Identification of long non-coding RNA-binding proteins in unrestricted somatic stem cells from human cord blood and human induced pluripotent stem cells

Dr. rer. nat. Jessica Schira, Molecular Proteomics Laboratory, (MPL, Prof. Dr. Kai Stühler), Institute for Molecular Medicine, Medical Faculty, Heinrich Heine University Düsseldorf

1. State of the art

Long non-coding RNAs (lncRNAs) lacking protein-coding potential are increasingly recognized as gene expression regulators acting in *cis* or *trans*. RNA profiling revealed that lncRNAs are expressed in a tissue specific manner. However, the biological role of lncRNAs and their mode of action have just begun to be understood and need further elucidation (1). Most lncRNAs seem to have epigenetic or posttranscriptional function, e.g., by interacting with transcription factors (TFs), chromatin or chromatin-modifying complexes (2) as shown for a lncRNA termed *HOTAIR*. *HOTAIR* is encoded within the HOXC cluster and binds to the polycomb repressive complex 2 (PRC2) which is required for transcriptional silencing of the HOXD locus (3). It is further suggested that lncRNAs are scaffolds to assemble selected proteins in cell type specific RNA-protein-complexes (4) to collectively coordinate gene expression programs (5). Little information is available regarding the functionality of lncRNAs in stem cell maintenance as well as during neural lineage differentiation. A shRNA mediated systematic loss of lncRNAs function in mouse embryonic stem cells (ESC) revealed that lncRNAs exert a strong impact on pluripotency maintenance by regulating TFs such as Nanog and Oct-4 (5). The lncRNA *TUNA* was shown to affect global gene expression of mouse and human ESC with impact on self-renewal and neural differentiation (6). Pull-down assays of *TUNA* revealed that it binds the proteins PTBP1, hnRNP-K and nucleolin, which are all implicated in the maintenance of pluripotency. The *TUNA*-protein complex further associates with promoters of the TFs Nanog and Sox2, thereby likely participate in regulation of pluripotency and differentiation (6). Moreover, lncRNAs act as “sponges” by binding and repressing microRNAs (miRNAs) as described for *linc-RoR* (*r*egulator *o*f *r*eprogramming) (7). The miRNA hsa-miR-145 is a key translational repressor of the core TFs Sox-2, Oct-4 and Nanog and binding of *linc-RoR* to miR-145 blocks its function and thus maintains pluripotency. The core TFs were further described to bind to the promoter of *Haunt*, a lncRNA that regulates HOXA gene expression in precise orchestration with the *Haunt* DNA locus both with opposing roles (1). *Haunt* lncRNA prevents aberrant expression of HoxA by attenuating long-range chromatin interactions between the *Haunt* enhancer DNA and the HOXA region likely through differentiation suppression by direct chromatin association. The *Haunt* locus activates HoxA transcription by potential HoxA enhancers. In addition, *HOTAIR*, *TUNA* and *linc-RoR* maintain self-renewal in cancer stem cells indicating that these lncRNA play important roles in various stem cell populations (8-10). Taken together, *TUNA*, *linc-RoR* and *Haunt* orchestrate stem cell maintenance and differentiation at various regulatory levels either by directly (*TUNA*) or indirectly (*linc-RoR*) regulating core TFs expression and translation. In addition, pluripotency and differentiation is controlled by core TFs which regulate other lncRNAs (i.e. *Haunt*) as a further regulatory level. Since interaction with proteins is essential for regulatory function of lnc-RNAs, determination of the functional contribution of each lncRNA-protein-interaction during developmental processes will lead to a more detailed understanding how gene expression patterns are regulated (5). Expression of lncRNAs in stem cells derived from human umbilical cord blood (hUCB) is rarely investigated. In addition to mesenchymal stem cells (MSC), hUCB contains unrestricted somatic stem cells (USSC), a

CD45-negative population of multipotent stem cells with a broader differentiation potential compared to adult MSC (11). USSC and MSC can be distinguished by the expression of delta-like 1/preadipocyte factor 1 (DLK-1/PREF-1) (12) and a specific Hox gene expression pattern. Herein, USSC reflect the HOX expression pattern of human H9 ESC and do not express HOXA and HOXD (13). Furthermore, USSC differentiate into cells of neuronal lineage *in vitro* (14).

