

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl Entwicklungsgenetik

Identification of molecular hallmarks and regulators during neuronal maturation in adult hippocampal neurogenesis using RNA-Sequencing

Elisabeth Uttenthaler

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. E. Grill

Prüfer der Dissertation: 1. Univ.-Prof. Dr. W. Wurst

2. Univ.-Prof. Dr. D.C. Lie

Die Dissertation wurde am 14. 03. 2016 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 18. 07. 2016 angenommen.

Danksagung

Zu allererst möchte ich mich bei Prof. Dr. W. Wurst für die Möglichkeit bedanken, an seinem Institut meine Doktorarbeit anfertigen zu können, und bei Prof. Dr. E. Grill für die Bereitschaft den Vorsitz meines Rigorosums zu übernehmen.

Mein ganz besonderer Dank gilt meinem Betreuer Prof. Dr. Chichung Lie für seine hervorragende Betreuung und Unterstützung und dafür, dass er immer an mich und das Projekt geglaubt hat.

Des Weiteren möchte ich mich bei Prof. Dr. L. Bally-Cuif und Prof. Dr. H. Bading bedanken, die mir als Mitglieder meines Thesis Committees wertvollen fachlichen Input gegeben haben. Außerdem bedanke ich mich bei Prof. Dr. T. Klopstock für die Möglichkeit das LCM Mikroskop mitbenutzen zu dürfen, und bei Dr. W. Beisker für die Hilfe beim FACS. Für die hilfreiche Unterstützung bei bioinformatischen Analysen möchte ich mich bei Dr. N. Müller, Dr. T. Aneichyk (HMGU, München), Dr. N. Nagarajan, B. Chia (Genome Institute of Singapore), Prof. Dr. H. Song und Dr. Z. Zhonghua (Johns Hopkins University School of Medicine, Baltimore) bedanken.

Entscheidend mitgewirkt an diesem Projekt haben außerdem Marcela Covic und Filippo Cernilogar, die sowohl theoretische als auch praktische Hilfeleistungen gegeben haben.

Darüber hinaus möchte ich mich bei allen Leuten der AG Lie für die motivierende und freundschaftliche Arbeitsatmosphäre bedanken, insbesondere bei Birgi, Tobi, Kathrin K. und Sarah, die über das Labor hinaus, sehr gute Freunde fürs Leben geworden sind.

Mein größter Dank gebührt meinem geliebten Mann Andi, meinen Kindern Hannah und Sarah, und unseren Familien für ihr Verständnis und ihre Unterstützung und dafür, dass sie mir stets den Freiraum gegeben haben, um diese für mich so überaus wichtige Arbeit abzuschließen. Danke für eure unendliche Liebe und Unterstützung! Ihr seid das wichtigste in meinem Leben!

Table of Contents

1	ZUSAMMENFASSUNG	6
2	ABSTRACT	8
3	INTRODUCTION	9
	3.1 Adult neurogenesis.....	9
	3.1.1 Neurogenic niches.....	10
	3.1.2 Development of newborn neurons in the adult hippocampus	13
	3.1.3 Functional significance of adult hippocampal neurogenesis.....	16
	3.1.4 Transcriptome analysis of newborn neurons within the two neurogenic niches.....	17
	3.2 The role of long non-coding RNAs	20
	3.2.1 Putative functions of lncRNAs	23
	3.2.2 Antisense transcription	28
	3.2.3 The role of lncRNAs in adult neurogenesis.....	29
	3.3 RNA-Seq.....	30
	3.3.1 Directional RNA-Seq.....	32
	3.3.2 Transcriptome analysis of single cells or a limited low number of cells	33
	3.3.3 Techniques to isolate cells from complex heterogeneous tissues.....	34
	3.3.4 Different amplification technologies for a limited number of cells.....	38
	3.3.5 Comparison of different amplification methods	46
	3.4 Objectives of this study	49
4	RESULTS.....	50
	4.1 Labelling and isolation of newborn neurons	50
	4.2 Directional amplification protocol	54
	4.3 Quality control and technical performance	58
	4.3.1 Quality of raw reads.....	58
	4.3.2 Mappability of reads	59
	4.3.3 Comparison of two biologically identical samples prepared with different library preparation kits and sequenced at different sequencing depths.....	60
	4.3.4 Mapping to genomic regions, directionality, reads distribution and expression level	62
	4.3.5 Reproducibility and sensitivity.....	65
	4.3.6 Validation of specific maturation stages	67
	4.4 Differentially expressed genes	68
	4.4.1 Genes and cellular processes enriched at 6 dpi.....	70
	4.4.2 Genes and cellular processes enriched at 12 dpi.....	72

4.4.3	Genes and cellular processes enriched at 28 dpi.....	76
4.5	Dynamic expression of RefSeq annotated long non-coding RNAs.....	80
4.6	Antisense expression	84
4.7	Novel lncRNA transcripts	85
5	DISCUSSION	90
5.1	Cellular processes involved in the maturation of adult-born neurons of the adult hippocampus	90
5.2	Long ncRNAs involved in the maturation of adult-born neurons of the adult hippocampus	94
5.2.1	Which novel lncRNAs are functional?	94
5.2.2	Linking lncRNA to function	96
5.3	Strengths, limitations and perspectives	97
5.3.1	Sensitivity, expression levels and sequencing depth	99
5.3.2	Why are more genes detected at 28 dpi compared to 6 and 12 dpi samples?	101
5.3.3	Reasons for sample heterogeneity	101
5.3.4	Limitations of the protocol	105
5.3.5	Future technological advancements	108
5.3.6	Outlook and significance	110
6	MATERIALS and METHODS	111
6.1	Materials.....	111
6.1.1	Reagents for RNA amplification.....	112
6.1.2	Equipment and material.....	113
6.1.3	Software	113
6.1.4	Primer.....	114
6.1.5	Buffers, solutions and media	116
6.2	Methods	117
6.2.1	Animals.....	117
6.2.2	Plasmid production	118
6.2.3	Retrovirus preparation	120
6.2.4	Stereotactic injection	121
6.2.5	Fluorescence-activated cell sorting (FACS).....	122
6.2.6	Croysectioning and LCM of retrovirally labeled cells	123
6.2.7	RNA amplification and library preparation.....	124
6.2.8	qRT-PCR	130
6.2.9	Bioinformatic analysis	131

7	ABBREVIATIONS.....	134
8	REFERENCES.....	139
9	APPENDIX	159
	Eidesstaatliche Erklärung	159
	CURRICULUM VITAE.....	160

1 ZUSAMMENFASSUNG

Neuronale Stammzellen des hippocampalen Gyrus dentatus bilden ein ganzes Leben lang neue Nervenzellen. Es ist bekannt, dass die genaue Regulation der Reifung und funktionellen Integration von neugeborenen Nervenzellen für Neurogenese-abhängige Gedächtnisprozesse besonders entscheidend ist. Die zugrundeliegende molekulare Regulation der adulten Neurogenese ist jedoch noch weitestgehend unbekannt.

Das Ziel der vorliegenden Dissertation war die Charakterisierung des Transkriptom von spezifischen Reifungsstadien während der *in vivo* adulten hippocampalen Neurogenese, mit dem Ziel der Aufklärung der Mechanismen, welche die Reifung von unreifen Nervenzellen kontrollieren. Hierzu wurde i) eine Methodik entwickelt, die es erlaubt, neugeborene Nervenzellen aus dem adulten Hippokampus zu bestimmten Entwicklungszeitpunkten *in vivo* zu markieren und zu isolieren; und ii) eine RNA-Seq Methode entwickelt, die es erlaubt, eine geringe RNA Menge (< 100 Zellen) direktional zu sequenzieren.

Diese Untersuchung enthüllte Gene und zelluläre Prozesse, welche maßgeblich an der Reifung, Differenzierung und Integration von neugeborenen Nervenzellen im adulten Hippokampus beteiligt sind. Eine entscheidende Rolle spielen dabei unter anderem Transkriptionsfaktoren, Translationsfaktoren, Metabolismus, Zytoskelett und RNA Transport.

Darüber hinaus erlaubte die direktionale Amplifikation in Verbindung mit hoher Sequenziertiefe die Charakterisierung von annotierten und nicht-annotierten langen nicht-kodierenden RNAs (lncRNAs), die vor allem bei der Regulation von transkriptionellen und epigenetischen Prozessen eine Rolle zu spielen scheinen. Beinahe ein Drittel der exprimierten kodierenden Gene wies eine Expression auch auf dem Gegenstrang auf, i.e., antisense Transkripte. Des Weiteren wurden 21229 neue, noch nicht annotierte lncRNAs identifiziert, von denen 15% differenziell exprimiert wurden. Interessanterweise befinden sich diese differenziell exprimierten lncRNAs in der Nähe von protein-kodierenden Genen, welche bei der Neurogenese eine Rolle spielen. Da lncRNAs oftmals mit benachbarten Genen reguliert werden und funktionell miteinander verbunden sind, bekräftigt dies die Annahme, dass diese lncRNA-mRNA Paare bei der Regulation der adulten hippocampalen Neurogenese eine funktionelle Rolle spielen könnten.

Diese Arbeit liefert eine Datenbank zur Aufklärung neuer regulatorischer Netzwerke während der neuronalen Reifung *in vivo*, und trägt darüber hinaus dazu bei, Mechanismen aufzuklären, welche an der Pathogenese von neurodegenerativen und neuropsychiatrischen Erkrankungen beteiligt sind. Dieses Wissen wird dazu beitragen, Strategien zu entwickeln, welche auf die

hippokampale Neurogenese Einfluss nehmen, um dem kognitiven Abbau während des Alterns und der Neurodegeneration entgegenzuwirken. Letztlich kann die entwickelte Methode dazu beitragen, dass komplexe genetische Programme, die mit zellulärer Funktion, Entwicklung und Erkrankung in Verbindung stehen, in unterschiedlichen biologischen Systemen untersucht werden können.

2 ABSTRACT

Neural stem cells in the mammalian hippocampal dentate gyrus generate neurons throughout life. While it has been firmly established that precise regulation of maturation and functional integration of newborn neurons is central to adult hippocampal neurogenesis-dependent cognition, the molecular regulation of this key developmental step remains largely unknown.

The aim of the present study was the transcriptomic characterization of distinct maturation stages of *in vivo* adult hippocampal neurogenesis, in order to elucidate the mechanisms controlling the maturation of immature neurons to mature neurons. This was achieved by i) establishing a methodology for labelling and isolation of stem cell-derived neurons from the adult hippocampus at distinct developmental stages; and ii) establishing a method for directional RNA-Seq for a limited quantity of RNA (< 100 cells).

This study revealed important genes and cellular processes that are involved in the proper maturation, differentiation and integration of newborn neurons during adult hippocampal neurogenesis, characterized by regulatory switches in transcription factors, translation machinery, metabolism, cytoskeleton and RNA transport machinery.

Moreover, the directional amplification combined with deep sequencing allowed the characterization of already annotated and not yet annotated long non-coding RNAs (lncRNAs), which are emerging regulators of transcription and epigenetic modifications. Almost one third of expressed coding genes had expression on the antisense strand as well. Moreover, 21229 novel, not yet annotated lncRNAs could be identified of which about 15% were differentially expressed in the course of maturation. Notably, these differentially expressed lncRNAs were found in close proximity to protein-coding genes involved in neurogenesis-related functions, thereby supporting a possible functional role of these lncRNA-mRNA pairs in the regulation of adult hippocampal neurogenesis.

In sum, this database provides a roadmap for the discovery of regulatory networks for *in vivo* neuronal maturation, and will thereby support the identification of mechanisms contributing to the pathogenesis of neurodevelopmental and neuropsychiatric disorders. Such knowledge is expected to contribute to the development of strategies that target hippocampal neurogenesis to counter cognitive decline in ageing and neurodegeneration. Finally, the methodology developed will be highly valuable to decipher complex genetic programs associated with cellular function, development and disease of many biological systems.

3 INTRODUCTION

3.1 *Adult neurogenesis*

Just a few decades ago it was a common belief that the mature mammalian brain is not capable to generate new neurons. The adult central nervous system (CNS) was considered to be postmitotic, without the possibility of regeneration. “In the adult centres, the nerve paths are something fixed, ended and immutable. Everything may die, nothing may be regenerated” (Santiago Ramon y Cajal 1913) (Balu & Lucki 2009). This central dogma in neuroscience was first challenged by the discovery of mitotic cells in the dentate gyrus of the hippocampus in adult rats (Altman & Das 1965). It was shown that some granule cells of the hippocampal dentate gyrus were labeled after the injection of thymidine- H^3 , a nucleoside that is taken up by cells synthesizing DNA before cell division. However, this data was initially not accepted by the scientific community, as there was lack of evidence that these cells were neurons. At that time, neuron-specific immunohistochemical markers did not exist yet, and the cellular phenotype was evaluated purely on the basis of morphological criteria.

The interest in studying adult neurogenesis was rekindled by the discovery of the production and functional integration of new neurons in the adult central nervous system (CNS) of songbirds (Goldman & Nottebohm 1983). It was shown that canaries replace new neurons in the high vocal center seasonally when they learn new songs. The study of neurogenesis and neuronal replacement in birds paved the way for neurogenesis research in mammals, and the connection between neurogenesis and learning.

Eventually, a combination of the cell cycle marker thymidine- H^3 with neuron-specific immunohistochemical markers clearly demonstrated the existence of neurogenesis in adult rats (Cameron et al 1993). This was the first detailed characterization of newborn neurons and their location in the adult dentate gyrus of rats. In 1998, Eriksson and colleagues demonstrated that neurogenesis persists also in humans throughout life (Eriksson et al 1998). This evidence was further confirmed by an independent method that measures the incorporation of the C^{14} isotope, which allows for retrospective birthdating of cells in humans (Spalding et al 2013, Spalding et al 2005). Similar as C^{14} is used in archeology to retrospectively date biological material, the amount of C^{14} isotope can be measured to

determine the timepoint when the DNA was synthesized and cells were born in different brain regions.

Many studies followed demonstrating that life-long continuous neurogenesis exists in almost all mammals examined, including humans (Bergmann et al 2015, Bond et al 2015, Brus et al 2013). A recent comparison of adult hippocampal neurogenesis among 71 mammalian species showed that robust levels of hippocampal neurogenesis persist across all species examined (as measured by DCX-immunohistochemistry), except for two cetacean species (whales and dolphins) (Patzke et al 2013). It was further demonstrated that the immunohistochemical marker profile as well as the age-related decline in neurogenesis was similar in humans and rodents, indicating that the qualitative and quantitative course of neurogenesis is evolutionary conserved between rodents and humans (Knoth et al 2010).

3.1.1 Neurogenic niches

In the context of stem cells, a niche is “a specific location in a tissue where stem cells can reside for an indefinite period of time and produce progeny cells while self-renewing” (Ohlstein et al 2004).

The neurogenic niche is characterized by a cellular and molecular microenvironment ideally suitable for the generation and integration of new neurons. Surrounding astrocytes, ependymal cells and close proximity to vasculature are important factors that build up a gradient of nutrients and growth factors within the neurogenic niche which determine vital properties of stem cells including self-renewal, proliferation, fate determination and differentiation (Lie et al 2004, Morrison & Spradling 2008, Suh et al 2009).

In mammals at least two brain regions are capable of generating new neurons throughout life (Fig 1A):

- i) The subventricular zone (SVZ) of the lateral ventricle
- ii) The subgranular zone (SGZ) of the dentate gyrus in the hippocampus

Besides these two regions, robust levels of adult neurogenesis recently were found also in the striatum, but only in humans, as measured by the carbon-14 (C^{14}) dating approach (Ernst et al 2014). In the following I will focus on the two “canonical” regions, i.e., the SVZ of the

lateral ventricle and the SGZ of the dentate gyrus, where neurogenesis has been consistently reported in many mammalian species.

3.1.1.1 The subventricular zone (SVZ) of the lateral ventricle

In the adult subventricular zone (SVZ), activated radial glia-like stem cells give rise to transit amplifying progenitor cells, which generate neuroblasts. These neuroblasts migrate in chains through tubular structures formed by astrocytes via the rostral migratory stream (RMS) towards the olfactory bulb (OB) (Fig 1A). Neuroblasts mature into olfactory inhibitory interneurons – granule cells and periglomerular cells and make local contacts with mitral and tufted cells (Lledo et al 2006). The majority become GABAergic granule neurons, a minority (5 %) become dopaminergic periglomerular neurons, and a very small percentage develop into glutamatergic juxtglomerular neurons (Brill et al 2009).

More than 30 000 neuroblasts exit the rodent SVZ each day (Alvarez-Buylla et al 2001).

3.1.1.2 The subgranular zone (SGZ) of the dentate gyrus of the hippocampus

Radial glia-like (RGL) stem cells reside in the subgranular zone (SGZ) of the dentate gyrus in the hippocampus and have the long-term ability to self-renew and give rise to neurons and astrocytes (Bonaguidi et al 2011). The SGZ is defined as the region in the lower third of the granule cell layer (GCL), at the border between the hilus and the GCL. Upon activation, radial glia-like stem cells (type 1) give rise to intermediate progenitor cells (IPC) with first glial (type 2a) and then neuronal (type 2B) phenotype. After neuronally committed neuroblasts (type 3) have exit the cell cycle, immature granule neurons migrate towards the granule cell layer (GCL) and extend dendrites towards the molecular layer and their axon to CA3. Within 4-6 weeks they become mature granule neurons with an extensive dendritic tree, and become integrated in the hippocampal circuitry (Kempermann et al 2015) (Fig 1B).

Using the C¹⁴ dating approach it was calculated that 700 new neurons are added in each hippocampus per day in adult humans (Spalding et al 2013). In young adult rats and mice the numbers are approximately 10 times higher, estimating that approximately 9000 new granule neurons are generated per day in rats (Cameron & McKay 2001) and 2000 – 4000 in

mice (Kempermann et al 1997) as measured by the S-phase marker BrdU. The relative proportion of newborn to old neurons that is added in the dentate gyrus per day is 0.004% in humans, and 0.03 – 0.06% in 2-month-old mice (Kempermann et al 1997, Spalding et al 2013).

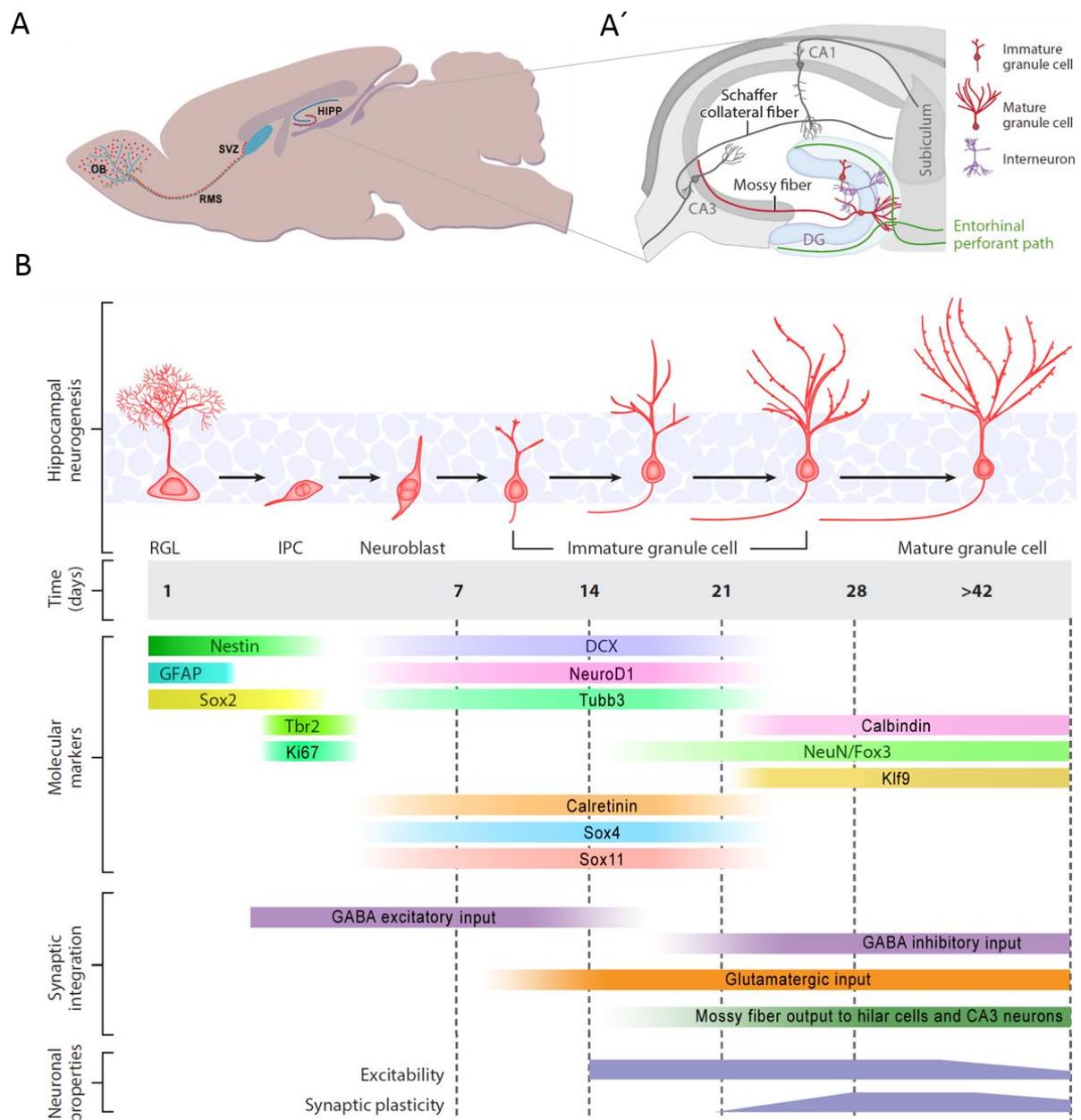


Fig 1. Adult neurogenesis in the adult mouse hippocampus.

