt.

Unter Zuhilfenahme der RNA-seq Daten konnten globale Muster identifiziert werden. Die Daten veranschaulichten, dass sich verschiedene molekulare Prozesse ursächlich für Differenzierungsverlangsamungen zeigten.

Anreicherungsanalysen zeigten starke Auswirkungen in den Bereichen Entwicklungssignalwege, Signaltransduktionswege, Tp53-abhängiger Zell Zyklus Arrest, Zelladhäsion oder Energiemetabolismus.

Der Signalweg zum Nonsense vermittelten RNA-Abbau wurde insbesondere angereichert gefunden. Obwohl dieser zum Teil bereits beschriebene Auswirkungen auf Stammzelldifferenzierung gezeigt hat, sind die zugrundeliegenden Prozesse bisher kaum verstanden. Zur weiteren Aufklärung wurden von Martin Leeb die Gene *Smg6* und *Smg7* ausgeschaltet und zum Zeitpunkt 0 und 24 Stunden wurden RNA-Sequenzierungen durchgeführt. Aufwändige Analysen der Transcriptionsänderungen lieferten mutmaßliche Ziele in den Bereichen negative Regulation von *Tp53*, Zellidentität, Interaktionen mit *Nanog*, Zellwachstum and Differenzierung von Adipozyten.