2. Preliminary work

Our previous work focused on the regeneration promoting potential of USSC and the miRNA impact on their neuronal lineage differentiation. We demonstrated that USSC promote functional locomotor recovery and axonal regeneration after transplantation into a rat spinal cord injury model (15). Furthermore, proteome approaches revealed that USSC secrete several trophic factors which potentially promote axonal regeneration and regeneration-associated processes (16). Our evaluation of the impact of miRNAs during neuronal lineage differentiation of USSC (14) revealed a downregulation of a set of 18 miRNAs, mostly from the miR-17-92 family. Bioinformatic target predictions followed by experimental validations unfolded a network of miRNAs affecting neuronal differentiation-related pathways such as axon guidance and WNT signaling (17). Indeed 35 % of experimentally tested predictions were confirmed in Firefly luciferase reporter assays including neurobeachin, neurogenic differentiation 1 and others (17). In addition, the cell cycle arrest of USSC during neuronal lineage differentiation is mediated by downregulation of miR-17, -20a, and -106b, all targeting a set of pro- and anti-proliferative target genes. This results in an increase of the intracellular activity of E2F transcription factors to govern G(1)/S transition (18). We further demonstrated that hsa-miRNA-145 is highly expressed in naïve USSC and during their neural lineage differentiation (17) suggesting that *linc-RoR* which targets hsa-miRNA-145 might not be expressed in USSC. To further define miRNA networks involved in neural lineage differentiation our ongoing project in the framework of the SPP1738 centers on comparison of differentiations of USSC and human induced pluripotent stem cells (iPSC) to define important regulatory checkpoints in both differentiation paradigms. We recently already analyzed naïve USSC and iPSC in a shotgun proteomic approach revealing that, beside other predicted RNA-binding proteins, *TUNA*-interacting factors PTBP1, hnRNP-K and nucleolin are present in both cell populations.

3. Objectives

As almost every functional RNA molecule recruits proteins, the identification of lncRNA-protein interactions is an essential step to understand lncRNA functions (19). Although individual proteins associating with lncRNAs have been described, detailed characterization of lncRNA-protein complexes and their function in human pluripotent (ESC, iPSC) as well as in somatic stem cells is still lacking. In particular, no data currently exist regarding expression of lncRNAs in USSC. Thus the experimental goal of this application is to

1) evaluate and compare the expression of lncRNA in naïve USSC and iPSC,

2) investigate protein interactions with selected candidate lncRNAs in USSC and iPSC cells,

3) validate single lncRNA binding proteins by Western Blot analysis and immunocytochemical stainings.

In this approach, we will identify, characterize and quantify protein-lncRNA complexes in USSC and human iPSC with an accurate and sensitive mass spectrometry method (RAP-MS) after affinity based enrichment of the candidate lncRNAs. Herein we will choose lncRNAs which are exclusively expressed in only USSC or human iPSC as well as lncRNA being expressed in both populations. The basic comparative characterization of lncRNA expression in USSC and human iPSC together with the

identification of proteins bound to these lncRNAs will provide the groundwork for the future analysis of lncRNA expression and function during neuronal differentiations of both cell types. Furthermore, RAP-MS will deliver the proof-of-principle for identification of lncRNA-protein interactions, which will be a highly valuable tool for the members of the SPP1738.

4. Work program

4.1 Expression patterns of lncRNA in naïve USSC and human iPSC

To determine lncRNA expression patterns in both stem cell populations we will perform an unbiased approach using lncRNA array (RT^2 lncRNA PCR Array Human IncFinder, Qiagen) covering 84 lncRNAs including *HOTAIR*, *TUNA* and *linc-RoR* which are expected to be highly expressed in iPSC. Since USSC express low amounts of Hox genes, we hypothesize *HOTAIR* likewise expressed in USSC. In contrast, since hsa-miR-145 is highly abundant in USSC, we assume that linc-RoR might be weakly expressed in USSC. Total RNA will be isolated from both cell populations by mirVana™ Total RNA Isolation Kit (Thermo Fisher Scientific), reversely transcribed and analyzed for lncRNA expression using the RT^2 lncRNA PCR Array Human IncFinder on an ABI7900HT Fast Real Time PCR system (Applied Biosystems). At least three different USSC lines and two different human iPSC lines will be used for all experiments. From these expression data we will select lncRNAs exclusively expressed in either USSC or human iPSC and lncRNAs expressed in both cell types as candidates for subsequent evaluation of protein interactors.