A) A sagittal section view of an adult mouse brain highlighting the two regions of ongoing adult neurogenesis: i) the subgranular zone (SGZ) of the dentate gyrus (DG) within the hippocampus (HIPP) and ii) the subventricular zone (SVZ) of the lateral ventricle from where newborn neurons migrate along the rostral migratory stream (RMS) to the olfactory bulb (OB). Newborn neurons are marked in red.

A') Anatomy of the hippocampus showing the connections of neurons within the trisynaptic circuit. Mature granule neurons receive inputs from the entorhinal cortex via the perforant path, and extend axonal projections to pyramidal cells of the Cornu ammonis area (CA)3 via the mossy fiber pathway. CA3 pyramidal cells in turn pass on the information through Schaffer collateral projections to neurons in the CA1 subfield. CA1 neurons project to the Subiculum and finally back to the entorhinal cortex, from where the information is redistributed to neocortical areas (Henke 2010, Neves et al 2008, Song et al 2012a). Besides excitatory glutamatergic neurons and pyramidal neurons within the CA3 and CA1 region, inhibitory interneurons that release GABA are important for feedforward and feedback inhibition (Temprana et al 2015) and dictate the activation and self-renewal mode of quiescent adult neural stem cells (Song et al 2012b). **B)** Summary of the developmental maturation of adult hippocampal neurogenesis within the first 6 weeks, including developmental stages, marker expression, synaptic integration and neuronal properties associated with different maturation stages (modified from Christian et al 2014).

3.1.2 Development of newborn neurons in the adult hippocampus

The development of newborn neurons proceeds in a stereotypic manner in the dentate gyrus (Fig 1B).

During the first week after cell division, newborn neurons transform from fastly dividing progenitor cells to postmitotic immature neurons. Within the first week they start making first neurite outgrowth and become polarized. While newborn neurons mature they migrate a short distance into the inner granule cell layer. By using clonal lineage tracing of active radial glia-like (RGL) neural stem cells and their progeny, it was shown that 7 dpi newborn neurons already are capable of migrating tangentially along blood vessels away from the RGL cell. Radial migration from the SGZ towards the GZ is observed at later developmental stages (Sun et al 2015a). Newborn neurons are strongly regulated by neuronal activity and neurotransmitters. Functional GABA receptors are present as early as in type 2 progenitor cells and responsive to GABAergic inputs from hippocampal interneurons within four days of birth (Song et al 2013, Tozuka et al 2005). One-week-old immature neurons are sensitive to locally diffusing GABA released by interneurons (tonic GABAergic activation), which is depolarizing and critical for their development (Espósito et al 2005, Ge et al 2006, Toni & Schinder 2016). Notably, the suppression of interneuron activity in the adult dentate gyrus decreases survival of newborn progeny (Song et al 2013), again suggesting that tonic GABAergic activation by interneurons is critical for development and survival of newborn neurons. Indeed, a very critical period for survival seems to be the transition from amplifying

progenitor cells to neuroblasts, as the majority of the newborn cells undergo death by apoptosis in the first 1 to 4 days after neuronal birth (Sierra et al 2010).

During the second week, newborn neurons start to extend their axonal processes into the hilus towards CA3 and their dendritic processes towards the molecular layer (Sun et al 2013). Electrophysiological recordings have shown that they receive direct, slow GABAergic synaptic input, which is still excitatory (Espósito et al 2005, Ge et al 2006). GABA is known as the major inhibitory neurotransmitter in the adult brain, but due to the fact that newborn neurons express a chloride importer (NKCC1), which results in high intracellular chloride concentrations, stimulation of GABA receptors results in cell membrane depolarization. This depolarization is crucial for the proper development of newborn neurons, as an experimental conversion of the GABA-induced depolarization (excitation) into hyperpolarization (inhibition) impairs their development, dendritic spine formation and synaptic integration *in vivo* (Ge et al 2006). First glutamatergic synaptic responses emerge in 11-14-day-old newborn neurons, which they receive from hilar mossy cells and the entorhinal cortex (Chancey et al 2014, Espósito et al 2005, Ge et al 2006, Zhao et al 2006). At the same time, by 10-11 days after birth, axonal projections of newborn neurons reach the CA3 region, but not yet establish functional axonal synapses (Sun et al 2013, Zhao et al 2006).

During the third week, newborn neurons continue to extend their dendritic processes further into the molecular layer and develop a highly branched dendritic tree. In addition, they extend their axons more distally into the CA3 region. After an initial phase of rapid axonal and dendritic growth between 7 and 21 days, there is a phase of modest growth for at least two months (Sun et al 2013). Cells start to express the chloride exporter KCC2 which is believed to be responsible for the conversion from depolarization to hyperpolarization by GABA (Ge et al 2006). Using paired-recording (Markwardt et al 2011), optogenetics (Song et al 2013, Temprana et al 2015) and rabies virus-mediated retrograde transsynaptic tracing (Deshpande et al 2013, Li et al 2013, Vivar et al 2012) it is possible to study the network connectivity of newborn neurons within the hippocampus. Besides early inhibitory GABAergic interneuron innervation by different types of local interneurons (Li et al 2013, Markwardt et al 2011, Song et al 2013), newborn neurons receive first excitatory glutamatergic input by hilar mossy cells, followed by excitatory input from the medial and

lateral entorhinal cortex and cholinergic input by septal cholinergic cells (Chancey et al 2014, Deshpande et al 2013, Vivar et al 2012). At the output side, axons establish functional synapses with hilar interneurons, mossy cells and CA3 pyramidal cells as early as 16 days after neuronal birth and release glutamate as their main neurotransmitter (Gu et al 2012, Toni et al 2008, Toni & Schinder 2016). Thus, 16-day-old adult-born neurons already exhibit functional synaptic inputs and outputs and can participate in neural processing.

After the third week, newborn neurons progressively increase their dendritic arborization and continuously generate more spines. Spine density significantly increases between 21 and 28 days and reaches a plateau by 56 days (Zhao et al 2006). 30 days after birth, new neurons receive a diversity of inputs similar to mature granule neurons, including axosomatic, axodendritic and axospinous synapses, as shown by electron microscopy (Toni et al 2007). Even though the gross maturation is finished by the end of 4 weeks, dendritic spines and mossy fibre boutons continue to mature, and their connectivity changes up to 8 weeks after birth. When newborn neurons become integrated, they preferentially contact multiple synapse boutons of already existing synapses. Electron microscopy analyses suggest that dendritic spines from newborn neurons compete with dendritic spines from older neurons, and eventually replace them (Toni et al 2007). A similar competition was described on mossy fibre axon terminals with dendritic shafts from CA3 pyramidal neurons (Toni et al 2008). This shows that competition with mature granule neurons takes place both at the synaptic input and the synaptic output side and that synaptic partners are shared with pre-existing neurons for a couple of weeks, until they are replaced.

Within this competition, an advantage of newborn neurons over mature granule neurons might be the higher degree of plasticity. Electrophysiological recordings have demonstrated that 4-8 week old newborn neurons show increased synaptic plasticity and excitability. Long-term potentiation (LTP) exhibits a lower threshold for induction and a larger amplitude in newborn neurons compared to mature granule neurons within 4-8 weeks (Dieni et al 2013, Ge et al 2007, Marín-Burgin et al 2012, Mongiat & Schinder 2011, Schmidt-Hieber et al 2004). After 2 months, newborn neurons (labeled in adult 6 week old mice) become indistinguishable from neurons born during embryogenesis (labeled in pups at postnatal day 7) (Laplagne et al 2006).

The activity-regulated synapse formation could be a selection mechanism for the integration of neurons at the appropriate time and place. Indeed, the 3rd week after neuronal birth is another critical period of survival coinciding with the time when cells develop synapses and integrate into the network (Aasebø et al 2011, Tashiro et al 2006). Due to the fact that the survival of new neurons is regulated in an experience-dependent manner (Fabel et al 2009, Tashiro et al 2007), it was postulated that the survival or death decision might be information-specific, dependent on the synaptic input they receive before they get integrated into the network. It was shown that cells lacking glutamatergic synaptic input activity, mediated by cell-specific deterioration of the glutamatergic NMDA-receptor, died after the second week of neuronal age (Tashiro et al 2006). This period might be especially sensitive for the process of learning and memory. Similar as during embryonic development, neurons are generated in excess in the adult brain but only a fraction is needed for proper development and function. Thus, only cells needed for learning or the encoding of a new memory trace may receive synaptic input and become integrated into a very flexible and highly connected network (“use it or lose it”) (Shors et al 2012). Ultimately, less than a third of newly generated neurons survive and are integrated into the hippocampal network (Kempermann et al 2003, Kuhn 2015, Sierra et al 2010).

3.1.3 Functional significance of adult hippocampal neurogenesis

Newborn neurons have been implicated in a wide range of behaviours, including novelty recognition, pattern separation, spatial learning, anxiety behaviours, and antidepressant response (reviewed in Abrous & Wojtowicz 2015, Wu et al 2015). Very well accepted is the functional involvement of newborn neurons in the encoding of subtle changes in space and time, the discrimination of highly similar representations and contexts, a process that is called pattern separation (Clelland et al 2009, Creer et al 2010, Sahay et al 2011a, Sahay et al 2011b, Tronel et al 2012). Interestingly, while young adult-born neurons (< 4 weeks) were shown to be involved in pattern separation-mediated encoding, old hippocampal neurons are involved in pattern completion-mediated recall (Nakashiba et al 2012). Besides the role in encoding detailed and vivid information, adult-born neurons actively participate in

consolidation (Kitamura et al 2009), retrieval processes (Arruda-Carvalho et al 2011, Gu et al 2012, Kee et al 2007, Liu et al 2012) and “forgetting” (Akers et al 2014).

Moreover, altered levels of adult neurogenesis have been linked to several neurodegenerative and psychiatric disorders (Apple et al 2016, Kang et al 2016, Winner & Winkler 2015). While stress and depression contribute to decreased neurogenesis, antidepressant treatment has been shown to reverse this loss of hippocampal neurogenesis (Mahar et al 2014). Notably, adult hippocampal neurogenesis is required for the efficacy of specific classes of antidepressants (Sahay & Hen 2007). After disruption of adult neurogenesis the ameliorative effect of the antidepressant fluoxetine, a selective serotonin reuptake inhibitor, was blocked (Santarelli et al 2003). However, also neurogenesis-independent effects of antidepressants have been reported and not all antidepressants require neurogenesis for their actions (David et al 2009). Hence, whether adult hippocampal neurogenesis is linked to antidepressant treatment or might even be an etiological factor for depression is still controversially debated (DeCarolis & Eisch 2010, Eisch & Petrik 2012, Miller & Hen 2015, Sahay & Hen 2007). Further investigation of the precise functional impact of adult neurogenesis on disease onset, symptoms and progression will provide new therapeutic strategies for the treatment of mental and neurodegenerative diseases.

3.1.4 Transcriptome analysis of newborn neurons within the two neurogenic niches

Newborn neurons within the SGZ niche have been characterized in great detail with regard to marker expression, morphology, synaptic integration and electrophysiological properties (Christian et al 2014, Zhao et al 2008) (Fig 1B). An important missing piece towards the understanding of the mechanisms regulating neural stem cell differentiation into functional neurons, however, is the knowledge of the regulatory networks that control distinct developmental phases in adult neurogenesis. Therefore, the overall aim of this project is the characterization of the molecular networks at distinct stages of neuronal maturation.

Recent work has begun to tackle this issue by performing microarray-based transcriptome analysis from LCM-dissected SGZ and GZ (Miller et al 2013) or from FACS- isolated neural stem (NSC) and progenitor cells using transgenic mouse models (Bracko et al 2012) or

surface markers (Beckervordersandforth et al 2010). These studies provided a general picture of the SGZ niche, and the identification of new regulatory pathways for early stages of adult SVZ and hippocampal neurogenesis, respectively. However, the regulatory networks controlling maturation and functional integration of newborn neurons have eluded systematic analysis, in part because of the lack of FACS-compatible markers and transgenic mouse models to separate neuronally fate-committed cells of distinct maturational stages with high temporal resolution.

3.1.4.1 Putative solutions for transcriptomic analysis of maturation

Solutions to the problem of isolating immature neurons at distinct maturation stages are i) to use a retroviral birthdating strategy or ii) to profile hundreds of single cells combined with bioinformatic clustering techniques which allow the classification of developmental stages thereafter.

The retrovirus birthdating strategy allows to label a cohort of cells from the same developmental stage with high temporal resolution, as the replication deficient recombinant murine-moloney-leukemia virus (MMLV) integrates only into dividing cells, i.e., activated neural stem cells and highly dividing intermediate progenitor cells in the adult dentate gyrus. Thus, the corresponding labelled cells represent cells of the same developmental stage. This strategy has become a methodological mainstay to birth-date adult-born neurons and to study the morphology (Sun et al 2013, van Praag et al 2002, Zhao et al 2006), synaptic integration (Deshpande et al 2013) or electrophysiological properties (Laplagne et al 2006, Schmidt-Hieber et al 2004) of newborn neurons, or even to manipulate signaling pathways (Ge et al 2006, Jagasia et al 2009, Mu et al 2012) in newborn neurons of the adult brain at particular developmental stages. However, the isolation of retrovirally labelled cells from the adult brain has not been established so far and will be one specific aim of this project.

Bioinformatic clustering tools have been established recently that order single-cell gene expression dynamics along a pseudotemporal timeline, thereby generating different developmental subgroups (Shin et al 2015, Trapnell et al 2014). Such computational ordering strategies allow the unbiased statistical analyses of multi-dimensional single-cell datasets from continuous biological processes without the need of *a priori* selection of cells based on

specific markers. Recently, single-cell RNA-Seq analyses of adult neural stem cells (NSCs) in both SVZ (Llorens-Bobadilla et al 2015, Luo et al 2015) and SGZ (Shin et al 2015) combined with bioinformatic clustering tools have started to provide a holistic picture of molecular signatures of quiescent adult NSCs and molecular dynamics of NSC activation and neurogenesis. Even though this strategy has the potential to resolve transcriptional dynamics of later stages of immature neuron development as well (see 5.3.3.1), so far such analyses have only been applied on neural stem cells and early progenitor cells. Moreover, these three studies are not directional and do not (Shin et al 2015) or only marginally (Llorens-Bobadilla et al 2015, Luo et al 2015) describe non-coding RNA (ncRNA) expression in their datasets, which are emerging regulators of transcription and epigenetic modifications.

3.2 The role of long non-coding RNAs

For half a century the central role of RNA was believed to serve as a template for protein synthesis, termed messenger RNA (mRNA) (Brenner et al 1961, Jacob & Monod 1961). With the discovery of miRNAs it became clear that RNAs can also have regulatory functions (Lagos-Quintana et al 2001). Nowadays, several classes of short regulatory RNAs have been discovered and described in great detail, such as microRNAs (miRNAs), short interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), and other short RNAs (reviewed in Carthew & Sontheimer 2009, Gangaraju & Lin 2009, Ghildiyal & Zamore 2009, Hirose et al 2014, Malone & Hannon 2009, Siomi & Siomi 2009, Siomi & Siomi 2010, Winter et al 2009). However, the complete picture of gene regulation is just now starting to be revealed (Morris & Mattick 2014). The understanding of the architecture, activity and regulation of the mammalian genome has been revolutionized by the development of new sequencing technologies and findings from large-scale consortia such as ENCODE (ENCODE 2012) and FANTOM (Carninci et al 2005, Katayama et al 2005), which focus on the characterization of functional genomic elements. This work revealed that a large fraction of the genome is transcribed into what was previously regarded as “dark matter” – non-coding RNAs (ncRNAs) that do not encode information about proteins (Djebali et al 2012, ENCODE 2012). By analyzing a large set of RNA-Seq data including 127 datasets from 23 human tissues it was shown that while only 2% of the genome constitute protein-coding genes, 85% of the genome is transcribed (Hangauer et al 2013). However, this already large number may still be underestimated, as only a subset of cell states have been assayed (Kellis et al 2014). Interestingly, the relative amount of non-protein-coding sequence increases consistently with organismal complexity (Mattick 2011, Taft et al 2007), indicating that some ncRNAs may have functional roles. Among all classes of ncRNAs, long ncRNAs (lncRNAs) represent the most prevalent and functionally diverse class (Mercer et al 2009, Wilusz et al 2009).

LncRNAs are defined as RNA genes greater than 200 nucleotides in length with no coding potential (Kapranov et al 2007, Ponting et al 2009, Wilusz et al 2009). The limit of 200 nt in length distinguishes long ncRNAs from short regulatory RNAs. LncRNAs are transcribed from intergenic or intragenic regions and vary in length from 200 nt to over 100 kb (Kapranov et al 2007).

LncRNAs are a heterogeneous group in terms of their genomic context, size, regulation, mechanisms of action and functional repertoire. Many attempts have been made to group them based on genomic context, sequence and structure conservation, association with protein-coding genes or other DNA elements (e.g., enhancer-associated, promoter-associated), association with repeats, biochemical pathways or different biological states, or association with subcellular structures (e.g., chromatin-associated) (St. Laurent et al 2015). As more lncRNAs are being functionally characterized, soon a more precise definition and functional classification based on their effect exerted on DNA sequences, mechanism of functioning and their targeting mechanism will be possible (Ma et al 2013).

Most of the characterized long ncRNAs are generated by the same transcriptional machinery as mRNAs: they are transcribed by Pol II and show the same histone modifications associated with transcription initiation (histone H3K4 trimethylation at their 5' end) and elongation (histone H3K36 trimethylation in the gene body) (Guttman et al 2009). Several lncRNAs are processed like mRNAs and have a 5' terminal methylguanosine cap and 3' polyadenylation signals (Ng et al 2013, Qureshi & Mehler 2012). The expression of lncRNAs can be modulated by transcriptional and epigenetic regulatory factors. They can be subject to post-transcriptional processing, intracellular and intercellular transport, and associate with a myriad of RNA binding proteins and molecules. Accordingly, lncRNAs harbor sequence motifs and elements that regulate their expression (e.g., transcription factor binding sites, miRNA binding sites), post-transcriptional processing (e.g., splicing sites, polyadenylation sites), binding (e.g., binding domains) and folding into secondary or higher order structures (Qureshi & Mehler 2013).

LncRNAs are predominately localized in the nucleus and associated to chromatin (Derrien et al 2012b), suggesting that lncRNAs may have a significant impact on DNA sequences and are devoted to gene regulation in the nucleus.

Many studies found the expression level of lncRNAs to be on average approximately 10-20-fold lower than protein-coding transcripts (Derrien et al 2012a, Derrien et al 2012b, Lv et al 2013). However, the lower expression levels of lncRNAs may not necessarily result from a general low copy number but may result from restricted expression in only a subpopulation of cells. For some individual lncRNAs it was shown that the expression level can be as high as those of protein-coding genes in some cell lines (Djebali et al 2012). This fact forces the

need of analysis at the single-cell level in order to obtain a complete picture of expression and to be able to distinguish between causes and consequences of lncRNA expression. What is more, it reinforces the need to use deep sequencing-based technologies to identify these low expressed ncRNAs.

Long ncRNAs have been observed in a large diversity of species, including animals, plants, yeast, prokaryotes and even viruses. Intriguingly, the proportion of noncoding DNA increases with developmental complexity (Kapusta & Feschotte 2014, Mattick 2011, Ponting et al 2009, Sabin et al 2013, Taft et al 2007, Ulitsky & Bartel 2013), suggesting that lncRNAs are one major driving source for the rapid evolution of developmental complexity and cognitive development (Mattick 2011). LncRNAs are poorly conserved among different species and it is suggested that this poor conservation may be the result of recent and rapid adaptive selection (Pang et al 2006, Ponjavic et al 2007). Interestingly, transposable elements occur in more than two thirds of mature lncRNA transcripts, covering approximately 30% of total lncRNA sequence in humans (Kapusta et al 2013), indicating that retrotransposition of functional modules may be a major source for the rapid evolution of lncRNAs. Moreover, transposable elements have been found in particular at transcription start sites of lncRNAs (Kapusta et al 2013, Kelley & Rinn 2012), suggesting that transposable elements play a role in the regulation of lncRNA transcription, thus possibly being responsible for the highly cell- and tissue-specific expression of lncRNAs.

The poor evolutionary conservation among different species and the low level of expression have questioned the functionality of these lncRNAs, raising the possibility that most lncRNAs are just non-functional byproducts of transcription from neighbouring loci (Ebisuya et al 2008, Ponjavic et al 2007, Ponting & Hardison 2011, Ponting et al 2009, Struhl 2007).

Even though just a few lncRNAs have been functionally characterized, there are compelling indices of their functionality:

- i) Tissue- and cell-specific expression patterns and subcellular localization (Cabili et al 2015, Cabili et al 2011, Derrien et al 2012b, Djebali et al 2012, Gloss & Dinger 2016, Goff et al 2015, Hu et al 2014, Kadakkuzha et al 2015, Mercer et al 2008)
- ii) Dynamic expression and alternative splicing during differentiation (Dinger et al 2008, Hu et al 2012, Ramos et al 2013, Yan et al 2013)

- iii) Conservation of RNA secondary structure (Mercer & Mattick 2013, Torarinsson et al 2006, Torarinsson et al 2008). Higher conservation of promoters, splice junctions and exons than neutrally evolving ancestral repeat sequences, albeit at lower levels than protein-coding genes (Derrien et al 2012b, Ponjavic et al 2007)
- iv) Association with particular chromatin signatures that are indicative of actively transcribed genes (Guttman et al 2009, Khalil et al 2009, Lv et al 2013, Ramos et al 2013)
- v) Effect on gene expression as shown by gain and loss-of-function experiments and knock-out models (Clark & Blackshaw 2014, Guttman et al 2011, Khalil et al 2009, Sauvageau et al 2013)
- vi) Regulation of their expression by key morphogens, transcription factors and hormones (Bhan et al 2013, Guttman et al 2009, Mattick 2011)
- vii) Altered expression or splicing patterns in cancer and other diseases (reviewed in Batista & Chang 2013, Gutschner & Diederichs 2012, Niland et al 2012, Qureshi & Mehler 2013)

Moreover, genome wide association studies (GWAS) of cancer risk loci revealed that 80 to 90 % of cancer-associated SNPs (single nucleotide polymorphisms) occur in noncoding regions of the genome (Cheetham et al 2013, Freedman et al 2011).

A large fraction of lncRNAs is specifically expressed in the brain (Lv et al 2013, Mattick 2011, Mercer et al 2008) and thus, many studies have focused on the involvement of lncRNAs in neuronal development, function and disease (Antoniou et al 2014, Aprea & Calegari 2015, Clark & Blackshaw 2014, Fatica & Bozzoni 2014, Hu et al 2012, Iyengar et al 2014, Ng et al 2013, Qureshi & Mehler 2013, Qureshi & Mehler 2012, Vučićević et al 2014, Wu et al 2013).

3.2.1 Putative functions of lncRNAs

3.2.1.1 Transcriptional and epigenetic regulation

Long ncRNAs have been mostly implicated in epigenetic and transcriptional regulation (Batista & Chang 2013, Bonasio & Shiekhattar 2014, Rinn & Chang 2012, Sabin et al 2013). Functionally characterized lncRNAs are involved in the regulation of genome organization

and gene expression (Quinodoz & Guttman 2014, Rinn & Guttman 2014). They can regulate gene transcription through transcriptional interference and chromatin remodeling (Fig 2). With regards to transcriptional interference, lncRNAs can influence the transcription activity of target genes through promoter binding to block pre-initiation complex formation, or by interacting with transcription factors (Kornienko et al 2013, Ma et al 2013).

On the epigenetic level, lncRNAs can influence chromatin modifications by recruiting chromatin modification complexes (e.g., PRC = polycomb repressive complexes). Even though the repression of transcription is the principal outcome of Polycomb complexes, it is still elusive how they are recruited to specific genomic loci in mammals, as they do not contain sequence specific DNA binding proteins (Schuettengruber et al 2007, Schwartz & Pirrotta 2007, Simon & Kingston 2009). lncRNAs can bridge that gap because they can fold up to bind proteins and at the same time are inherently sequence-targeted. Due to their propensity to fold into thermodynamically stable secondary and higher order structures (Mercer & Mattick 2013, Wan et al 2011), lncRNAs have been suggested to serve as a scaffold for the chromatin-modifying machinery, while at the same time being able to provide sequence specificity and guide them to the correct gene loci by interacting either with DNA or the nascent transcript (Guttman & Rinn 2012, Spitale et al 2011). This would then explain how the same protein complexes can act on different sequences in different cells. Indeed, large-scale and genome-wide studies of RNA-protein interactions have shown that a large number of lncRNAs interact with chromatin-modifying complexes, such as PRC2 (Khalil et al 2009, Zhao et al 2010). By using RIP-Seq (RNA-immunoprecipitation Seq) more than 9.000 PRC2-associated transcripts have been identified (Zhao et al 2010).

Until now, most lncRNAs have been identified to associate with repressive chromatin marks, however, some also serve as activators of gene expression, e.g., by recruiting the H3K4 methyltransferase MLL1 (Bertani et al 2011, Wang et al 2011).

With regards to their site of action, lncRNAs can be classified into *cis*-regulating lncRNAs, thereby regulating the expression of genes in close genomic proximity, or in *trans*-regulating lncRNAs, regulating the expression of distant genes.

3.2.1.2 Post-transcriptional regulation

In addition to the well-established function of lncRNAs in transcriptional and epigenetic gene regulation, lncRNAs can also promote and inhibit the post-transcriptional processes of mRNA splicing, degradation and translation (Yoon et al 2012a) (Fig 2).

Some lncRNAs are involved in alternative splicing, e.g., by binding and sequestering several splicing factors to nuclear speckles, thereby altering splicing factor localization and activity (Tripathi et al 2010). In hippocampal neurons, the regulation of splicing factors by the lncRNA MALAT1 is important for synapse formation. Knock-down of MALAT1 reduces synaptic density, whereas overexpression increases it, suggesting a role in synapse assembly through transcription-coupled alternative splicing (Bernard et al 2010).

In addition it was shown that some lncRNAs interact with miRNAs and “sponge” them, thus preventing their function and stabilizing the target mRNAs (Cesana et al 2011, Salmena et al 2011). Besides functioning as a sponge, some antisense lncRNAs may bind to mRNA to mask the binding sites of miRNA and thus stabilize the mRNA (Faghihi et al 2008, Faghihi et al 2010). Still, it is suggested that the interaction of lncRNAs and miRNAs is reciprocal and that lncRNAs are themselves subject to miRNA-mediated destabilization (Jalali et al 2013, Paraskevopoulou & Hatzigeorgiou 2016).

Moreover, lncRNAs may exert their effects even at the level of translational regulation. Recently, lncRNAs have been identified to inhibit (Yoon et al 2012b) and to promote translation (Carrieri et al 2012), respectively.

Based on these examples, lncRNAs have been classified according to their mode of action into four archetypes (Wang & Chang 2011). lncRNAs can act as i) signals integrating temporal, spatial, developmental and stimulus-specific cellular information; ii) decoys with the ability to sequester a range of RNA and protein molecules, thereby precluding the access of regulatory proteins to DNA and inhibiting their functions; iii) guides for genomic site-specific and more widespread recruitment of transcriptional and epigenetic regulatory factors; and iv) scaffolds for macromolecular assemblies with varied functions.

Notably, lncRNAs can exert several functional roles at the same time, i.e., they can be induced by different endogenous and exogenous signals (thus working as a signal), but at the same time recruit chromatin modifying complexes to repress or activate expression of a set of genes (thus working as a guide or scaffold).

As transcription of lncRNAs is very cell-specific and timely controlled (Cabili et al 2015, Cabili et al 2011, Derrien et al 2012b, Djebali et al 2012, Gloss & Dinger 2016, Hu et al 2014, Mercer et al 2008) lncRNAs may be capable of integrating developmental cues, interpreting cellular contexts and in turn respond to diverse stimuli. lncRNAs may co-operate with chromatin modifiers, transcription factors and miRNAs to ensure precise gene expression at both transcriptional and post-transcriptional levels (Hu et al 2012) (Fig 2). As such, lncRNAs can act as key regulatory nodes in multiple transcriptional pathways. Due to their specific expression and function, lncRNAs are now being discussed as potential biomarkers and therapeutic targets in cancer and other diseases (Fatima et al 2015, Yarmishyn & Kurochkin 2015).

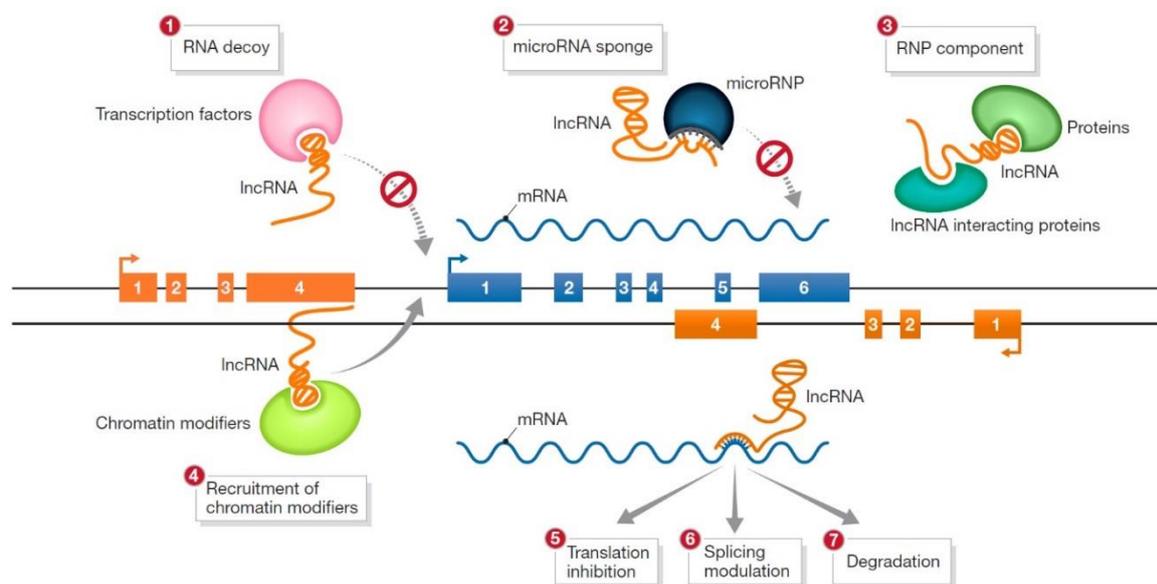


Fig 2. Mechanisms of long non-coding RNA function.

lncRNAs (orange) can be transcribed from intergenic or genic regions and exert several functional roles. They can act as a decoy (1) or sponge (2) to prevent the binding of transcription factors or miRNAs to target sites. They can also act as a scaffold for the composition of ribonucleoprotein (RNP) complexes (3) or recruit chromatin modifiers and guide them to specific target sites (4). Moreover, they can be involved in post-transcriptional processes like translation (5), splicing (6) or degradation (7) (modified from Hu et al 2012).

3.2.1.3 Catalogization of lncRNAs in databases

Due to a dramatic increase of genome-wide sequencing studies, the identification of lncRNAs increased dramatically in the past few years. Several lncRNA databases have been developed in order to organize the increasing number of lncRNAs (Table 1). These databases rely on experimental and computational evidences and integrate this information with other “omics” information.

The LNCipedia database (www.lncipedia.org), which curates human lncRNA data from a number of different sources, is currently the most comprehensive database, currently containing 113,513 human annotated transcripts (Volders et al 2015). For mouse and also other species, the most comprehensive database is NONCODE (Xie et al 2014). NONCODE v4.0 (www.bioinfo.org/noncode/) is a database of literature documented lncRNAs, currently containing a total of 210,831 lncRNA entries from 1,239 different organisms. Within this database 67,628 lncRNA transcripts are documented for mouse and 95,135 for human (Xie et al 2014).

Still, only a small fraction of these documented lncRNAs (just a few hundred) have been functionally annotated according to the lncRNAdb (www.lncrnadb.org), a database of functionally annotated lncRNAs (Amaral et al 2011, Quek et al 2015). This means that the functional significance of the majority of lncRNAs has yet to be elucidated.

Table 1

Database	Description	Species	number of lncRNAs
NONCODE v4.0 (Xie et al 2014)	A database of literature documented long ncRNAs, including detailed annotation, expression and potential functional annotations	1239 different organisms	95135 human, 67628 mouse transcripts (56018 human, 46675 mouse genes)
Lncipedia (Volders et al 2015)	A database for annotated human lncRNA transcript sequences and structures, including structural information, miRNA binding sites and protein-coding potential	human	113513 transcripts (63038 genes)
LncRNAtor (Park et al 2014)	A database for functional investigation of lncRNAs, including annotation, sequence analysis, gene expression, protein binding and phylogenetic conservation	human, mouse, zebrafish, fly, worm, yeast	34605 transcripts
lncNome (Bhartiya et al 2013)	A comprehensive database of human lncRNAs, including sequence and structure analysis, RNA-protein interactions, miRNA binding sites, methylation and histone modifications	human	18855 transcripts (11790 genes)
ChIPBase (Yang et al 2013)	A database for decoding the transcriptional regulation of lncRNA and microRNA genes from ChIP-Seq data	human, mouse, dog, chicken, fly, worm	38293 transcripts
Starbase v2.0 (Li et al 2014)	A database for protein-RNA interactions and miRNA-RNA interactions derived from large-scale CLIP-Seq data	human, mouse, C.elegans	> 700,000 interactions human, > 100,000 interactions mouse
DIANA-lncBase (Paraskevopoulou et al 2013)	A database of experimentally verified and computationally predicted microRNA targets on lncRNAs	human and mouse	56110 transcripts
lncRNAMap (Chan et al 2014)	A database of human lncRNA-derived siRNA-target interactions, deciphering the putative regulatory functions of lncRNAs as miRNA decoy and encoding siRNAs	human	> 20,000 lncRNA transcripts
Linc2go (Liu et al 2013)	A human lincRNA function annotation database based on ceRNA hypothesis, including miRNA-lincRNA interactions and miRNA-mRNA interactions	human	2 198 132 human miRNA-lincRNA interactions; 1 218 961 miRNA-mRNA interactions
lncRNADisease (Chen et al 2013)	A database for long-non-coding RNA-associated diseases, including 166 diseases	human	480 experimentally supported and 1564 predicted human lncRNA-disease associations
lncRNAdb v2.0 (Quek et al 2015)	A database of functionally annotated lncRNAs (literature based, manually curated), very detailed annotation and information	many eukaryotic species	287 lncRNAs
Functional lncRNA db (Niazi et al 2012)	A database of lncRNAs manually extracted from the literature, including description of cellular function, miR precursor, repeats	human, mouse, rat	204 lncRNAs
NRED (Dinger et al 2009)	A database of lncRNA expression, including evolutionary conservation, secondary structure evidence, genomic context links and antisense relationships	human, mouse	11926 mouse lncRNA probes, 1287 human lncRNA probes
NATsDB (Zhang et al 2007)	A database of natural antisense transcripts, including information on gene structures, poly(A) signals and tails, conservation, repeat elements, expression and disease association	several eukaryotic species	7356 human, 6806 mouse sense-antisense pairs
PlncDB (Jin et al 2013)	A database of plant lncRNAs, including genomic information, expression and epigenetic modifications	plant	16227 lncRNAs

Table 1. Comparison of different lncRNA databases with respect to their functional features, species analyzed, and number of lncRNAs (modified from Fritah et al 2014, Iwakiri et al 2016).

3.2.2 Antisense transcription

More than 30% of annotated transcripts in humans have antisense transcription, as discovered by direct RNA-Sequencing (Ozsolak et al 2010). However, antisense transcripts are on average more than 10-fold lower in abundance than sense transcripts (Derrien et al 2012b, Ozsolak et al 2010) and are preferentially located in the nucleus (Derrien et al 2012b), while protein-coding mRNAs accumulate in the cytoplasm (Djebali et al 2012).

Natural antisense transcripts (NATs) have been shown to influence almost all stages of gene expression, including transcriptional initiation, co-transcriptional processes and post-transcriptional processes (reviewed in Pelechano & Steinmetz 2013). They can exert their function either through the act of transcription but also through the non-coding RNA that is produced (Kornienko et al 2013). In addition, they can affect gene expression either in *cis* (on the same genomic locus) or in *trans* (on different DNA strands). The fact that both antisense and sense transcripts are transcribed from the same region suggests that antisense transcripts function more frequently in *cis* than other classes of ncRNAs. Mechanistically it is suggested that antisense transcripts act as regulatory hubs, generate self-regulatory circuits that allow fast adaptation to new conditions, or may be responsible for rewiring regulatory networks, thereby switching transcription factor functions between activators and repressors (Pelechano & Steinmetz 2013). Indeed, it was shown in yeast that genes that have to respond in a switch-like manner between on-off states, such as stress response and environment-specific genes, are enriched for antisense expression (Xu et al 2011). Mechanistically it is suggested that sense and antisense transcripts reciprocally inhibit each other up to a certain threshold. In the absence of sufficiently strong activating signals, the antisense transcript is expressed and inhibits sense expression. Upon activation and environmental changes that reach a certain threshold, the sense gene promoter gets disinhibited and consequently switched on (Xu et al 2011). Even though this self-regulatory circuit is an appealing model, the reciprocal inhibition of sense-antisense pairs seems to be dependent on the expression level. While highly expressed sense transcripts tend to negatively correlate with the expression of antisense transcripts, lowly expressed genes show a positive correlation between sense and antisense transcription (Ozsolak et al 2010). Thus, the mode of reciprocal regulation has to be examined carefully for each transcript pair and condition.

In addition to the function as a regulatory switch, antisense transcripts have been shown to function as a modular scaffold for protein complexes, similar as other lncRNA classes. In fact, guidance of Polycomb proteins by antisense transcripts is likely to be common, as PRC2 directly interacts with more than 9000 lncRNAs, of which 3000 are antisense transcripts (Zhao et al 2010).

The reliable detection of antisense transcripts requires strand-specific approaches. Due to the overlapping nature of antisense transcripts to protein-coding sense transcripts, many studies mistakenly identified antisense transcripts as sense transcripts, without strand-specific approaches.

3.2.3 The role of lncRNAs in adult neurogenesis

Recently, the first systematic analysis of the lncRNA expression profile associated with neuronal fate determination and differentiation of adult neural stem cells in the adult SVZ has been reported (Ramos et al 2013). In this study custom lncRNA microarrays were generated based on RNA-Seq and RNA-CaptureSeq datasets from regions of ongoing adult neurogenesis, resulting in the identification of 8992 lncRNAs. Hybridization of these arrays with expression libraries derived from FAC-sorted stem and precursor cells, and neuroblasts of the adult SVZ system identified several lncRNAs that are differentially expressed between these early developmental stages, strongly suggesting that stage specific lncRNAs participate in the control of stem cell activity and differentiation. Indeed, depletion of the lncRNA Six3os, which is highly expressed in subventricular zone stem cells and downregulated in neuroblasts, resulted in increased gliogenesis at the expense of neurogenesis in cultured neural stem cells, thereby providing proof-of-principle that lncRNAs regulate adult neural stem cell fate (Ramos et al 2013).

Thus, lncRNAs are evolving as an additional layer of regulation in adult neural stem cell biology. It is expected that the regulatory role of lncRNAs will not be confined to the regulation of neural stem cell fate but will also influence later steps in adult neurogenesis such as neuronal maturation and synaptic integration.

3.3 *RNA-Seq*

High throughput sequencing of cDNA (RNA-Seq) allows the detection and quantification of expressed transcripts in biological samples. Compared to microarray gene expression analysis, RNA-Seq provides an unbiased assessment of the full range of transcripts with a greater dynamic range and sequence coverage (Wang et al 2009). Thus, it offers the great advantage to identify differentially expressed genes with a greater dynamic range and to be able to identify novel transcripts and alternative splicing in an unbiased manner. Because of these advances, Next-Generation Sequencing has been declared as the method of the year in 2007 (2008, Chi 2008).

The three main second-generation sequencing platforms are Roche/454 platform (launched in 2005), Illumina/Solexa system (launched in 2006) and Applied Biosystems/SOLID system (launched in 2007). These platforms differ in their way of template pre-amplification (emulsion PCR amplification on beads or solid phase amplification) and their sequencing chemistry (pyrosequencing, sequencing with reversible terminators or sequencing by ligation). While pyrosequencing (Roche) and sequencing with reversible terminators (Illumina) are sequencing-by-synthesis reactions, SOLiD is based on a sequencing-by-ligation reaction (reviewed in Mardis 2008a, Mardis 2008b, Metzker 2010).

In the present study, libraries have been sequenced using Illumina sequencing, which relies on solid phase amplification followed by sequencing with reversible terminators.

During sequencing, base calls are recorded by the camera, which detects fluorescent events generated during the incorporation of labeled nucleotides. Since most imaging systems have not been designed to detect single fluorescent events, templates need to be amplified prior to sequencing, i.e., the clonal amplification of identical templates. During solid phase amplification, fragments of DNA are ligated to adapters, which bind at one end to a solid surface already coated with a dense layer of adapters. Its free end “bends over” and hybridizes to a complementary adapter on the surface, which initiates the synthesis of the complementary strand. This amplification is also called “bridge amplification” and generates clusters of ~ 1000 copies of single-stranded DNA molecules distributed randomly on the surface.

During Illumina sequencing, fluorescently labeled nucleotides are incorporated that have a 3' terminator on the base to ensure that only one nucleotide is being incorporated at a time. After the chemicals are washed out, the fluorescence is imaged by a camera, and then the terminator is cleaved so that the next nucleotide can be incorporated in a next sequencing cycle. However, as sequencing cycles progress, the clonal strands get out of phase with one another due to chemistry inefficiencies, leading to increased fluorescence noise and causing base-calling errors. Therefore, read lengths are typically short (currently approximately 100 nt), but have been increased from 35 nt reads at the very beginning. Longer sequencing reads are favoured, as they can be mapped to the genome more easily, thus, unambiguously being mapped also to repetitive elements.

After the reads are sequenced and their quality assessed, good quality reads are mapped to the genome, the reads are normalized and quantified and subsequently used for differential expression analysis. Several bioinformatic packages have been developed for each analytical step (reviewed in Conesa et al 2016, Garber et al 2011).

Quantification of mRNA transcripts is often reported as RPKM (reads per kilobase per million mapped reads) or FPKM (fragments per kilobase per million mapped reads), the equivalent for paired-end sequencing. RPKM/FPKM normalize for differences in gene length and thereby provide meaningful comparisons of expressed transcripts in terms of molar equivalents (Mortazavi et al 2008, Zeng & Mortazavi 2012). Since RPKM normalization is currently widely used, I used RPKM normalization for the quantification of reads mapping to different genomic regions, in order to be able to compare expression levels with other publications.

For differential expression analysis RPKM/FPKM normalization was shown to be less robust (Dillies et al 2013), and especially in samples with 3'-biases normalization approaches that adjust for transcript length are problematic (Stegle et al 2015). Compared to RPKM normalization methods, DESeq and TMM (Trimmed Mean of M values) are more robust normalization methods in the context of differential expression analysis (Dillies et al 2013). In the present study, differential expression analysis was performed by generating read counts using HTseq (Anders et al 2015) followed by normalization using DESeq (Anders & Huber 2010, Anders et al 2013, Love et al 2014) and edgeR (TMM) (Robinson & Oshlack 2010).