4.2 RAP-MS of protein-lncRNA complexes

To characterize protein-lncRNA complexes we will perform RNA antisense purification (RAP) and subsequent quantitative mass spectrometry (RAP-MS) (20) which allows to define intracellular protein-lncRNA complexes without genetic tagging. The method will be modified regarding protein quantification strategy. Instead of stable isotope labeling of intracellular proteins (SILAC) for relative protein abundance quantification we will perform label-free quantification which is less time consuming and cost-intensive. We thereby exclude variances in labeling efficiency and the method is routinely performed at the MPL. Protein-lncRNA complexes will be UV-crosslinked *in vivo* and purified by antisense biotinylated 90 nt DNA oligonucleotide probes (at least 2 independent probes for each lncRNA). Coupling of the probe to streptavidin-coated magnetic beads (Thermo Scientific) allows pull-down of protein-lncRNA complexes which subsequently will be analyzed by nano liquid chromatography-tandem mass spectrometry (LC-MS/MS) (see 4.3). Since the probes can potentially pull-down DNA fragments, appropriate controls will be used as follows. Capture of protein-lncRNA complexes from non-crosslinked cells will be used to exclude non-specific binding of proteins during experimental procedure. In addition, scrambled versions of probes and RNase treated samples before RAP will be used to approximate RNA-independent noise.

4.3 Identification of proteins by LC-MS/MS

For unbiased identification and quantification of protein binding to candidate lncRNAs in USSC and iPSC, LC-MS/MS analysis will be performed. Samples will be processed according to standard workflow (16) which includes in-gel digestion of proteins with trypsin and separation of peptides using the nano LC system. Peptides will be transferred into the mass spectrometer (LTQ Orbitrap Elite or Q Exactive™ Hybrid Quadrupole-Orbitrap, Thermo Fisher Scientific) via electrospray. Identification of proteins will be performed using Proteome Discoverer (Thermo Fisher Scientific) including Mascot search engine. The false discovery rate will be set to 1% ($p \leq 0.01$).

4.4 Quantitative analysis of lncRNA binding proteins

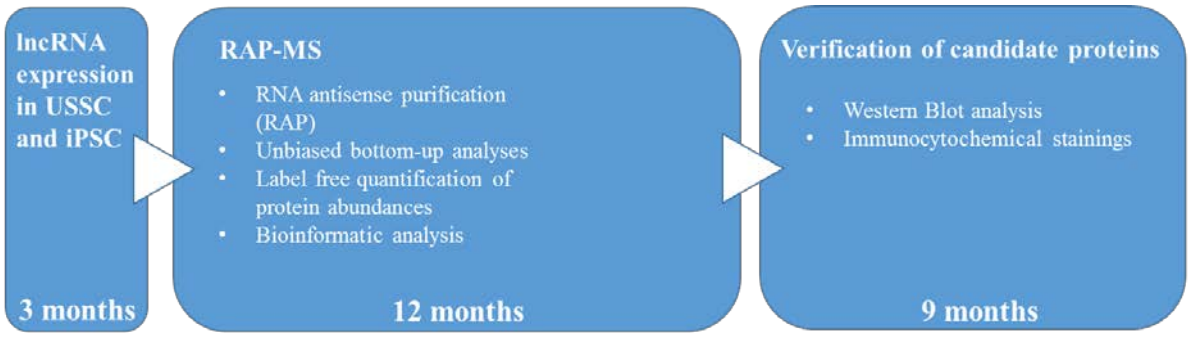
To identify proteins of interest interacting with candidate lncRNAs, label-free quantification will be performed using Progenesis QI for proteomics 2.0 (Nonlinear Dynamics) determining MS peak

intensities which enables quantification even in presence of background proteins. A normalization factor will be calculated for comparison of individual runs. Proteins of interest have to be identified with at least 2 unique peptides in the mass spectrum. Statistical analysis will be performed using ANOVA with a significance threshold of 5 % ($p \leq 0.05$). Comparison of protein abundances with controls will proof candidate proteins in each differentiation stage. Relative protein abundance of candidate proteins will then be compared across the differentiation stages to calculate stage specific protein abundances. Furthermore, molecular functional networks will be determined using Ingenuity Pathway Analysis and / or Reactome.

4.5 Verification and Localization of lncRNA binding proteins

To further confirm candidate proteins, Western Blot analysis of single candidate proteins will be performed. To investigate if the candidate proteins form multiprotein complexes or exist as single lncRNA-bound protein, candidates will be immunoprecipitated by specific antibodies, treated with RNase and analyzed by Western Blot. To determine intracellular localization of candidate proteins, immunocytochemical stainings will be performed.

4.6 Workflow and expected time frame



5. Requested funding

Reagents	costs (€)
Medium for cultivation of iPSC and USSC	1500
Reagents for buffers (LC-grade)	1500
Dynabeads® MyOne™ Streptavidin C1 (Thermo Scientific)	5500
biotinylated DNA nucleotide probes (Eurofins)	2500
UV crosslinker (Bio-link BLX 254, PeqLab)	2200
mirVana™ Total RNA Isolation Kit	480
RT² First Strand Kit (Qiagen)	250
RT² SYBR Green ROX qPCR Mastermix (Qiagen)	260
RT2 lncRNA PCR Array Human IncFinder (Qiagen)	1100
DNase / RNase	300
MS-compatible trypsin (Promega)	500
Antibodies & IgG control antibodies for Western Blot / stainings	2000
Proteinase / RNase inhibitors (Roche)	800
Acclaim PepMap RSLC, 75µm x 25 µm columns (Thermo Scientific)	2600
Acclaim PepMap 100, 75µm x 2cm columns (Thermo Scientific)	2100
Pico Tip Emitter (MS Will GmbH)	1100
total (24 months)	**24690**

6. Significance statement and future perspective

The intended approach will provide proof of principle to study protein-lncRNA interactions in human induced pluripotent stem cells and USSC by quantitative mass spectrometry. The pilot project applied for here will deliver valuable results, indispensable for deeper understanding of protein-lncRNA networks. The workflow which will be developed is likely to be transferable to other cell model systems and thus the established state of the art technique will be provided to other members of the SPP1738. In addition, the approach described here will provide a basic dataset regarding lncRNA function in our cell system and therefore constitutes the prerequisite for future its extension into the field of neuronal differentiations.

Data obtained by this approach will be published by the applicant as a senior author. This will further strengthen the applicant's publication track record in the highly specialized field of proteomics in neurobiology. This is most valuable as she plans to apply as a principal investigator for a future DFG proposal in the framework of the SPP1738's second funding period alongside the current PIs of the project. Since parts of this future proposal will aim at functional validation of lncRNAs and RNA binding proteins during neural lineage differentiation of both cell populations, the study applied for here will not only generate essential preliminary data needed for this upcoming application but will also extend the group's scientific and technical expertise in the field of noncoding RNAs in neuronal differentiation and development.

7. Scientific equipment

The MPL (Kai Stühler) is fully equipped and highly experienced in analysis of MS-based protein-protein interactions (21) which enables the identification and label-free quantification (22) of lncRNA-protein interactions. The department of Neurology (Hans Werner Müller) is equipped with a confocal laser scanning microscope (LSM 510, Zeiss) for immunocytochemical stainings and an ABI7900HT Fast Real Time PCR system (Applied Biosystems) for quantitative RT-PCR analyses. The Institute for Transplantation Diagnostics and Cell Therapeutics (Hans-Ingo Trompeter) will provide human iPSC cultures which are routinely cultured for the SPP1738 project of Hans-Ingo Trompeter, Hans Werner Müller and Kai Stühler.

8. References

1. Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., Orkin, S. H., and Shen, X. (2015) Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. *Cell stem cell* 16, 504-516
2. Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., and Rinn, J. L. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 106, 11667-11672
3. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323
4. Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689-693
5. Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., Young, G., Lucas, A. B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J. L., Root, D. E., and Lander, E. S. (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295-300