It has to be noted that HTseq is designed specifically for differential expression analysis, but not for transcript discovery, as only reads mapping unambiguously to a single gene are counted, whereas reads aligned to multiple positions or overlapping with more than one gene are discarded (Anders et al 2015). Therefore, for the analysis of transcript isoform expression or the identification of novel transcripts the software Cufflinks (Trapnell et al 2010) has been used, which estimates, rather than counts, expression levels of transcripts (Dillies et al 2013).

3.3.1 Directional RNA-Seq

Typically, conventional RNA-Seq library preparation procedures are not directional, as the shearing of double-stranded cDNAs leads to the loss of sequence orientation. However, the direction of transcription is crucial for resolving overlapping genetic features and detecting antisense transcription, which is critical for *de novo* transcript discovery. Therefore, alternative library preparation methods have since been developed that yield strand-specific reads.

In a comparison of different strand-specific RNA-Seq protocols starting with > 100 ng of poly(A)⁺ RNA from *S. cerevisiae* (Levin et al 2010), the dUTP second strand marking protocol (Parkhomchuk et al 2009) and the RNA ligation method (Lister et al 2008) have been the best-performing protocols with respect to strand specificity, coverage and accuracy. Using the dUTP method, dUTP (instead of dTTP) is incorporated during synthesis of the second-strand cDNA, which helps to distinguish the second strand cDNA from the first strand (Parkhomchuk et al 2009). In the RNA ligation method, 5' and 3' adaptors are sequentially ligated in predetermined directions to single-stranded RNA or the first-strand cDNA (Lister et al 2008).

Recently, a new ligation-free, terminal-tagging protocol has been developed (Pease & Sooknanan 2012) to circumvent the laborious and inefficient ligation procedures and coverage bias of both 5' and 3' ends of the RNA ligation method. In the new protocol, RNA is fragmented and reverse transcribed using random primers containing a 5'-tagging sequence. The 5'-tagged cDNA is then tagged at its 3'-end by the terminal-tagging reaction to yield di-tagged, single-stranded cDNA. This procedure allows rapid and efficient library

preparation, with high quality, strong directionality and good transcript coverage (Pease & Sooknanan 2012). Another advantage is that it allows paired-end sequencing, while the RNA ligation method only allows single-end sequencing. Still, currently these directional library preparation kits are not suitable for single-cell library preparation and require at least 500 pg to 50 ng of RNA as input (Epicentre, ScriptSeq v2 RNA-Seq library preparation kit, 2011).

3.3.2 Transcriptome analysis of single cells or a limited low number of cells

Typically, transcriptome analyses have been conducted using millions of cells, however these approaches mask cellular heterogeneity and dynamics, and cannot be applied to study rare and heterogeneous cell populations like stem cells, brain cells or cancer cells.

Single-cell analysis allows for the precise description of individual and rare cells and their developmental dynamics, a fact that would be masked on a population level. Especially lncRNAs are lowly, but specifically expressed in certain cell types at distinct developmental stages and many of them would be masked on a population level (Cabili et al 2015, Cabili et al 2011, Derrien et al 2012b, Mercer et al 2008, Wills et al 2013, Yan et al 2013, Zhang et al 2014). Moreover, studying cells at the single-cell level offers unique opportunities to dissect the interplay between extrinsic cues and cell intrinsic mechanisms in a complex multi-dimensional environment of neural networks. The advances in single-cell analysis greatly accelerate our understanding of the complexity of gene expression, regulation and networks. Thus, it is not surprising that single-cell analysis is the new frontier in 'omics' research (Wang & Bodovitz 2010).

The strength of single-cell analysis is definitely the in-depth characterization of single cell dynamics. However, it was shown that gene expression in individual cells is often stochastic (Raj et al 2006) and cell-to-cell variation a result of technical noise. Even highly expressed genes will show considerable cell-to-cell variation which is even more pronounced in low expressed genes. To decrease this noise one can either i) analyze large numbers of single cells or ii) pool several cells from the same cell type. Since the first option is very cost- and laborintensive, "multiplexing" has been developed. During multiplexing, each sample obtains a different barcode. These samples are pooled and sequenced together in a single

lane. Based on the barcode, samples can be distinguished thereafter. Multiplexing allows to sequence several biological samples simultaneously in a single sequencing lane and thus, saving money and time. Multiplexing can be achieved by adding barcoded primers either at the cDNA synthesis step (Hashimshony et al 2012, Islam et al 2011, Jaitin et al 2014, Shishkin et al 2015) (referred to as “early multiplexing”) or later at the final PCR amplification step of the library production (Nakamura et al 2015, Picelli et al 2014, Ramsköld et al 2012, Sasagawa et al 2013) (“late multiplexing”). While multiplexing saves money and time, it goes at the expense of sequencing depth and the ability to detect lowly expressed genes (discussed in 5.3.1).

The second option showed that pooling 30-100 cells approaches the information content and reproducibility of RNA-Seq from large amounts of input material (Marinov et al 2014). In the present study I opted to isolate and pool 30 to 100 cells per brain replicate in order to be able to i) obtain a resolution at almost single-cell level and ii) reduce for biological and technical noise.

RNA-Seq of a limited number of cells requires the successful combination of two independent techniques: i) the isolation of individual cells of interest from culture or tissue and ii) the successful amplification of RNA/cDNA for RNA-sequencing (RNA-seq).

3.3.3 Techniques to isolate cells from complex heterogeneous tissues

The isolation of newborn neurons from the SGZ at distinct maturation stages is a major challenge as the relative proportion of newborn to old neurons that is added in the dentate gyrus per day is only 0.03 – 0.06% (Kempermann et al 1997, Spalding et al 2013). In total, the adult murine dentate gyrus consists of 500.000 neurons (Abusaad et al 1999), of which approximately 200 cells (0.04%) are labelled by the replication deficient recombinant murine-moloney-leukemia virus (MMLV).

In the following section I will describe different single-cell isolation techniques and discuss their practicality for the isolation of rare cell types from intact tissue.

3.3.3.1 Micromanipulation

After dissociation of tissues into single cells, individual cells can be aspirated from a cell population using a glass micropipette. Due to the direct manual control under a microscope, their low cost and easy handling, micromanipulation has been widely used for the first single-cell analyses (Hashimshony et al 2012, Picelli et al 2013, Ramsköld et al 2012, Tang et al 2009). A disadvantage is its low throughput and high effort of manual handling. Moreover, the cells need to be dissociated from the intact tissue, which may change the gene expression. In addition, cells can be misidentified in suspension, as their original spatial context is lost (Macaulay & Voet 2014, Shapiro et al 2013).

3.3.3.2 FACS: flow- or fluorescence-activated cell sorting

Due to its high throughput (thousands of cells per second), FACS-based cell sorting is the most commonly used technique for the analysis of bulk cells. Cells in suspension are passed through a laser beam and sorted based on light scattering properties (such as size and granularity) or based on cell type specific markers (currently up to 17 individual markers can be used simultaneously) (Chattopadhyay & Roederer 2012). For single-cell analysis, single-cells can be directly sorted into 96- or 384-well plates within a few minutes.

A major disadvantage is that it requires a large number of cells in suspension as starting material, which makes it difficult to isolate low-abundance cell subpopulations. In addition, the spatial context of the cells within their original tissue is lost, and enzymatic cell dissociation can change gene expression (Spaethling & Eberwine 2013). Moreover, the rapid flow in the machine might damage the cells, making it difficult to isolate RNA from these cells (Macaulay & Voet 2014, Shapiro et al 2013).

FACS-based isolation is suited for the isolation of a substantial number of cells with a high target-to-noise ratio. Given that only 0.03 – 0.06% of neurons are added per day to the dentate gyrus in 2-month-old mice (Kempermann et al 1997), this number is too low for successful isolation via FACS.

3.3.3.3 Microfluidics

Microfluidic-based cell handling becomes more and more popular these days, as it allows semi-automated processing of multiple cells in parallel. It allows to perform cell culture, single-cell isolation, cDNA synthesis, amplification and NGS library preparation in a single device, thus, increasing throughput while at the same time reducing technical variations. A microfluidic peristaltic pump directs single cells in suspension to an isolated sorting chamber, where cell lysis and amplification reactions are performed in nanolitre-scale volumes (Lecault et al 2012, Sackmann et al 2014).

A major advantage of microfluidics is that samples are processed semi-automatically in parallel reaction chambers with lithographically defined nanoliter volumes. Hence, bias that is introduced during sample preparation (like variation in pipetting volumes, timing, mixing or reaction temperature) can be reduced, ultimately increasing accuracy and sensitivity as compared to tube-based approaches handling with microliter-reaction volumes (Wu et al 2014). By using a drop-based design in order to encapsulate the reagents, single-cell reaction volumes are now being reduced to even picolitre-reaction volumes (Rotem et al 2015). Moreover, by using barcoded oligonucleotides for first strand cDNA synthesis, early multiplexing is possible (Rotem et al 2015).

While at the beginning of single-cell RNA-Seq, most labs manually isolated single-cells (mostly by using glass micropipettes) (Hashimshony et al 2012, Picelli et al 2013, Ramsköld et al 2012, Tang et al 2009), recent methods more often use microfluidics to successfully describe the transcriptome of single-cells (Burns et al 2015, Darmanis et al 2015, Islam et al 2014, Kim et al 2015c, Klein et al 2015, Macosko et al 2015, Miyamoto et al 2015, Pollen et al 2014, Shalek et al 2014, Streets et al 2014, Suzuki et al 2015, Ting et al 2014, Zeisel et al 2015).

Still, this method requires to have the cells of interest already in suspension, which are then sorted into single-chambers. Thus, a preselection of cells of interest (e.g., by FACS or MACS) is a prerequisite to make use of the microfluidic device (Mahata et al 2014, Treutlein et al 2014, Tsang et al 2015).

3.3.3.4 LCM: Laser-capture microdissection

The laser capture microdissection technique offers the possibility to analyze specific cell populations within the context of their heterogeneous tissue microenvironment *in vivo*. Laser capture microdissection is a technique to isolate single cells or regions from fixed tissues or cryosections under direct microscopic visualization. After visualization of a specific cell population based on their morphology or by immunohistochemistry (Burbach et al 2004), using transgenic mouse models (Khodosevich et al 2007, Rossner et al 2006) or retroviral-based labelling, cells can be extracted from the intact tissue section by a UV laser, which are then collected or catapulted into the epitube cap (Erickson et al 2009, Espina et al 2006).

The major advantage of LCM is that it allows the isolation of morphologically distinguishable cell types from tissue without the need of prior dissociation and thus, preserves their 3D structure. Cells keep their positional information and are still connected to other cells, which is important for the proper transcriptome analysis, as cells in suspension lose their positional context and connections to other cells, which can vastly change gene expression (Shin et al 2014, Spaethling & Eberwine 2013). Thus, LCM is the only present technology that allows isolation of cells without perturbing the molecular state of the cells (Espina et al 2006). Moreover, the direct microscopic visualization allows direct control over the labelling procedure and position of the cell, thus, minimizing the error of isolating incorrectly labelled false-positive cells.

However, due to the close proximity to other cells within the network, even single microdissected cells might contain contaminations of neighbouring cells. In addition, cells might have been cut at the cryosectioning stage, leading to incomplete recovery of the total amount of RNA. Moreover, the RNA content residing in distal compartments (e.g., axons or dendrites) might not be captured by LCM (Batish et al 2012).

3.3.4 Different amplification technologies for a limited number of cells

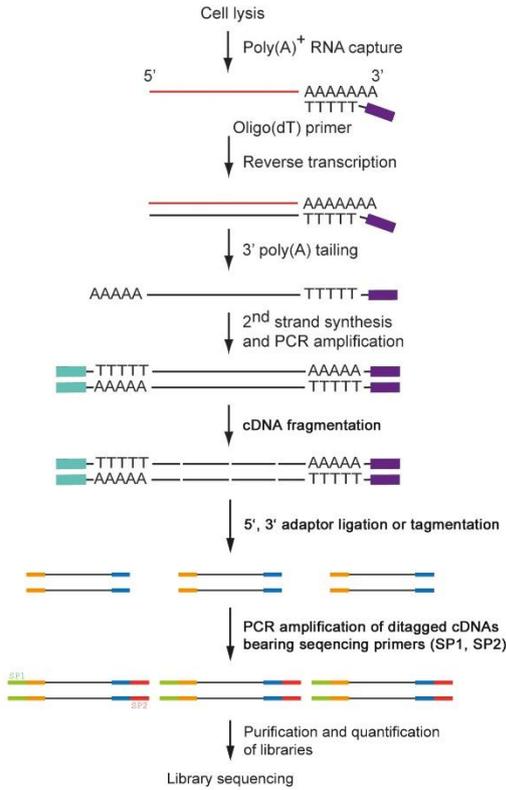
A typical mammalian cell contains approximately 10 pg of total RNA. Only 1-5% of this fraction are polyadenylated protein-coding mRNAs, corresponding to only about 0.1 pg mRNA or 50,000 to 300,000 mRNA molecules per cell (Islam et al 2012, Marinov et al 2014). However, RNA-Seq usually requires microgram amounts of total RNA, which corresponds to millions of mammalian cells. Thus, for 1 µg of mRNA, approximately 10 million cells or a representative 10 million-fold amplification is needed. The challenge here is to efficiently isolate such a small number of molecules, amplify them without introducing too much bias and accurately quantify the number of copies of each transcript.

RNA-seq of a limited number of cells requires a procedure that can amplify very limited starting material sufficiently for large-scale analysis, with high reproducibility, high fidelity and high coverage/sensitivity. The quantitative and qualitative performance of single-cell RNA-Seq has mainly been evaluated with respect to i) sensitivity and ii) accuracy/reproducibility. Sensitivity is typically measured by counting the number of genes detected within a cell or by counting the number of genes detected in single cells divided by the number of detected genes in non-amplified samples started with bulk cells (Sasagawa et al 2013). Precision/reproducibility is being measured by how well the results can be reproduced on replicate samples. The sensitivity and accuracy of single-cell RNA-Seq is limited by the efficiencies of each sample-processing step. The two main technical challenges in single-cell RNA-Seq are the efficiency of cDNA synthesis (which sets the limit of detection) and the amplification bias (which reduces quantitative accuracy) (Islam et al 2014). It is estimated that only 5-25% of all transcripts are converted to cDNA (Islam et al 2012). Moreover, PCR amplification is prone to introduce biases based on the nucleic acid composition, ultimately altering the relative abundance of transcripts in the sequencing library. Therefore, it is recommended to introduce as few sample-processing steps and PCR cycles as possible.

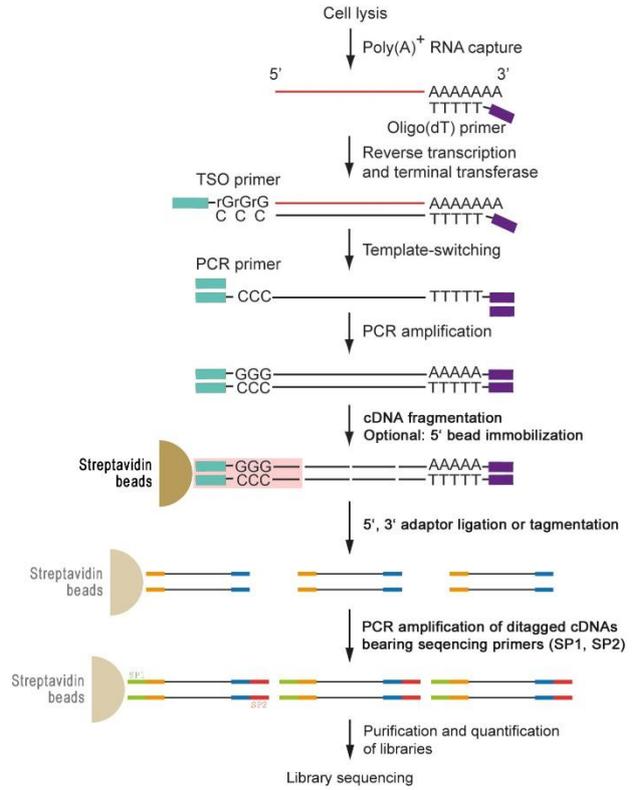
There are three major approaches that are used for cDNA amplification: i) methods involving exponential amplification by polymerase chain reaction (PCR), ii) methods involving linear amplification by T7 RNA polymerase and iii) methods involving rolling circle amplification (Hebenstreit 2012, Kolodziejczyk et al 2015, Liu et al 2014a, Macaulay & Voet 2014, Saliba et al 2014, Spaethling & Eberwine 2013) (Fig 3).

A PCR amplification

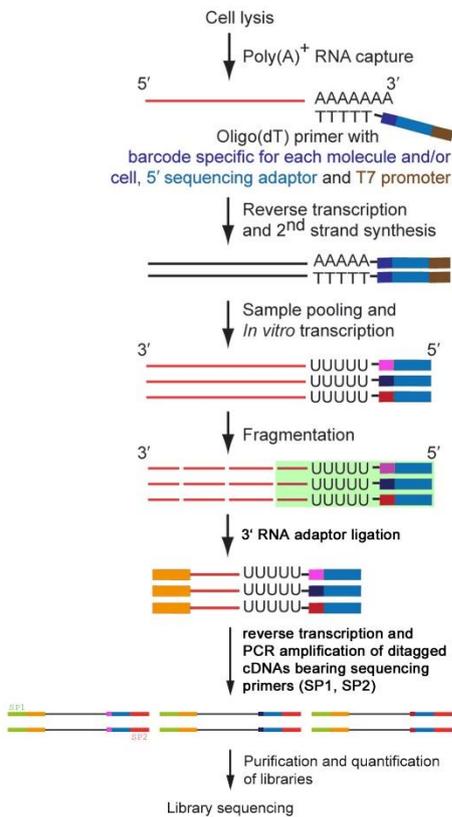
(i) Homopolymer tailing



(ii) Template-switching



B In vitro transcription



C Rolling circle amplification

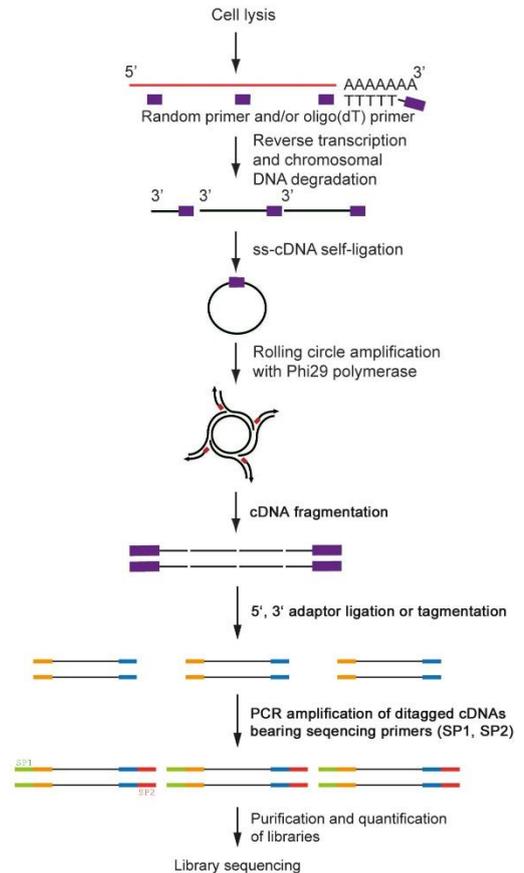


Fig 3. Different single-cell RNA-Seq amplification methods.

A) PCR amplification. Both homopolymer tailing (i) and template switching (ii) rely on PCR amplification based on the insertion of universal primer sequences at the 3' end (purple) and 5' end (turquoise). The 5' end adaptor can either be inserted by i) a polyA tailing reaction or ii) via a template-switching mechanism involving a template-switching oligonucleotide (TSO). After PCR amplification doublestranded cDNA is fragmented, 5' and 3' adaptors are inserted (yellow and blue) and sequencing primers are added by PCR (green and red). While the fragmentation of doublestranded cDNA leads to the loss of sequence orientation, it can be retained by capturing the 5' ends of amplified, biotinylated cDNAs using Streptavidin beads (ii), thus selecting only the 5'-most fragments for library construction (red highlighted box).

B) *In vitro* transcription (IVT). During cDNA synthesis a primer harbouring a T7 RNA polymerase enables the amplification of RNA via IVT. Typically, 3 rounds of IVT are needed for sufficient RNA amplification of single cells. If it is used with barcodes (dark blue, pink, red) enough RNA from different cells can be pooled that is sufficient for one round of IVT, however, allows the sequencing of the 3'-most ends only (green highlighted box). The strength of this protocol is that it is directional.

C) Alternatively, cDNA can be circularized and amplified by the Phi29 DNA polymerase, which allows amplification of full-length but non-directional libraries.

Red line, RNA; black line, cDNA; purple, universal primer sequence to label the 3' end of cDNA; turquoise, universal primer sequence to label the 5' end of cDNA; brown, T7 promoter; yellow/ blue, 5' and 3' sequencing adaptors; green/ red, sequencing primers; dark blue/pink/dark red, barcodes (modified from Saliba et al 2014).

3.3.4.1 PCR-based amplification

PCR based amplification procedures rely on the tagging of both 3' and 5' ends of cDNA with distinct universal primer sequences, which then can be targeted for PCR amplification (Fig 3A). The 3' end is tagged by the oligodT primer harbouring a universal primer sequence. The 5' end tag can either be inserted by i) a polyA tailing reaction (homopolymer tailing) or ii) via a template-switching mechanism involving a template-switching oligonucleotide (TSO) (see below).

The first single-cell RNA-Seq protocol relied on homopolymer tailing followed by PCR amplification (Tang et al 2010, Tang et al 2009). This protocol modified a quantitative single-cell microarray analysis protocol (Kurimoto et al 2006, Kurimoto et al 2007), thereby identifying 70% more genes than using microarray (Kurimoto et al 2006), depicting the higher sensitivity of RNA-Seq over microarray techniques (Tang et al 2009).

The main disadvantage of exponential PCR-based amplification is the accumulation of primer dimers and other nonspecific byproducts during PCR amplification, especially during later cycles of PCR (Kurimoto et al 2006). Excessive amounts of byproducts that were derived from the RT primer had to be removed by gel purification (Tang et al 2010, Tang et al 2009). Recently, alternative strategies have been developed to eliminate nonspecific products during PCR amplification, strategies that are implemented in the Quartz-Seq protocol (Sasagawa et al 2013). Quartz-Seq enables amplification in a single tube, and most importantly, without the need of gel purification (Sasagawa et al 2013).

A characteristic of all these PCR amplification methods is that they are 3'-biased due to the usage of oligo-dT primers. Even though the cDNA sizes could have been extended from 0.8 to 3 kb by extending the RT incubation time (5 min vs 30 min) (Kurimoto et al 2006, Tang et al 2009), still only 64% of genes were captured at full-length, while the 5' ends of the other transcripts have not been detected (Tang et al 2009).

The SC3-seq method (single-cell mRNA 3-prime end sequencing) takes advantage of this 3' bias and counts the very 3' ends only (Nakamura et al 2015). This amplification is based on the original Kurimoto protocol (Kurimoto et al 2006, Kurimoto et al 2007) with slight modifications. While the generation of full length cDNAs by reverse transcription is not an efficient process, and especially longer cDNAs are susceptible to PCR amplification bias, the amplification and sequencing of the very 3' ends would allow more precise quantification with a relatively small depth of sequence reads (Nakamura et al 2015). Therefore, the cDNA synthesis step was kept short in order to generate short cDNA stretches, and only the very 3' end read containing the barcode is being counted. Thus, one sequence tag corresponds to one transcript and does omit the need of normalization per gene length (Nakamura et al 2015). However, this is at the expense of the identification of new splice isoforms and the proper identification of new full-length transcripts. Moreover, due to high multiplexing and a low resulting sequencing depth (1 million reads per sample), only genes expressed with more than 10 copies can be successfully detected (Nakamura et al 2015).

A limitation of the described PCR-based amplification protocols is that the fragmentation of double-stranded cDNA leads to the loss of sequence orientation and thus, cannot be used for the identification of antisense transcripts.

3.3.4.1.1 *Template switching technologies*

Another PCR-based amplification approach is the template switching technology (Islam et al 2012, Islam et al 2011, Islam et al 2014, Picelli et al 2013, Picelli et al 2014, Ramsköld et al 2012) (Fig 3A ii). The template switching method that is also known as “switching mechanism at the 5′ end of the RNA transcript” (SMART) (Zhu et al 2001) relies on the intrinsic property of the moloney murine leukemia virus (MMLV) reverse transcriptase to add a few non-templated nucleotides (mainly cytosines) to the 3′ end of the first strand cDNA. After the priming of an oligonucleotide that contains (rG)₃ on its 3′ end (and a suitable 5′ sequence that will later serve as primer binding site for PCR amplification), the reverse transcriptase switches from the mRNA to the cDNA. This double-tagged cDNA can then be PCR-amplified. Importantly, the template switching will occur only if reverse transcription successfully reaches the 5′ end, thus ensuring that only full-length transcripts are being amplified. However, this is also the main disadvantage of the protocol, as any partially reverse transcribed mRNAs and many long mRNAs will fail to be detected.

The first SMART-based single-cell amplification protocol was named STRT (SMART-based single-cell tagged reverse transcription) (Islam et al 2012, Islam et al 2011). By using a template-switching oligonucleotide (TSO) that features a barcoding sequence and a common primer sequence, the cDNAs can be pooled after first strand cDNA synthesis and sequenced in a multiplex reaction, which reduces cost and time. In addition, a biotinylated STRT-primer was used for PCR amplification that enables the immobilization of the 5′ ends of amplified cDNAs on Streptavidin beads, and thus, ensures that only the 5′ ends will be sequenced and directionality will be kept. This was the first method that allowed strand-specific single cell RNA-Seq. Even though the template switching procedure in theory should retain the information of strandness, practically it was shown to be the least strand-specific approach compared to others using bulk cells (Levin et al 2010), most likely due to inefficient addition of non-templated C's to the cDNA end and reduced template switching efficiency when using long template switching oligos (TSO) (Islam et al 2014). Most likely this is also the reason why the authors of an updated STRT protocol do not claim to use a directional protocol and do not provide biological information on strand-specificity (Islam et al 2014). Therefore, strand specificity of the STRT protocol still needs to be proven. The strength of this method is that it can be used to pinpoint the exact location of the transcription start site

and thus, to analyze promoter usage in single cells, and may provide an alternative to nano-CAGE (cap analysis of gene expression) (Plessy et al 2010). However, it is not suitable for the analysis of alternative splicing.

The second SMART-based amplification protocol called “SMART-Seq” is a simplification of STRT (Ramsköld et al 2012). It does not use barcoding or 5′ end immobilization, but instead sequences full-length transcripts. Even though it is not strand-specific, due to its simple workflow and full-length coverage, this protocol is now available as a commercial Smart-Seq kit called “SMARTer Ultra Low RNA kit for Illumina sequencing” (Clontech). The main advantage of this protocol is the analysis of alternative transcript isoforms, even though it has to be kept in mind that some incomplete synthesized cDNAs are lost in the analysis.

Recently, an improved and updated method, called Smart-Seq2, was published (Picelli et al 2013, Picelli et al 2014). It allows multiplexing and introduced several optimizations (in reverse transcription, template-switching oligonucleotides and PCR pre-amplification) that resulted in overall improved sensitivity, accuracy and full-length coverage across transcripts compared to Smart-Seq. A limitation is still that it is not strand-specific.

3.3.4.2 IVT-based amplification

Linear isothermal RNA amplification was the first strategy that has been used to successfully amplify RNA (Van Gelder et al 1990). This actually promoted the birth of single cell analysis (Eberwine et al 1992). However, due to higher amplification efficiency of PCR based methods, PCR amplification is more prevalently used for single-cell transcriptome analyses.

The most commonly used mechanism for linear amplification is based on T7 RNA polymerase-mediated *in vitro* transcription (IVT), but also other bacteriophage promoter sequences (e.g., T3 or SP6) attached to the oligonucleotide primer can be used. The first methods attached the promoter sequence to the 3′ end of first strand cDNAs, resulting in RNA transcripts with antisense orientation (Eberwine et al 1992, Van Gelder et al 1990). Later, protocols have been described that attached the promoter sequence either to the 3′ or 5′ end of first-strand cDNAs, thus producing transcripts with antisense or sense orientation, respectively, and allowing for a proper distinction between sense and antisense transcripts after amplification (Che & Ginsberg 2003, Ginsberg 2005, Ginsberg & Che 2004).

As mentioned earlier, the IVT-based method is less frequently used mainly because the amplification procedure is less efficient. 20 cycles of PCR corresponds to approximately 1×10^6 fold amplification (Kurimoto et al 2006). The same amplification efficiency is achieved only after 2 rounds of *in vitro* transcription (IVT) (Eberwine et al 1992), which is much more tedious and time-consuming than PCR-based amplification. Each round of IVT can amplify the cDNAs only up to 2000-fold (Phillips & Eberwine 1996), which requires 3 rounds of amplification for single-cell amplification and corresponds to ~ 5 days work/cell vs. several hours for PCR amplification. In addition, the IVT-products are not full length (typically less than 1 kb) and with each round of amplification there is a 3' end biasing, as the amplification is based on oligo-dT primers. However, the advantage of the IVT strategy is its stringent specificity and ratio-fidelity while reducing accumulation of nonspecific products (Eberwine et al 1992). The linear mode of amplification is less efficient, but better preserves the relative abundances of mRNA transcripts (Morris et al 2011). Very importantly, the IVT-based amplification is strand-specific.

Up to now, the only single-cell amplification method that relies on IVT followed by RNA-Seq is called CEL-Seq (cell expression by linear amplification and sequencing) (Hashimshony et al 2012) (Fig 3B). The minimal requirement for IVT single round amplification is 400 pg total RNA. This limitation was overcome by barcoding and pooling samples before linearly amplifying mRNA in a single round of IVT (Hashimshony et al 2012). The first-strand oligo-dT primer is designed with a unique barcode, the 5' Illumina sequencing adaptor and a T7 promoter. After first and second strand cDNA synthesis, cDNA samples are pooled and consequently comprise sufficient template material for an IVT reaction. After 13 hours of IVT, amplified RNA is subjected to directional RNA library preparation (Illumina). The libraries are being sequenced with paired-end sequencing, where the first read recovers the barcode and the second read identifies the mRNA transcript.

It has to be considered that only the 3'-most ends of transcripts are being sequenced. This makes this protocol not applicable for the detection of new splice isoforms or novel transcripts. Moreover, caution must be taken when using such high multiplexing. Several cells need to be multiplexed in order to obtain sufficient amounts of RNA for one round of IVT. This means that each sample is being sequenced with very low sequencing depth, thus being hardly able to detect very low expressed genes, such as ncRNAs. While this method

can be used for the analysis of less complex genomes like *C. elegans* (Hashimshony et al 2012), caution should be taken when analyzing highly complex genomes from mouse or humans.

3.3.4.3 Rolling circle amplification

This amplification is based on the replicative polymerase from the *Bacillus subtilis* phage phi29 (ϕ 29) (Fig 3C). This polymerase is a highly processive polymerase with strong strand displacement activity that allows for highly efficient isothermal DNA amplification (Pan et al 2013). Since the Phi29 DNA polymerase requires relatively long DNA templates (> 3-4 kb) for efficient amplification, the full-length cDNA was circularized by CirLigase (Epicentre) or T4 DNA ligase before amplification. The major advantage of this amplification is that it can capture both the 5' end and 3' end sequences and thus, preserves full-length transcript coverage (Pan et al 2013). The only limitation was the incomplete representation of low abundance mRNAs, which suggests that low abundant transcripts are lost before or during cDNA circularization, most likely due to incomplete ligation or exonuclease action (Pan et al 2013). If random primers are used for cDNA synthesis, this amplification can be used for total transcript amplification in prokaryotes as well (Kang et al 2011). However then, genomic DNA needs to be removed before priming with random primers.

A summary of different single cell RNA-Seq methods is given in Table 2 (modified from Shapiro et al 2013).

Table 2

Single cell RNA-Seq methods

method name	single-cell amplification procedure	positional bias	strand-specific	reference
Tang et al., 2009	homopolymer tailing, PCR	weakly 3'	no	Tang et al., 2009; 2010
Quartz-Seq	homopolymer tailing, PCR	weakly 3'	no	Sasagawa et al., 2013
SC3-Seq	homopolymer tailing, PCR	3' only	no	Nakamura et al., 2015
Smart-Seq	template switching, PCR	nearly full-length	no	Ramsköld et al., 2012
Smart-Seq2	template switching, PCR	nearly full-length	no	Picelli et al., 2013; 2014
STRT-Seq	template switching, PCR, 5' bead immobilization	5' only	yes	Islam et al., 2011; 2012; 2014
CEL-Seq	<i>in vitro</i> transcription	3' only	yes	Hashimshony et al., 2012
PMA	rolling circle amplification	full-length	no	Pan et al., 2013
Uttenhaller et al., in prep	homopolymer tailing, PCR, <i>in vitro</i> transcription	weakly 3'	yes	Uttenhaller et al., in prep.

SC3-Seq, single-cell mRNA 3' end sequencing; Smart-Seq, switching mechanism at the 5' end of the RNA transcript sequencing;

STRT-Seq, SMART-based single-cell tagged reverse transcription; CEL-Seq, cell expression by linear amplification and sequencing; PMA, Phi29-mRNA amplification

3.3.5 Comparison of different amplification methods

Compared to homopolymer tailing, SMART-Seq (Ramsköld et al 2012) exhibits approximately one third lower sensitivity (7869 vs. 11784 detected genes >0.1 RPKM in single human embryonic stem cells at a similar sequencing depth) (Yan et al 2013), which may be explained by an imperfect efficiency of the reverse transcriptase to add 3' cytosines, especially to longer transcripts. Thus, any partially reverse transcribed mRNAs will fail to be amplified. This ineffectiveness was also seen in a comparison between CEL-Seq (Hashimshony et al 2012) and STRT-Seq (Islam et al 2012), resulting in lower sensitivity of STRT-Seq in mouse ES cells (5500 vs. 2200 detected genes >10 FPKM, respectively) (Hashimshony et al 2012). However, it has to be mentioned that the overall detected number of genes in MEFs reached similar levels using STRT-Seq as CEL-Seq (Hashimshony et al 2012).

In a direct comparison between Quartz-Seq (Sasagawa et al 2013), CEL-Seq (Hashimshony et al 2012) and SMART-Seq (Ramsköld et al 2012) in single mouse ES cells, Quartz-Seq displayed highest sensitivity, followed by CEL-Seq and SMART-Seq. While Quartz-Seq was able to reliably detect more than 6000 genes per ES cell (fpm >1 or tpm >1), SMART-Seq and CEL-Seq detected only 2906 and 4070 genes, respectively (Sasagawa et al 2013). In a comparison to a non-amplified sample starting with 1 µg of RNA, the sensitivity to detect genes in single cells compared to bulk cells was again higher for Quartz-Seq than for SMART-Seq (81% vs. 63%, respectively) (Sasagawa et al 2013).

These comparisons clearly show that homopolymer-tailing based PCR amplification is more sensitive and accurate than template-switching methods or linear-based amplification.

Still, the template switching amplification is widely used because it is commercially available (SMARTer Ultra Low RNA kit, Clontech) and because of the very good transcript coverage, which makes it more suitable for isoform expression analysis. Moreover, the optimizations by Smart-Seq2 improved both sensitivity, accuracy and full-length coverage, ultimately being able to detect as many genes as Quartz-Seq (Picelli et al 2013). Still, theoretically, the best amplification procedure that allows full-transcript coverage is the rolling circle amplification (Pan et al 2013). However, as for today, this amplification procedure has not been used in any other experimental settings and its applicability has to be tested further.

3.3.5.1 Comparison of strand-specific amplification protocols

Up to date, the only strand-specific single-cell amplification protocols are STRT-Seq (Islam et al 2012, Islam et al 2011) and CEL-Seq (Hashimshony et al 2012). Even though they are strand-specific protocols, both methods describe only a very small fraction of antisense reads overlapping exonic regions (less than 3% vs. 1% respectively (Hashimshony et al 2012, Islam et al 2011)). However, using Direct RNA-sequencing it was shown that approximately 8% of reads overlap an annotated transcript in antisense, accounting for more than 30% of annotated transcripts in humans having antisense transcription (Ozsolak et al 2010), which is also in agreement with numbers obtained from cDNA libraries (Chen et al 2004). This discrepancy can most likely be explained by the fact that the STRT and CEL-Seq approach are not able to detect very low expressed transcripts, which antisense transcripts belong to, since they are more than 10-fold lower in abundance than sense transcripts (Derrien et al 2012b, Ozsolak et al 2010).

Indeed, high multiplexing goes at the expense of sequencing depth and thus, limits the detectability of lowly expressed genes. It was shown that using CEL-Seq the detection rate of genes with 5 copies per cell is only 50 % (Hashimshony et al 2012). Using the barcoding STRT protocol in single cells only highly and medium-expressed genes of more than ten copies per cell will be reliably detected (Islam et al 2012). As shown for ES cells, the median copy number is just 2 copies/ cell (Islam et al 2011), which means that many transcripts will be missed by using the multiplexed CEL-Seq or STRT method. The low sensitivity of these methods will be overcome by higher sequencing depth.

Moreover, due to early barcoding, only the 5'- and 3'-most ends can be sequenced that contain the barcode, which allows the sequencing of only 1 read per transcript and does not provide information on the entire transcript sequence or alternative splicing.

Even though an optimization of the STRT-method should in principle also retain the information of strandness (Islam et al 2014), the authors do not claim to use a directional protocol neither do they provide information on strandness (Islam et al 2014). Before, SMART-Seq starting with high amount of input material (Cloonan et al 2008) was shown to be the least strand-specific protocol compared to others (Levin et al 2010), suggesting that

also in single-cell procedures the template-switching approach is not efficient and thus, sequence orientation can not be guaranteed.

In the present study a new directional single-cell RNA-Seq protocol will be described that allows for the first time a comprehensive description of both coding and non-coding transcripts in a directional manner from a limited number of cells (< 100 cells) isolated *in vivo*.

3.4 *Objectives of this study*

The overall aim of this project was the transcriptomic characterization of different maturation stages of *in vivo* adult hippocampal neurogenesis, in order to gain deeper insight into the molecular networks controlling distinct phases of adult neural stem cell differentiation.

While transcriptional control of early stages of adult neurogenesis, i.e., the transition from quiescent to activated neural stem cells and intermediate progenitor cells, has recently been extensively described in the adult SGZ (Shin et al 2015) and adult SVZ niche (Llorens-Bobadilla et al 2015, Luo et al 2015), later stages of immature neuron development, i.e., neuronal morphogenesis, synaptogenesis and circuit formation, have escaped analysis, due to the technical challenge to identify and isolate neurons of a specific maturation stage in the *in vivo* context. Therefore, one specific aim of this project was the establishment of a methodology for labelling and isolation of stem cell-derived neurons from the adult hippocampus at distinct developmental stages.

Besides the isolation of newborn neurons from the intact *in vivo* niche, one major challenge of this analysis was the limited number of adult-born neurons per dentate gyrus, which forced the need of pre-amplification prior to RNA-seq. Therefore, another specific aim of this project was the development of a pre-amplification strategy that allows the directional amplification of RNA, thus being able to discriminate between sense and antisense expression. Such information is highly important for the analysis of non-coding RNA transcripts, especially for those transcribed on the opposite strand of a coding gene (i.e., antisense transcripts). Since the involvement of long non-coding RNAs (lncRNAs) in the regulation of genetic programs is becoming more and more prevalent, the present study further aimed to provide a comprehensive description of the expression of already annotated lncRNAs, but also the discovery of novel, not yet annotated lncRNAs. Here, an emphasis was taken on the description of lncRNAs that are expressed at specific developmental stages of adult hippocampal neurogenesis. Moreover, a closer look was taken into the genomic context of stage-specific lncRNAs in order to be able to predict regulatory modules and pathways from co-expressed lncRNA-mRNA pairs.

4 RESULTS

To elucidate the molecular profiles of newborn neurons during early, intermediate and late maturation stages of adult hippocampal neurogenesis *in vivo*, a retrovirus birthdating strategy was used to label a cohort of neurons for isolation, followed by transcriptome analysis by RNA-Seq. While available transgenic mouse lines (e.g., Nestin, Sox2, Glast, DCX) or markers do not provide sufficient temporal resolution of different maturation stages, the labeling by murine Moloney leukemia (MML) retrovirus is highly temporal specific, as the MML only integrates into dividing cells, i.e., activated neural stem cells and highly dividing intermediate progenitor cells in the adult dentate gyrus. Thus, the corresponding labelled cells do represent cells of the same developmental stage.

Since only approximately 200 cells were retrovirally labeled per mouse, two major challenges had to be overcome, i.e., i) the isolation of such rare cells and ii) the amplification of RNA for RNA-Seq.

4.1 Labelling and isolation of newborn neurons

In a first attempt, I tried to isolate retrovirally labelled neurons via FACSorting. Isolation of DCX-positive neurons derived from the DCX dsRed transgenic mouse (Couillard-Despres et al 2006) was possible from the hippocampus (Fig 4A) and all other neurogenic regions harbouring DCX-positive cells (SVZ, RMS, OB) (Fig 4A'). However, the isolation of GFP-retrovirally transduced cells was not successful in my hands (Fig 4B), because the GFP-transduced cells were not bright enough and too low in number to set the gate for efficient and selective sorting of positive cells. Although the GFP positive cells were likely identified (Fig 4B, red rectangle), also the control mouse contained highly autofluorescent cells, ultimately yielding many false-negative and only a few true-positive GFP-expressing cells.

As an alternative, I tried the isolation of newborn neurons via laser capture microdissection (LCM). The main advantage of LCM is that the direct microscopic visualization allows direct control over the labelling procedure and position of the cells. In a first trial, a retrovirus encoding GFP was used to label newborn neurons. Thereafter, 12 μ m cryosections were made, followed by visual inspection of the transduced cells under the fluorescent microscope. As noticed before by other groups as well (Khodosevich et al 2007, Rossner et

al 2006), soluble fluorescent proteins such as GFP and RFP tend to leach out and to diffuse to neighbouring structures in unfixed tissue sections. In order to evaluate the usability of different fluorescent proteins, I inspected several tissue sections of Nestin-transgenic mouse lines expressing different fluorescent proteins. I also noticed that cytoplasmic GFP and mCherry leached out in 12 μm cryosections (Fig 4C and 4C'). Previous work recommended to circumvent the diffusion of fluorescent proteins by i) fixing brains by PFA and thereafter use a de-crosslinking procedure for RNA isolation (Khodosevich et al 2007) or ii) by using fluorescent markers targeted to subcellular structures such as the nucleus (Rossner et al 2006). Since the de-crosslinking procedure likely interferes with the RNA integrity, I favoured the usage of fluorescent proteins targeted to subcellular structures, but also here differences with regard to the subcellular labeling were found in cryofixed tissue. While fluorescent proteins targeted to the histone H2B also leached out (Fig 4C''), acceptable results were obtained using CFP targeted to the nucleus (Fig 4C'''). However also here it was not possible to explicitly identify the cell body and ultimately to isolate a pure and defined cell population as judged by the expression of marker genes by qRT-PCR (data not shown).