6. Lin, N., Chang, K. Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., Zhang, D., Han, T., Yang, C. S., Cunningham, T. J., Head, S. R., Duester, G., Dong, P. D., and Rana, T. M. (2014) An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Molecular cell* 53, 1005-1019
7. Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013) Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Developmental cell* 25, 69-80
8. Hou, P., Zhao, Y., Li, Z., Yao, R., Ma, M., Gao, Y., Zhao, L., Zhang, Y., Huang, B., and Lu, J. (2014) LincRNA-ROR induces epithelial-to-mesenchymal transition and contributes to breast cancer tumorigenesis and metastasis. *Cell death & disease* 5, e1287
9. Li, H., Zhu, L., Xu, L., Qin, K., Liu, C., Yu, Y., Su, D., Wu, K., and Sheng, Y. (2015) Long noncoding RNA linc00617 exhibits oncogenic activity in breast cancer. *Molecular carcinogenesis*
10. Padua Alves, C., Fonseca, A. S., Muys, B. R., de Barros, E. L. B. R., Burger, M. C., de Souza, J. E., Valente, V., Zago, M. A., and Silva, W. A., Jr. (2013) Brief report: The lincRNA Hotair is required for epithelial-to-mesenchymal transition and stemness maintenance of cancer cell lines. *Stem cells* 31, 2827-2832
11. Kogler, G., Sensken, S., Airey, J. A., Trapp, T., Muschen, M., Feldhahn, N., Liedtke, S., Sorg, R. V., Fischer, J., Rosenbaum, C., Greschat, S., Knipper, A., Bender, J., Degistirici, O., Gao, J., Caplan, A. I., Colletti, E. J., Almeida-Porada, G., Muller, H. W., Zanjani, E., and Wernet, P. (2004) A new human somatic stem cell from placental cord blood with intrinsic pluripotent differentiation potential. *The Journal of experimental medicine* 200, 123-135
12. Kluth, S. M., Buchheiser, A., Houben, A. P., Geyh, S., Krenz, T., Radke, T. F., Wiek, C., Hanenberg, H., Reinecke, P., Wernet, P., and Kogler, G. (2010) DLK-1 as a marker to distinguish unrestricted somatic stem cells and mesenchymal stromal cells in cord blood. *Stem cells and development* 19, 1471-1483
13. Liedtke, S., Buchheiser, A., Bosch, J., Bosse, F., Kruse, F., Zhao, X., Santourlidis, S., and Kogler, G. (2010) The HOX Code as a "biological fingerprint" to distinguish functionally distinct stem cell populations derived from cord blood. *Stem cell research* 5, 40-50
14. Greschat, S., Schira, J., Kury, P., Rosenbaum, C., de Souza Silva, M. A., Kogler, G., Wernet, P., and Muller, H. W. (2008) Unrestricted somatic stem cells from human umbilical cord blood can be differentiated into neurons with a dopaminergic phenotype. *Stem cells and development* 17, 221-232
15. Schira, J., Gasis, M., Estrada, V., Hendricks, M., Schmitz, C., Trapp, T., Kruse, F., Kogler, G., Wernet, P., Hartung, H. P., and Muller, H. W. (2012) Significant clinical, neuropathological and behavioural recovery from acute spinal cord trauma by transplantation of a well-defined somatic stem cell from human umbilical cord blood. *Brain : a journal of neurology* 135, 431-446
16. Schira, J., Falkenberg, H., Hendricks, M., Waldera-Lupa, D. M., Kogler, G., Meyer, H. E., Muller, H. W., and Stuhler, K. (2015) Characterization of regenerative phenotype of unrestricted somatic stem cells from human umbilical cord blood by functional secretome analysis. *Molecular & cellular proteomics : MCP*
17. Iwaniuk, K. M., Schira, J., Weinhold, S., Jung, M., Adjaye, J., Muller, H. W., Wernet, P., and Trompeter, H. I. (2011) Network-like impact of MicroRNAs on neuronal lineage differentiation of unrestricted somatic stem cells from human cord blood. *Stem cells and development* 20, 1383-1394
18. Trompeter, H. I., Abbad, H., Iwaniuk, K. M., Hafner, M., Renwick, N., Tuschl, T., Schira, J., Muller, H. W., and Wernet, P. (2011) MicroRNAs MiR-17, MiR-20a, and MiR-106b act in concert to modulate E2F activity on cell cycle arrest during neuronal lineage differentiation of USSC. *PloS one* 6, e16138
19. Chu, C., Spitale, R. C., and Chang, H. Y. (2015) Technologies to probe functions and mechanisms of long noncoding RNAs. *Nature structural & molecular biology* 22, 29-35
20. McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E.

S., Hess, S., Plath, K., and Guttman, M. (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232-236

21. Erkelenz, S., Poschmann, G., Theiss, S., Stefanski, A., Hillebrand, F., Otte, M., Stuhler, K., and Schaal, H. (2013) Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of vpr mRNA. *Journal of virology* 87, 2721-2734
22. Malzkorn, B., Wolter, M., Liesenberg, F., Grzendowski, M., Stuhler, K., Meyer, H. E., and Reifenberger, G. (2010) Identification and functional characterization of microRNAs involved in the malignant progression of gliomas. *Brain pathology* 20, 539-550