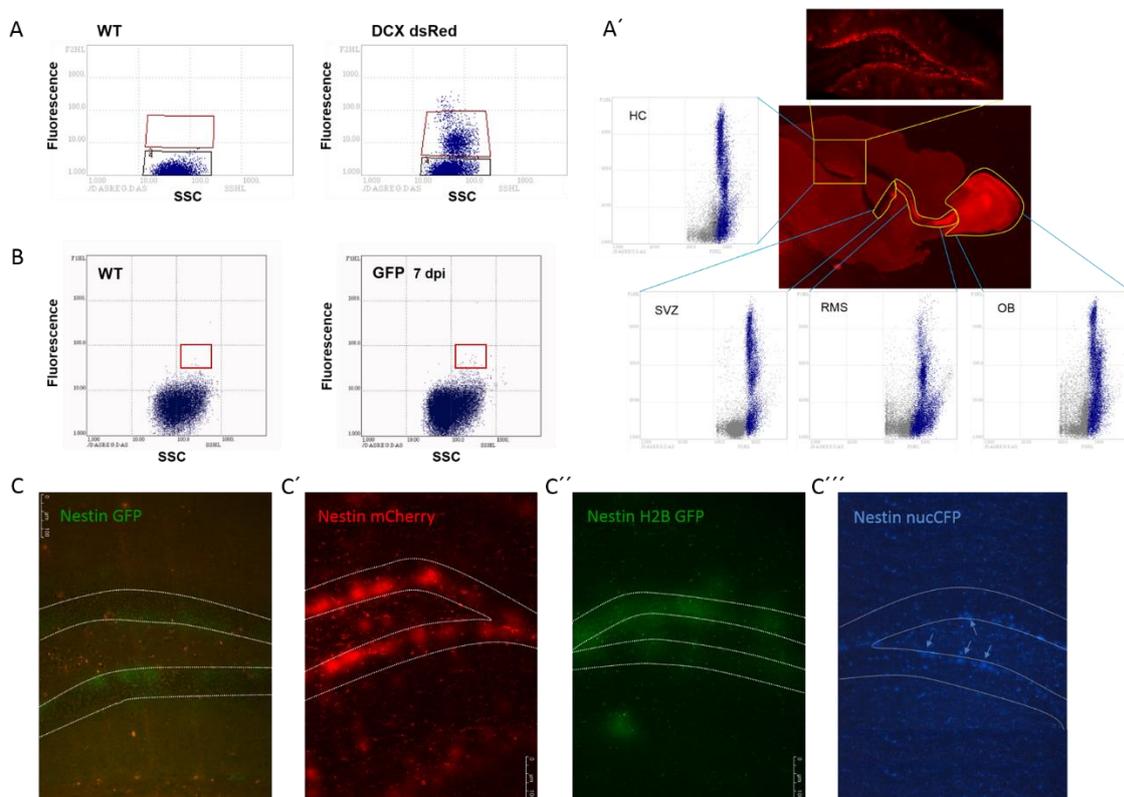


Fig 4. Approaches to label and isolate newborn neurons from the adult dentate gyrus with limited success.

A-A') FACSsorting of DCX⁺ cells derived from the DCX-dsRed transgenic mouse. **A)** Compared to a WT mouse, DCX⁺ cells can be accurately identified and sorted from the adult hippocampus (red rectangle). **A')** Using the DCX-dsRed transgenic mouse, it is possible to isolate DCX⁺ cells from all neurogenic niches, i.e., hippocampus (HC), subventricular zone (SVZ), rostral migratory stream (RMS) and olfactory bulb (OB). **B)** Compared to a WT mouse, GFP transduced cells isolated after 7 dpi can not be accurately identified (red rectangle) due to low numbers and weak signal intensity. SSC, side scatter. **C-C''')** Nestin transgenic mouse lines expressing different fluorescent proteins. Soluble cytoplasmic fluorescent proteins like GFP (C) and mCherry (C') diffused to neighbouring tissue, but also GFP targeted to the histone H2B (C'') leached out. More specific signal was obtained using CFP targeted to the nucleus (C''') (blue arrows), but still the isolation was not specific.

Next, I tried a retrovirus encoding for mitochondria-targeted dsRed for labeling (termed mito dsRed - a fusion protein consisting of dsRed and a mitochondrial target sequence that directs the fusion protein to the host cell's mitochondria). The successful labeling of mitochondria throughout the cell, allowed us to visualize the morphology of the cells in cryofixed tissue sections without facing the problem of diffusion or leakage of the signal (Fig 5B-B''). I isolated newborn neurons 6 days post injection (dpi), 12 dpi and 28 dpi in order to investigate early, intermediate and late maturation stages of adult hippocampal neurogenesis. While 6 dpi neurons just start making primary dendrites (Fig 5B), 12 dpi neurons already extend their immature and lowly branched dendrites towards the molecular layer (Fig 5B'), and 28 dpi neurons finally exhibit a highly branched dendritic tree (Fig 5B''). The mitodsRed labeling strategy finally allowed the visualization and isolation of newborn neurons by LCM. It has to be noted that only the cell soma could have been isolated, while RNAs residing in dendritic or axonal compartments could not be isolated by LCM. Since only approximately 200 cells were labeled per mouse and some cells were compromised in their integrity at the cryo-sectioning stage, only 30 to 100 cells could be collected per timepoint and mouse brain. These cells were captured in the epitube cap, and immediately lysed and processed for further amplification (see 4.2).

To determine the purity and specificity of the LCM-assisted isolation, qRT-PCR analysis of LCM isolated cells that were isolated at 7 dpi or 42 dpi was performed. The expression of characteristic marker genes correlated with the developmental stage. 7 dpi isolates showed high expression of the immature DG neuron marker doublecortin (DCX), whereas 42 dpi

isolates strongly expressed the mature DG neuron marker Calbindin. Both 7 dpi and 42 dpi samples showed only low expression of the stem cell and astrocyte marker Sox2 (Fig 5C and 5C'). These data demonstrate that LCM-assisted dissection is suitable for the specific isolation of virally-birthdated neurons. Therefore, the isolation of retrovirally labelled neurons by LCM became the method of choice for the transcriptome profiling approach (Fig 5A).

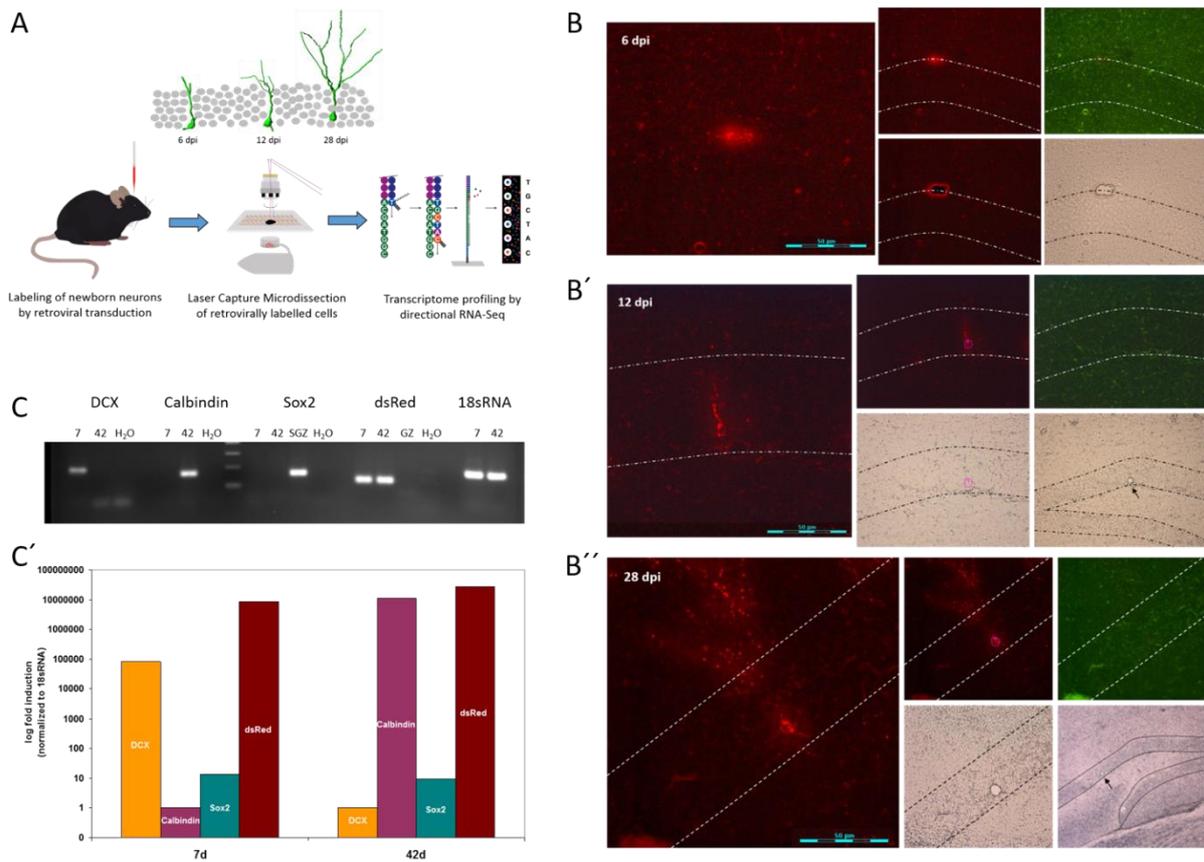


Fig 5. Successful isolation of retrovirally labeled newborn neurons from the adult mouse brain by LCM.

A) Outline of the workflow. Newborn neurons were labeled using a retrovirus encoding mitodsRed and isolated at different timepoints post injection (dpi) by LCM. After RNA amplification, transcriptome profiling was conducted by RNA-Seq. **B-B'')** LCM of single newborn neurons from the adult dentate gyrus at 6 dpi (B), 12 dpi (B') and 28 dpi (B''). The green channel provides evidence that the mitodsRed signal is specific only in the red channel and not caused by autofluorescence. After LCM the cells are accurately cut from the tissue (arrows in the brightfield picture). Note that only the cell soma could be isolated but not the dendrites. In total, 30 to 100 cells were collected per timepoint and mouse brain. **C-C')** Newborn neurons were specifically isolated by LCM and the expression of marker genes (DCX vs. Calbindin) correlated to the developmental stage (7 vs. 42 days). C) RT-PCR shows expression of the immature marker DCX in 7 day old immature neurons but not in mature 42 days neurons, whereas the immature marker Calbindin is only expressed in 42 days neurons and not in 7 days

immature neurons. Note that both immature and mature cell populations lack Sox2 expression, a marker for astrocytic and stem cells (the subgranular zone (SGZ) serves as a positive control). Since both cell populations were identified by the expression of mCherry, both cell populations show high expression of dsRed, whereas other regions of the granular zone (GZ) lack dsRed expression. 18sRNA serves as a loading control. C) For quantification, C_t values were normalized to 18sRNA and then the fold induction was calculated over Calbindin or DCX expression for 7 days or 42 days neurons respectively.

4.2 Directional amplification protocol

The second challenge that had to be overcome was the amplification of RNA for such a limited cell number. For RNA-Seq typically microgram amounts of total RNA are required, which corresponds to millions of mammalian cells. Given that one representative individual mammalian cell contains around 10 pg of total RNA and ~ 0.1 pg of mRNA, efficient amplification is essential (Tang et al 2011). Several single-cell RNA-Seq amplification protocols have been established so far (Table 2 in 3.3.4).

In addition to these single-cell amplification methods, I developed in collaboration with my colleagues a library amplification protocol that allows to keep directionality and thus, to allow the characterization of antisense transcripts even after amplification (Fig 6A). Two other directional single-cell RNA-Seq protocols have been described so far (Hashimshony et al 2012, Islam et al 2012, Islam et al 2011), however, this is the first protocol that allows sequencing over the entire gene body, and not only the very 3' end (Hashimshony et al 2012) or 5' end (Islam et al 2012, Islam et al 2011).

In the present study, a modification of the Kurimoto (Kurimoto et al 2007) and Tang protocol (Tang et al 2009, 2010) was used, with modifications to make it suitable for strand-specific RNA-Seq and cells isolated from cryofixed brain tissue.

As a first step, PCR-based amplification of mRNA was performed by using a homopolymer-tailing approach with distinct universal primer sequences (UP1, UP2) at the 5' and 3' end of cDNA, similar to as described before (Tang et al 2010, Tang et al 2009) with some modifications (see below). Since the fragmentation of the double-stranded cDNA leads to the loss of sequence orientation (Tang et al 2010, Tang et al 2009), an *in vitro* transcription (IVT) step was introduced after PCR amplification as a second step. Unlike the previous

addition of the T7 promoter to the 3' end (Kurimoto et al 2007), here the T7 promoter was added to the 5' end of cDNA, which allowed to generate amplified sense RNA transcripts by T7 RNA polymerase in an IVT step. Finally, these amplified RNAs were fragmented and subjected to directional library preparation either following the guidelines of the Illumina pre-release directional mRNA-Seq library preparation kit v1.0 (2009) or the ScriptSeq v2 library preparation kit by Epicentre (2011) (Fig 6A).

Several controls and optimizations were made in order to test the usability of this directional protocol. Firstly, I optimized the lysis conditions for cells isolated from intact tissue sections and evaluated the performance by qRT-PCR for selected genes after 20x PCR amplification. While for dissociated single cells 90 seconds lysis time was sufficient (Kurimoto et al 2007, Tang et al 2009), it was not sufficient to lyse cells from intact tissue. Therefore, the lysis temperature was increased from 70°C to 75 °C, and different cell lysis times for different marker genes were tested. Best results were obtained after 10 minutes lysis time for all markers tested (Fig 6B).

Secondly, I confirmed that mRNAs were specifically primed by the UP1 primer and that genomic DNA was not amplified (-RT control) (Fig 6C). The no-RT control contained all reaction components except for the reverse transcriptase. Reverse transcription, i.e., the production of cDNA should not occur in this control, hence no PCR amplification should be seen in the qRT-PCR, otherwise PCR products are most likely derived from contaminating DNA. Even though genomic DNA has not been eliminated by DNase I before first strand cDNA synthesis, the specific priming of polyadenylated transcripts followed by 20x PCR amplification yielded much more cDNA compared to genomic DNA, that the genomic DNA in the samples was negligible (Fig 6C).

As another control, I proved that false-priming did not occur (-primer control) (Fig 6C). The no-primer control contained all reaction components except for the UP1 primer. Self-priming/false-priming leading to reverse transcription in the absence of any primer poses a significant technical hurdle, especially in strand-specific protocols (Beiter et al 2007, Haddad et al 2007, Tzadok et al 2013). Unspecific priming depends on both the RT used as well as the temperature of the cDNA synthesis (Moison et al 2011). To increase RT reaction specificity, it was recommended to perform RT at 50°C minimum (Haddad et al 2007, Moison et al 2011, Tzadok et al 2013). In addition, the PCR amplification based on primer tags was

shown to largely prevent the amplification of unspecific products (Lim et al 2013). In the current experimental settings, reverse transcription was performed at 50°C and PCR amplification based on primer tags (UP1, UP2), which ultimately prevented self-priming/false priming (Fig 6C). Therefore, it can be concluded that amplified products were largely specifically primed and not caused by self-priming.

After the addition of the T7 promoter, byproducts shorter than 300 bp resulting from primer dimers were removed by agarose gel electrophoresis (Fig 6D-D'). The elimination of primer dimer products by gel purification is also a crucial step, as these dimer products result in high numbers of polyA/polyT sequencing reads and thus, dramatically influence the percentage of mappability of sequenced reads.

The PCR amplified cDNA products ranged in size between 300 to 2000 kb with a peak at approximately 500 bp (Fig 6E), which is in the range with previous amplification protocols (Tang et al 2010, Tang et al 2009).

Next, the PCR amplified cDNA products were subjected to *in vitro* transcription (IVT) by T7 RNA polymerase, thereby generating amplified sense RNA transcripts. As shown before by Kurimoto and colleagues, the initial ditagged PCR amplification followed by IVT keeps the information of directionality (Kurimoto et al 2006, Kurimoto et al 2007). Likewise, I could confirm that the size distribution of amplified transcripts did not change after IVT (Fig 6E and 6E') and that the IVT amplification was linear and the overall transcript abundance preserved after IVT (Fig 6E''). 4 hours of IVT were sufficient to obtain enough RNA material for the preparation of RNA-Seq libraries and for test qRT-PCRs. The PCR-IVT amplified products ranged in size between 100 and 3000 nt with a peak at 500 nt (Fig 6F).

After quality controls by the Bioanalyzer to control for RNA size distribution (Fig 6F), and test qRT-PCRs in order to confirm the expression of known marker genes, the amplified RNA was fragmented and subjected to the preparation of directional RNA-Seq libraries following the guidelines of i) the ScriptSeq v2 library preparation kit by Epicentre (2011) (Fig 6G-G') or ii) the Illumina pre-release directional mRNA-Seq library preparation kit v1.0 (2009) (Fig 6H-H').

Upon fragmentation the fragments ranged in size up to 500 nt with a peak below 200 nt (Fig 6G and 6H). The addition of 5' and 3' adaptors in a directional manner was accomplished

either by applying a terminal tagging reaction (Epicentre kit) (Fig 6A and 6G') or by serial ligation of adaptors (Illumina kit) (Fig 3H'). Thereafter, cDNAs were subjected to a final PCR amplification, thereby adding Illumina sequencing primers to the libraries. The Epicentre ready-to-sequence libraries ranged in size between 150 and 1000 bp with a peak at approximately 300 bp (Fig 6G'). The Illumina ready-to-sequence libraries ranged in size between 100 and 300 bp with a peak at approximately 150 bp (Fig 6H').

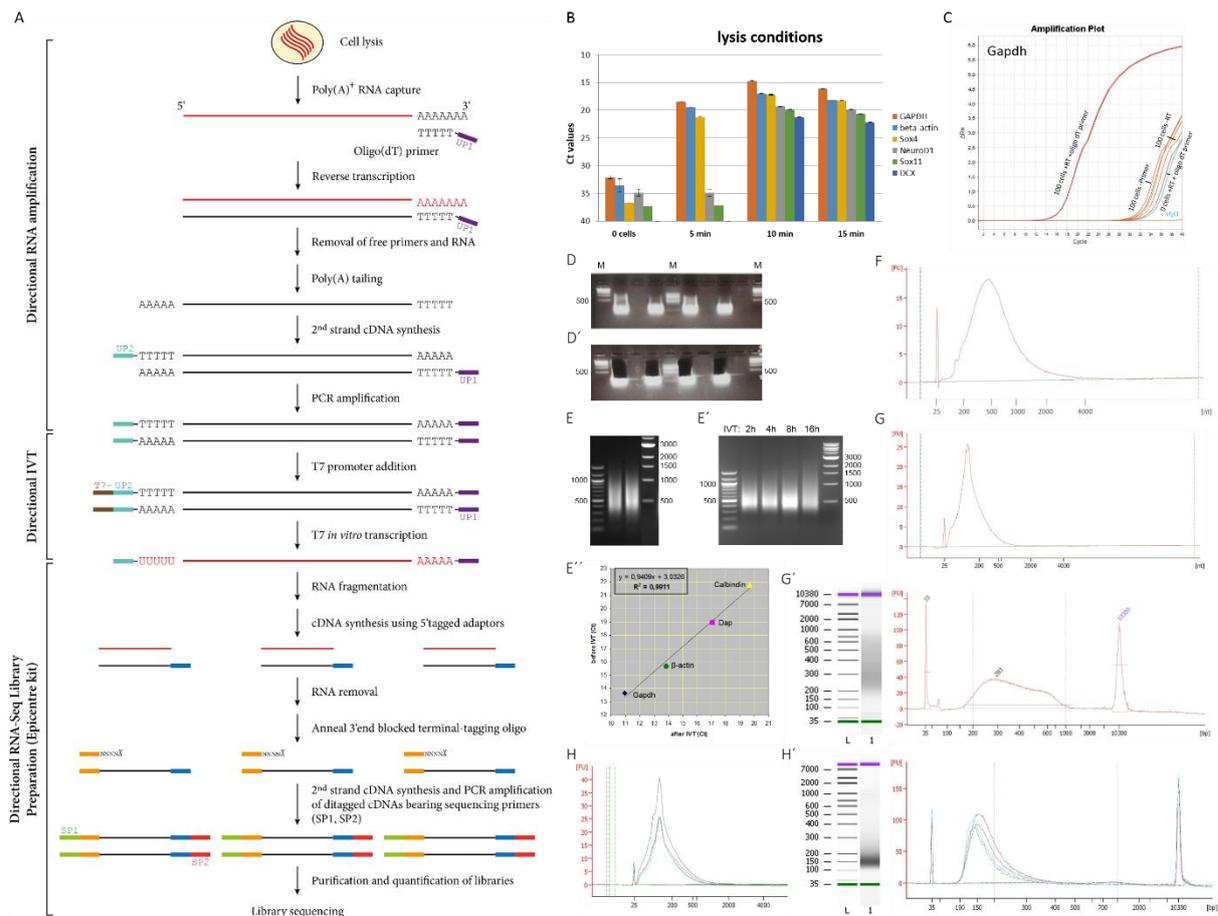


Fig 6. Validation of the directional amplification protocol.

A) Directional amplification protocol. The protocol can be divided into 3 main steps i) directional RNA amplification based on homopolymer tailing followed by PCR amplification, ii) directional IVT generating sense RNA transcripts and iii) directional library preparation following the ScriptSeq v2 RNA-Seq library preparation kit by Epicentre. PolyA⁺ RNA (red line) is captured using an oligo-dT primer harbouring a universal primer sequence (UP1) (purple). The 5' end of cDNA (black line) is marked by a homopolymer tailing reaction, which can be targeted by a second universal primer sequence (UP2) (turquoise). Based on these distinct UP1 and UP2 primers, doublestranded cDNAs are amplified by a PCR reaction (20x). In the second step, a T7 promoter sequence (brown) is added to the 5' end of cDNAs, which allows to generate amplified sense RNA by T7 RNA polymerase in an IVT step (4h). Finally, these amplified RNAs are fragmented and subjected to directional

library preparation. 5' and 3' adaptors (yellow/blue) are added by applying a ligation-free, terminal tagging reaction. Libraries are amplified using sequencing primers (green/red).

B) Optimization of lysis conditions. Lysis time has to be assessed for any cell type extracted from tissue. 100 cells were isolated from the DG and cells were lysed for 5, 10 or 15 minutes at 75°C. After 20 cycles of PCR amplification, expression of housekeeping genes and neuronal marker genes were tested by qPCR. Best results were obtained for 10 minutes of lysis. As a negative control, amplification was performed without any cells in the reaction (0 cells). **C)** Specific priming and amplification of polyadenylated transcripts. After 20x PCR amplification, Gapdh was significantly higher expressed than in the control reactions (a) without RT, (b) without oligo dT primer and (c) no cells in the initial cDNA synthesis step.

D-D') Elimination of primer dimer products by agarose gel purification. PCR amplified products were loaded on a 2% agarose gel (D). After 10 minutes of electrophoresis at 100 V the fraction above 300 bp was cut from the gel and purified (D'). M, marker.

E-E''') Preservation of size distribution and transcript abundance through the IVT reaction. Size distribution of amplified transcripts before (E) and after *in vitro* transcription (IVT) (E'). Amplified cDNA products ranged in size between 0,3 – 2 kb. This size distribution was preserved after varying timepoints of *in vitro* transcription (E'). E''') RNA was reverse-transcribed after IVT. qRT-PCR was performed on three endogeneous genes (Gapdh, β -actin and Calbindin) and one spike RNA (Dap). Ct values are plotted before IVT (y-axis) and after IVT (x-axis); the regression coefficient (R^2) is indicated in the box.

F-H') Directional library preparation for RNA-Seq using the Epicentre kit (G-G') or the Illumina kit (H-H'). An aliquot of the PCR-IVT amplified RNA was loaded on a Bioanalyzer RNA Pico chip before (F) and after fragmentation (G, H). 100 to 3000 nt long RNA transcripts (F) were fragmented to up to 500 nt (G, H). The ready-to-sequence libraries were loaded on a Bioanalyzer High Sensitivity DNA chip (G', H'). L, ladder.

4.3 *Quality control and technical performance*

4.3.1 **Quality of raw reads**

After sequencing, the quality of reads was determined for each library sample. Quality assessments are shown for one example (Fig 7). In a first quality measurement the Phred quality score was assessed. The Phred quality score is used to characterize the quality of DNA sequences by assessing the base calling accuracy. It is logarithmically linked to error probabilities. A Phred quality score of 10 means that 90% of base calls are correct (1 in 10 incorrect base calls), a score of 20 means that 99% are correct (1 in 100 incorrect base calls), a score of 30 means that 99.9% are correct (1 in 1000 incorrect base calls), and a score of 40 means that 99.99% are correct (1 in 10000 incorrect base calls). Phred scores below 20 are

considered unacceptable, while scores between 20 and 28 are considered acceptable, and scores above 28 refer to good base calling accuracy. For all samples tested the base calling was good over the entire sequence read (76 bp) with Phred scores being higher than 28 (green area) (Fig 7A). Over all sequences, the Phred score ranged between 29 and 40 with an average quality of 39 (Fig 7B), indicating that all sequences have a good quality and base call accuracy.

Besides the Phred quality score, the GC content and the distribution of all nucleotides was assessed. As expected, all 4 nucleotides (A, T, G, C) were distributed equally over the entire sequence length (Fig 7C). Moreover, the GC distribution over all sequences was very similar to the theoretical distribution (blue) with a peak at approximately 50% (Fig 7D). In addition, there was no overrepresentation of unspecific, repetitive sequences measured as k-mer sequences. Overall, these quality controls indicated that all samples were good quality reads that could be used for mapping and alignment.

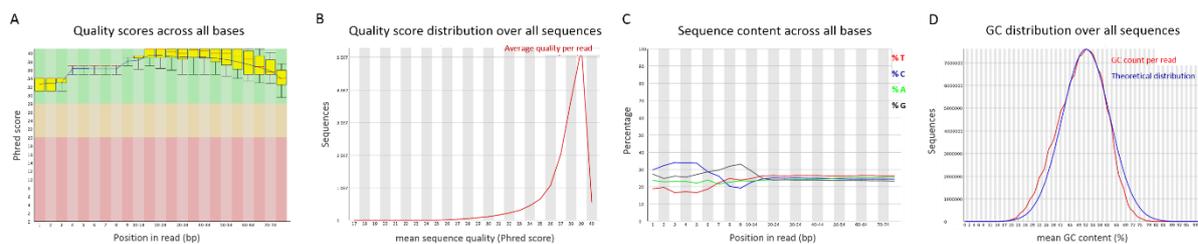


Fig 7. Quality control of sequencing reads.

A-B) Phred quality score per base (A) and per sequence (B). Base calling accuracy is good over the entire sequence length (green area) (A) and over all sequences with an average phred score of 39 (B).

C-D) Distribution of nucleotides per base (C) and per sequence (D). All 4 nucleotides are evenly distributed over the entire sequence length (C), and the GC content over all sequences (red) is similar to the theoretical distribution (blue) with a mean of 50 % (D).

4.3.2 Mappability of reads

Sequencing of samples was achieved in two runs. First samples were prepared in 2010 using the Illumina pre-release directional mRNA-Seq library preparation kit v1.0 (2009), a ligation-based approach. Due to optimizations of the amplification protocol, new libraries were prepared in 2012, however, this time using the Epicentre ScriptSeq v2 RNA-Seq library preparation kit (2011), which replaced the no-longer available Illumina pre-release

directional mRNA-Seq library preparation kit. While the first samples were sequenced on a Illumina GAII platform, generating approximately 35 million reads per sample, the second run samples were sequenced on a Illumina HiSeq2000 platform, generating approximately 200 million reads per sample. The reads were mapped to the mm10 mouse genome. The statistics of mapped reads per sample is summarized in Table 3. Note that the mappability of samples generated with the Epicentre kit was higher than compared to the Illumina kit (98.57 % (± 1.50 %) vs. 69.59 % (± 2.19 %)). Note that for all samples more than 90% of reads mapped uniquely to genomic regions.

Table 3

Mappability of reads

	RME180_6dpi3	RME181_6dpi5	RME182_6dpi6	RME187_12dpi2	RMN013_12dpi3	RMN014_12dpi5*	RME183_12dpi5	RME184_28dpi2	RME185_28dpi3	RME186_28dpi5
# of reads sequenced	188,403,384	200,004,037	207,498,688	233,776,231	35,175,257	35,533,753	210,126,365	200,812,736	220,945,510	208,349,590
# of reads that pass filtering	143,093,298	148,148,154	167,092,766	152,992,606	29,036,770	32,269,823	162,874,887	161,542,113	166,119,515	161,747,104
# of reads mapped	142,694,960	141,521,864	166,790,799	106,747,105	19,546,058	23,135,344	161,685,834	160,458,448	162,979,526	158,801,829
# of uniquely mapped reads	128,621,358	128,037,393	151,597,744	95,029,819	18,009,821	20,944,501	144,392,605	145,817,680	149,448,346	146,943,402
Mappability [%]	99,72	95,53	99,82	69,77	67,31	71,69	99,27	99,33	98,11	98,18
library preparation kit	Epicentre	Epicentre	Epicentre	Illumina	Illumina	Illumina	Epicentre	Epicentre	Epicentre	Epicentre

Table 3. Mappability of reads.

6 dpi samples are highlighted in blue, 12 dpi samples in yellow, and 28 dpi samples in green. Note that sample 12 dpi #5 was prepared with two different library preparation kits (Epicentre and Illumina) and sequenced at different sequencing depths.

4.3.3 Comparison of two biologically identical samples prepared with different library preparation kits and sequenced at different sequencing depths

In order to compare these two different library preparation kits, one sample (12 dpi # 5) was prepared using both sequencing kits and sequenced at different sequencing depths. Even though there was a high degree of correlation between the two samples (Spearman $r = 0.955$), the higher sequencing depth allowed a more in-depth characterization of especially very lowly expressed transcripts. A comparison of the number of genomic units detected at different rpkm cutoffs is given in Table 4. Among all genomic regions analyzed, the number of detected genomic units was higher in the sample with the higher sequencing depth. By looking at the distribution of the percentages according to different expression levels, it immediately became evident that especially lowly expressed genomic units had a lower

chance of detectability in the sample with the lower sequencing depth and prepared with the Illumina library preparation kit. While 4.8 % of exonic sense genomic units were detected below 0.1 rpkM in the low sequencing-depth sample, 6.44 % were detected in the high sequencing-depth sample. This effect was even more pronounced in the very lowly expressed intergenic class. While the difference at a cutoff of 0.1 rpkM was not so pronounced (78.6 vs 83.3 %, respectively), almost twice as many genomic units could be detected below 0.001 rpkM in the high sequencing-depth sample (22.9 vs 42.9 %, respectively). Overall, there was no difference in detectability up to 0.1 rpkM. However, the high sequencing-depth sample was able to detect almost twice as many genomic units expressed below 0.001 rpkM (7 vs 12.25 %, respectively). This direct comparison of the same biological sample sequenced at different sequencing depths clearly shows that a higher sequencing depth is necessary to detect also low abundant transcripts, like different classes of lncRNAs. Indeed, the ENCODE consortium has shown that genes that are expressed at low levels (i.e., fewer than 10 FPKM) could only be accurately quantified with 80 million mapped reads (ENCODE 2011). Therefore, it is recommended to sequence at least 100 million reads per sample for new transcript discovery (Zeng & Mortazavi 2012). Although such a high sequencing depth causes a higher duplication level (Fig 8A and 8A'), it is still necessary to achieve a saturation of the relative error (Fig 8B and 8B'). A higher duplication level means that many fragments were sequenced several times due to the high degree of sequencing, thus not yielding much gain by "oversequencing" (Fig 8A'). Still, the saturation curve predicts that there is an ongoing reduction of the relative error, ultimately improving the precision of the reads (Fig 8B'). The relative error is calculated by resampling several rpkM values when using 5%, 10%,...95%, 100% of total reads, thereby calculating when the sequencing depth was saturated. According to the saturation curve, a saturation of the relative error starts at approximately 60 % (Fig 8B'). Given that 60% of 161 million reads corresponds to approximately 100 million reads, the recommendation of sequencing 100 million reads (Zeng & Mortazavi 2012) can be confirmed by this empirical data. Altogether, the higher degree of sensitivity for especially lowly expressed genomic units (Table 4), and the saturation analysis (Fig 8B') suggested to use a high sequencing depth, especially for the discovery of novel transcripts.

Table 4
Comparison of two biologically identical samples sequenced at different sequencing depths

	RMN014 12 dpi #5 *	RME183 12 dpi #5
total reads mapped	23,135344 x 10 ⁶	161,685834 x 10 ⁶
<i>Number of genomic units</i>		
# of genomic units exonic sense	34511	34953
# of genomic units exonic antisense	15582	31329
# of genomic units intergenic	11415	11864
# of genomic units intronic sense	21917	27645
# of genomic units intronic antisense	15621	23418
# of genomic units total	99046	129209
<i>Percentage of genomic units according to different expression levels</i>		
exonic sense genomic units < 0.1 rpkm [%]	4,8	6,44
intergenic genomic units < 0.1 rpkm [%]	78,63	83,33
intergenic genomic units < 0.001 rpkm [%]	22,92	42,91
all genomic units < 0.1 rpkm [%]	43,96	42,32
all genomic units < 0.001 rpkm [%]	7,01	12,25

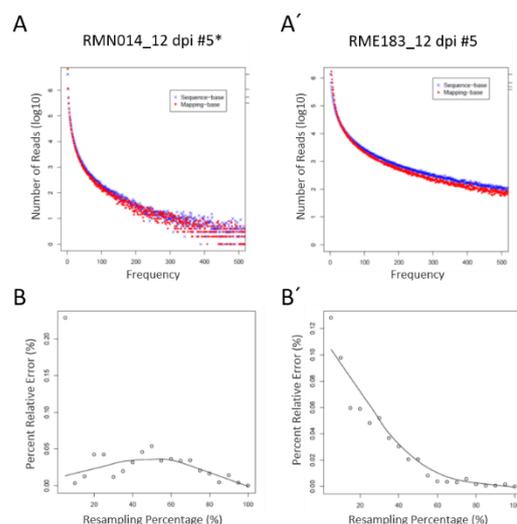


Table 4 and Fig 8. Comparison of two biologically identical samples prepared with different library preparation kits and sequenced at different sequencing depths.

Table 4. Total numbers and percentages of different genomic units detected at different rpkm cutoffs.

Fig 8 A-A') Reads duplication rate. Reads with identical sequence (blue) and reads mapped to the exactly same genomic location (red). Note that the duplication level in (A') is higher than in (A) due to higher sequencing depth. **B-B'**) Saturation analysis. While there is no saturation in (B), the saturation starts at approximately 60% in (B').

Since both biologically identical samples showed a high degree of correlation (Spearman $r = 0.955$), and samples have to be independent from each other for statistical analysis, we excluded sample RMN014_12dpi5* from the following analyses.

4.3.4 Mapping to genomic regions, directionality, reads distribution and expression level

Central mapping statistics are summarized in Table 5. In sum, more than 90% of the reads mapped to the genome. Among mapped reads, more than 90% mapped to unique loci, 3.5% to multiple loci, and 5.7% yielded no primary hits. Overall, 87.5% were non-spliced reads and 12.5% spliced reads. The number of spliced reads is likely to be higher, but since the libraries have a 3' end bias, splicing events at the 5' end can not be captured. Looking at the gene body coverage, reads mostly mapped to the 3' end of transcripts, with only a minor fraction capturing the 5' end (Fig 9A). Since the amplification is based on oligo dT primers, such a transcript coverage is expected and in the range with other single-cell RNA-seq studies using

the Tang-amplification protocol (Luo et al 2015). Yet, such a bias prevents a comprehensive analysis of possible transcript isoforms.

This 3' end bias was also apparent when looking at the reads distribution to different genomic regions (Fig 9B and Table 5). A detailed reads distribution analysis of unique reads showed that the largest fraction mapped to the 3'UTR. In total, 62% of reads were mapping to exonic regions (CDS, 5'UTR, 3'UTR). The remaining reads mapped to non-protein coding regions, with the largest fraction mapping to intergenic regions (17%), 14% mapping to intronic regions and 7% mapping antisense to exonic and intronic regions (5.5% and 1.3% respectively) (Fig 9B and Table 5). These numbers are in the range with the read mappings from mouse cerebral bulk cells after polyA+ selection starting with 1 μ g RNA, i.e., 59%, 15% and 23% mapping to exonic, intronic and intergenic regions, respectively (Cui et al 2010), indicating that the amplification preserved the overall transcript abundances.

Since this RNA amplification protocol was designed to keep directionality, a very important parameter to test is strandness. By comparing the strandness of reads to the strandness of the reference genome it was calculated that on average more than 94% of reads indeed were strand-specific (Table 5). Both the Illumina and the Epicentre libraries obtained the same percentage of mappability, indicating that both library preparation methods are efficient in keeping directionality. Moreover, the total number of reads mapping to the antisense strand was roughly the same as obtained by Direct RNA-Sequencing in human liver, i.e., 7% vs 8% respectively (Ozsolak et al 2010), indicating that this amplification protocol can be used efficiently for directional single-cell RNA-Seq.

Next, the distribution of the expression levels of both protein-coding RNAs and different classes of lncRNAs was examined. Consistent with previous observations (Derrien et al 2012a, Derrien et al 2012b, Marinov et al 2014) lncRNAs are much lower expressed than protein-coding transcripts. While protein-coding exonic sense reads showed a median expression of 11 rpkm, for exonic antisense transcripts the median was only 0.7 rpkm. The median expression was even lower for intronic sense and antisense transcripts (0.03 and 0.005 rpkm, respectively). Intergenic transcripts displayed the lowest expression with a median of 0.002 rpkm (Fig 9C). Protein-coding exonic sense transcripts displayed expression values between 0.001 and 1000 rpkm, confirming a dynamic range of 6 orders of magnitude, and also different classes of lncRNAs reach a dynamic range of 4 to 5 orders of magnitude

(Fig 9C). Overall, this represents a good sensitivity to detect both high and low expressed transcripts for both coding and ncRNAs.

Consistent with earlier estimates (Marinov et al 2014, Suzuki et al 2015), the peak expression of protein-coding genes (exonic sense) was between 10 to 50 rpkm (23.4%). Still, more than half of the transcripts (53.4%) were expressed below 10 rpkm. Almost 30% were expressed below 1 rpkm and even 13.6% below 0.1 rpkm (Fig 9C).

The picture is even more dramatic looking not only at protein-coding transcripts but at all expressed transcripts. The ENCODE consortium has plotted all expressed human genomic regions dependent on the expression level. Among 75% of transcribed genomic regions, 50% are expressed below an fpkm of 0.1. and 70% below 1 transcript per cell (Kellis et al 2014). Consistent with these numbers, 48.2% of polyadenylated transcripts in the present dataset were expressed below 0.1 rpkm and 68.1% below 1 rpkm (Fig 9C'), indicating that the overall expression levels were not distorted by this amplification protocol.

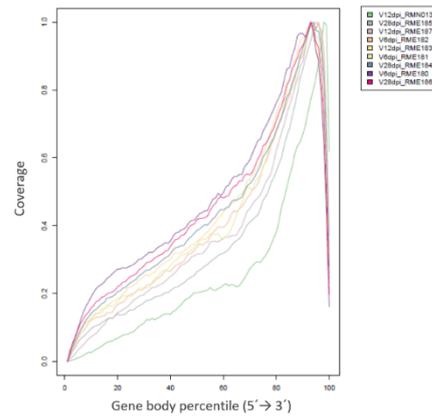
Altogether, these findings strongly suggest that both the overall reads distribution in sense and antisense as well as expression levels were not distorted by this amplification protocol. One major limitation though, is the pronounced 3' end bias of reads, making it difficult to make assumptions on new splice isoforms. Therefore, this dataset mainly focused on the identification of known and novel transcripts, but not on isoform expression.

Table 5

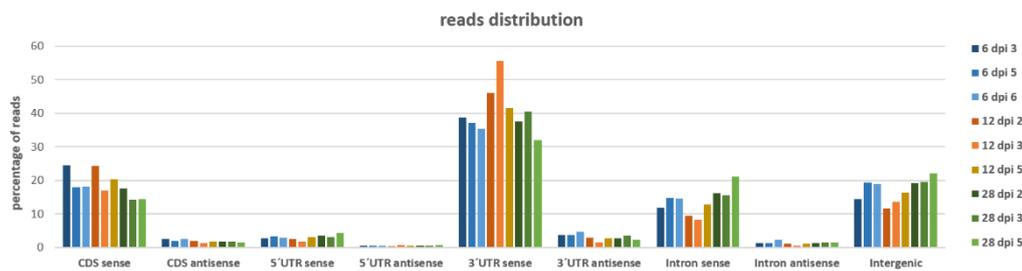
Summary of read mapping ^a	
total reads mapped per sample	135,69 x 10 ⁶ (± 47,36 x 10 ⁶)
mappability [%]	91,89 (± 13,32)
unique [%]	90,79 (± 1,20)
non-unique [%]	3,47 (± 0,59)
non-primary hits [%]	5,74 (± 0,65)
splice reads [%]	12,51 (± 3,08)
non-splice reads [%]	87,49 (± 3,08)
Directionality ^b	
strand-specific reads [%]	94,14 (± 3,11)
Read distribution ^c	
Exonic region sense (CDS, 5'UTR, 3'UTR) [%]	62,19 (± 7,82)
Intronic region sense [%]	14,14 (± 3,78)
Intergenic region [%]	17,23 (± 3,41)
Antisense total [%]	6,76 (± 1,67)

^aRaw reads mapped to the mouse genome (mm10)
^bStrandness of reads compared to the strandness of reference gene
^cThe distribution of the uniquely mapped reads

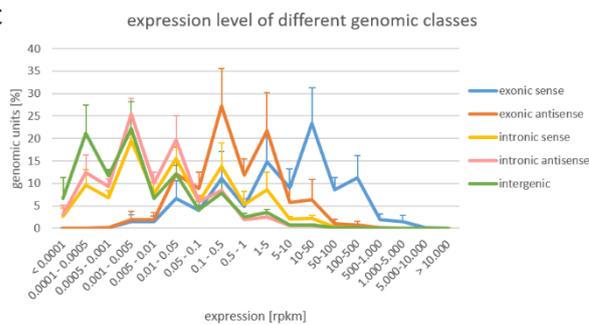
A



B



C



C'

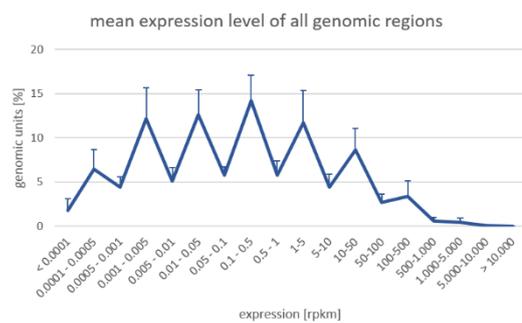


Table 5 and Fig 9. Summary of read mapping.

Table 5. Percentages of mappability, directionality and reads distribution to different genomic regions.

Fig 9 A) Gene body coverage. Note that there is a 3' end bias in all samples due to oligo-dT priming of polyA tails. **B)** Reads distribution profile mapping to different genomic regions in sense or antisense. 6 dpi samples in blue, 12 dpi samples in orange/yellow, 28 dpi samples in green. **C-C')** Expression level of different genomic classes (C) and overall mean expression level of all genomic regions (C').

4.3.5 Reproducibility and sensitivity

Other parameters to evaluate the quantitative and qualitative performance of single-cell RNA-Seq studies are sensitivity and precision/reproducibility. Precision/reproducibility is being measured by how well the results can be reproduced on replicate samples.

Unsupervised hierarchical clustering analysis showed that samples obtained from the same timepoint were more closely associated with each other than with samples from other timepoints (Fig 10A). Moreover, 6 dpi samples and 12 dpi samples were more closely associated than each was to 28 dpi samples, which may reflect the fact that 6 dpi and 12 dpi neurons are in the phase of extensive axonal and dendritic growth (Sun et al 2013) and are not yet synaptically connected within the hippocampal circuit (Deshpande et al 2013, Gu et al 2012, Toni et al 2008, Toni & Schinder 2016, Vivar et al 2012).

Also the principal component analysis (PCA) showed that samples from the same timepoint clustered together (Fig 10B).

Sensitivity is typically measured by counting the number of genes detected within a cell (Sasagawa et al 2013). Even though this analysis is comprised of 30 to 100 cells, the number of detected genes is still a valuable parameter to judge the amplification efficiency. In 6 dpi and 12 dpi samples the mean number of detected RefSeq coding genes per sample was more than 8000 (8680 vs 8218, respectively), while in 28 dpi samples the number was even higher (11682) (Fig 10C and 10C').

Even though a better measurement of sensitivity would be to calculate the number of genes detected in single cells divided by the number of genes detected in non-amplified samples started with bulk cells (Sasagawa et al 2013), such a measurement is impossible here as only a limited number of newborn neurons exists within the adult hippocampus.

The current numbers of expressed genes per sample are in the range with other single-cell RNA-Seq studies sequenced at high sequencing depth (> 20 million reads) (Picelli et al 2013, Sasagawa et al 2013, Tang et al 2009, Yan et al 2013), suggesting that the overall sensitivity is good.

Taken together, these findings strongly suggest that this amplification protocol followed by high sequencing achieves reproducible and highly sensitive results.

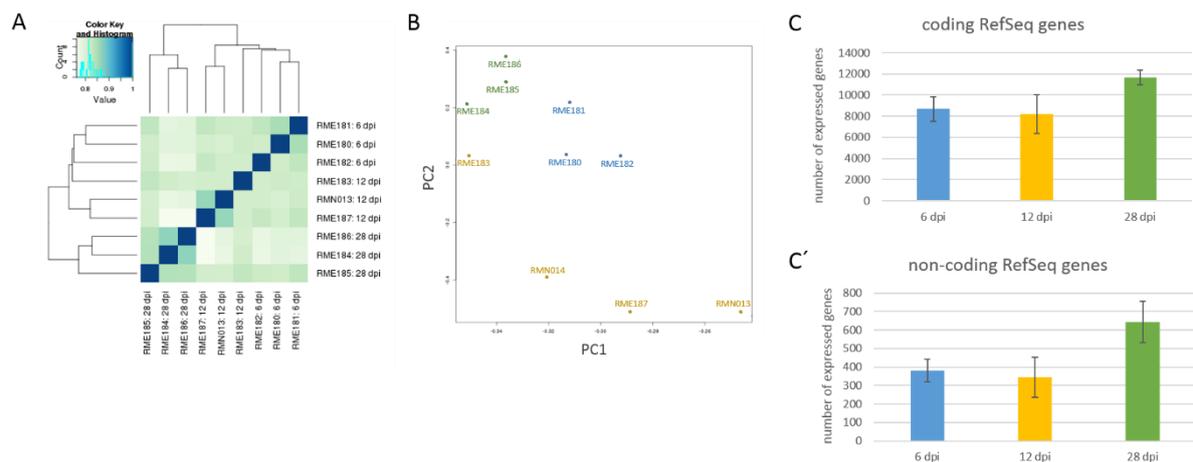


Fig 10. Reproducibility and sensitivity.

A) Unsupervised hierarchical clustering analysis. Samples from the same timepoint were more closely associated than with samples from other timepoints. **B)** Principal component analysis (PCA). 6 dpi (blue), 12 dpi (yellow) and 28 dpi (green) samples clustered in the PCA analysis, respectively. **C)** Sensitivity analysis. Number of detected RefSeq coding (C) and RefSeq non-coding (C') genes for 6 dpi (blue), 12 dpi (yellow) and 28 dpi (green) samples.

4.3.6 Validation of specific maturation stages

In order to confirm the expression of known and accepted marker genes at specific maturation stages, i) the read distribution on several accepted marker genes on the IGV (integrative genomics viewer) was analyzed (Fig 11A) and ii) the count expression was plotted (Fig 11B). Housekeeping genes, such as *Gadph* or *Ubiquitin B/Ubb*, were universally expressed across all samples. Stem cell markers such as *Sox2* or *Nestin* were absent in two (out of three) 6 dpi samples. A remaining expression is most likely due to a yet incomplete shut down in early immature neurons (Kempermann et al 2004, Liu et al 2010, Steiner et al 2006). Immature neuronal genes, such as *NeuroD1*, *DCX*, *Tubb3*, *Tubb5*, *Ndrd2*, *Sox4*, *Sox11*, *Lhx2*, *Limk1* were highest expressed at 6 and 12 dpi (Breuss et al 2012, Brown et al 2003, Gao et al 2009, Liu et al 2011, Meng et al 2002, Mu et al 2012, Parent et al 1997, Subramanian et al 2011), while there was a significant decrease at 28 dpi. Finally, known mature neuronal genes like *Klf9* and *Calbindin* showed highest expression at 28 dpi (Kuhn et al 1996, Scobie et al 2009). Consistently, synaptic transmission genes like the ionotropic glutamate receptors *Grin2a* and *Grin3a* and the metabotropic glutamate receptor *Grm1* were highest expressed at 28 dpi. Also the voltage gated calcium channel *Cacnb4* was highly expressed at

28 dpi and could potentially become a new marker gene to label synaptically integrated neurons in the adult hippocampus. Notably, *Cacnb4* was shown to be highly expressed in the granule cell layer of the adult dentate gyrus (similar to NeuN), and its expression increased in parallel with axon outgrowth and synapse formation (Ferrándiz-Huertas et al 2012, Tadmouri et al 2012), suggesting an involvement of *Cacnb4* in late-phase differentiating neurons.

Taken together, this analysis shows that it is possible to isolate immature neurons with high temporal resolution from the intact DG niche *in vivo* allowing us to provide a comprehensive description of their transcriptional profiles in the course of adult hippocampal neurogenesis.

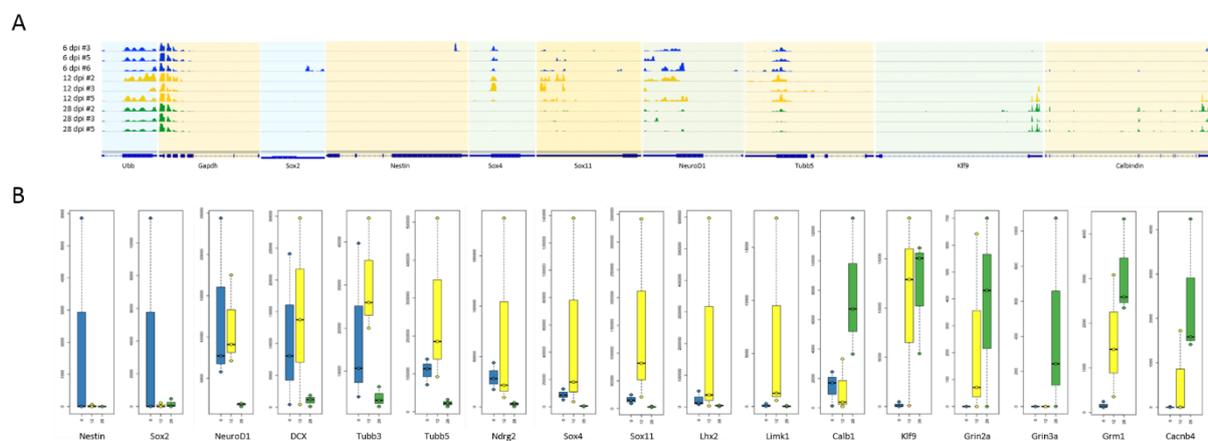


Fig 11. Expression of marker genes corresponds to developmental stage.

A) Read coverage of individual samples along selected genomic loci marking different developmental stages of adult hippocampal neurogenesis, including housekeeping genes (Ubiquitin B/Ubb, Gapdh), known stem cell markers (Nestin, Sox2), known immature neuronal markers (Sox4, Sox11, NeuroD1, Tubb5) and known mature neuronal markers (Klf9, Calbindin). 6 dpi samples are marked in blue, 12 dpi samples in yellow, 28 dpi samples in green. B) Count expression of known marker genes for stem cell markers (Nestin, Sox2), known immature neuronal markers (NeuroD1, DCX, Tubb3, Tubb5, Ndr2, Sox4, Sox11, Lhx2, Limk1), known mature neuronal markers (Calbindin, Klf9), synaptic transmission genes (Grin2a, Grin3a, Grm1) and potential new mature markers (*Cacnb4*). Boxplots are represented for 6 dpi (blue), 12 dpi (yellow) and 28 dpi (green) samples.

4.4 Differentially expressed genes

Next, I sought to identify genes with dynamic expression over the course of maturation. In total, 4339 genes (including 318 non-coding RefSeq genes) were found to be differentially expressed between time-points. 1414 genes and 837 genes were significantly upregulated

and downregulated, respectively, from 6 to 12 dpi. From 12 to 28 dpi, 1647 genes were upregulated and 648 genes became downregulated (Fig 12A).

Genes with significant changes in their temporal expression were grouped into clusters based on similarity in their temporal expression patterns. This analysis generated 6 clusters: genes highly expressed at a single timepoint or highly expressed at two timepoints (Fig 12B).

To gain insight into stage-associated dominant regulatory, developmental, and cell biological processes, functional categorization of dominant overrepresented terms using DAVID functional annotation clustering was performed (Fig 12C).

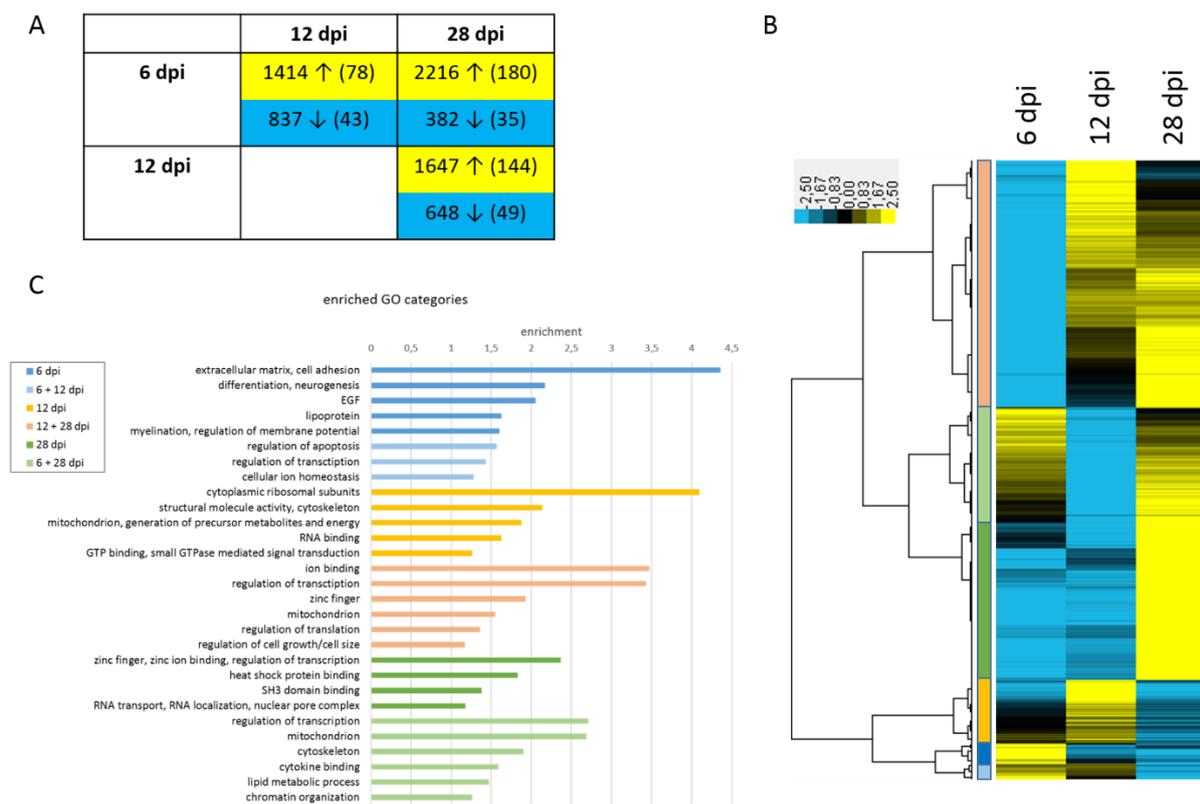


Fig 12. Differentially expressed genes.

A) Number of up- (yellow) and downregulated (blue) annotated RefSeq genes between timepoints (annotated non-coding RefSeq genes are shown in brackets). **B)** Differentially expressed genes grouped in clusters according to their expression during differentiation. **C)** DAVID-functional annotation analysis of enriched GO categories in each cluster.

In situ hybridization data from the Allen Brain Atlas indicated that a number of genes within these clusters, showed strong enrichment or even specific expression in the dentate gyrus. Moreover, a number of genes was limited to the two neurogenic germinal zones of the adult mouse brain (Lein et al 2007), strongly suggesting their function in neurogenesis associated processes (Fig 13B, 14B, 15B).

General cellular processes such as the regulation of transcription was enriched in 4 out of 6 clusters, suggesting that neuronal maturation and differentiation is tightly regulated by transcription factors (Beckervordersandforth et al 2015). In total, 374 transcription factors were found to be differentially expressed. Also mitochondrial regulation was enriched in several clusters, emphasizing the importance of mitochondrial regulation in newborn hippocampal neurons (Shin et al 2015, Steib et al 2014).

4.4.1 Genes and cellular processes enriched at 6 dpi

DAVID functional annotation clustering of 6dpi high genes revealed a clear enrichment of genes involved in neurogenesis and differentiation (Fig 13A). This cluster included transcription factors involved in differentiation (Nfic, Epas1, Pax4, Rara, Uncx, Sall3), genes related to early neuronal development and growth (e.g., Camk1, Gap43, Nnat, Fgf1, Arhgap22) and components of the Wnt-signaling pathway (Wnt3, Fzd9, Lrp5), the latter group underlining the importance of Wnt-signaling in controlling neuronal fate determination in adult neurogenesis (Kuwabara et al 2009, Lie et al 2005).

The highest enriched GO term for highly expressed genes at 6 dpi was “extracellular matrix and cell adhesion” (e.g., Hspg2, Spock1, Bgn) (Fig 13A'). The extracellular matrix (ECM) is an essential component of the neurogenic niche and plays an important role in the control of stem cell maintenance, proliferation, and differentiation (Faissner & Reinhard 2015). While numerous studies have highlighted the importance of transcription factors, morphogens, cytokines and growth factors as intrinsic and extrinsic factors of stem cell regulation and differentiation, less attention has been paid to the constituents of the molecular environment of the niche. The high expression of ECM related genes in the 6dpi population implies that the neurogenic lineage itself significantly contributes to the creation of a neurogenic environment.

Among those ECM and cell adhesion genes, Hspg2/perlecan was identified, a large multidomain heparin sulfate proteoglycan that binds to and cross-links many ECM components and cell-surface molecules. Perlecan-null mice die before birth due to impaired cartilage development and show impaired telencephalic development due to delayed cell cycle progression and decreased progenitor differentiation (Girós et al 2007). In a perlecan-rescue model by expressing recombinant perlecan specifically in the cartilage of perlecan-null mice, Kerever and colleagues were able to study the role of Hspg2/perlecan in the adult brain. Within the adult SVZ, the number of both neural stem cells and new neurons integrating into the olfactory bulb are reduced, owing to insufficiency in FGF2 signaling, suggesting that Hspg2/perlecan promotes neural stem cell self-renewal and adult neurogenesis (Kerever et al 2014).

Also Spock1 (also known as Testican-1), a chondroitin/heparin sulfate proteoglycan, is expressed abundantly in hippocampal neurons (especially in the subgranular zone of the dentate gyrus) (Fig 13B) as well as ependymal cells lining the ventricles (Iseki et al 2011, Marr et al 2000), suggesting to play a role in the neurogenic niches of the adult brain.

Further studies are needed to address the questions which mechanisms are controlling the composition and structure of ECM assemblies in different parts of the nervous system and how the ECM affects cell behavior. Knowing the exact composition of the ECM in the SGZ of the adult hippocampus will open new avenues for the development of cell-based therapies.

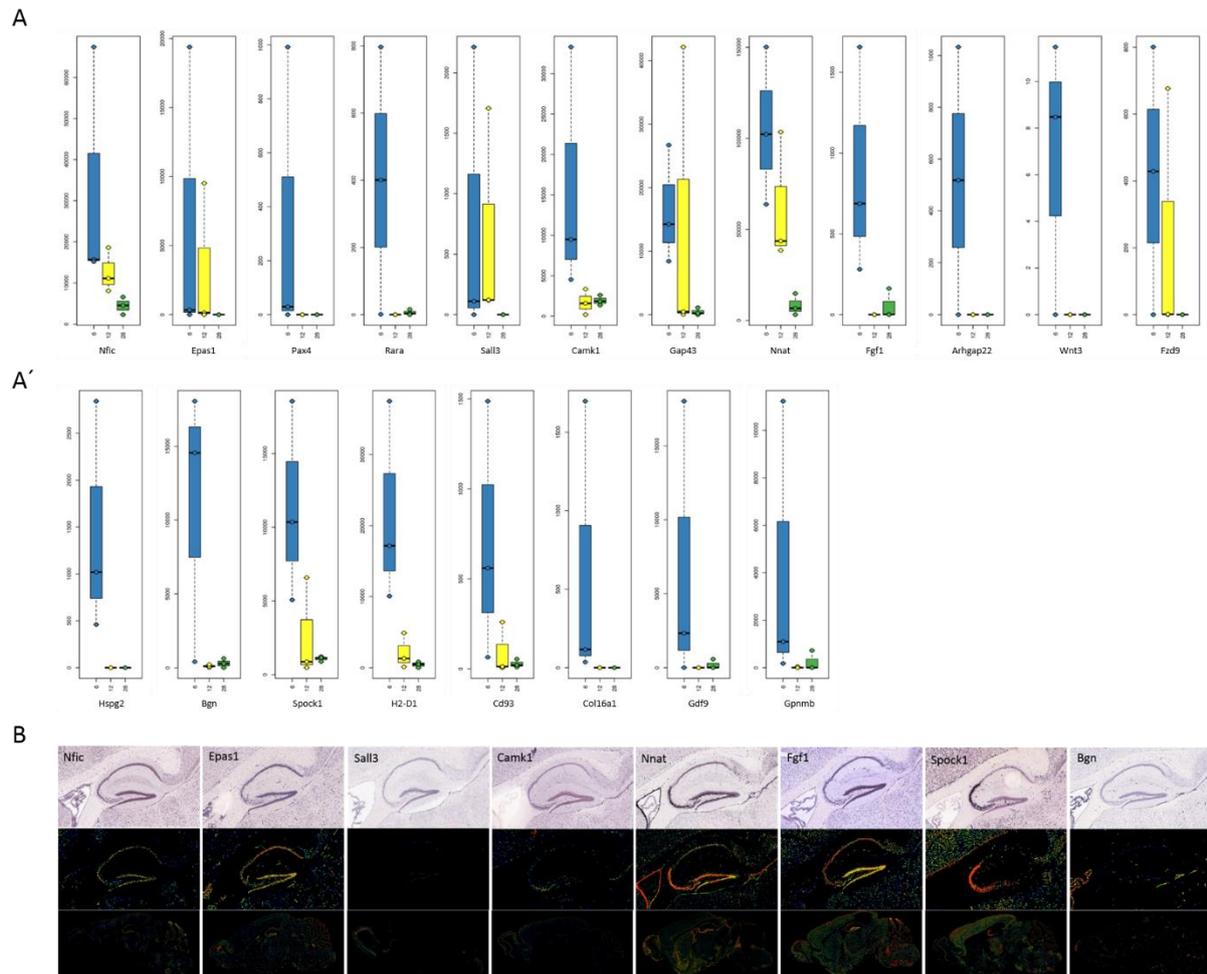


Fig 13. Highly expressed genes at 6 dpi.

A-A') Genes belonging to the GO cluster “neurogenesis” (A) and “extracellular matrix” (A'). Count expression is depicted for 6 dpi samples (blue), 12 dpi samples (yellow) and 28 dpi samples (green). **B)** *In situ* hybridization expression data for selected genes in the hippocampus and in the total mouse brain taken from the Allen Brain Atlas (Lein et al 2007).

4.4.2 Genes and cellular processes enriched at 12 dpi

At 12 dpi the highest enriched GO term belonged to the category of cytoplasmic ribosomal subunits (Fig 14A). Intriguingly, an increase of ribosomal subunits was shown from the transition of quiescent neural stem cells to Nestin-positive early neuroblasts within the hippocampal neurogenesis niche (Shin et al 2015), already suggesting that the priming of the protein synthesis machinery starts very early of adult hippocampal neurogenesis. Together with this data, it is suggestive that several ribosomal genes become upregulated in early immature neuroblasts, peak in immature neurons at around 12 dpi, and become

downregulated again in mature hippocampal neurons. Moreover, the specific expression within the hippocampus and the SVZ/RMS of some ribosomal genes (Rpl18a, Rpl31, Rps10) further suggests a specific regulatory function in the course of adult neurogenesis (Fig 14B).

Even though ribosomal proteins are generally considered as abundant and ubiquitously expressed genes, which exert a constitutive rather than a regulatory function (Klinge et al 2012), it was shown that individual ribosomal genes are expressed in a cell type- and tissue-specific manner, and to be involved in the regulation of tissue patterning during embryonic development by conferring transcript-specific translational control (Kondrashov et al 2011, Shi & Barna 2015, Xue & Barna 2012).

Functionally, an upregulation of ribosomal genes either suggests that i) the protein synthesis rate is increased in immature dentate granule neurons or ii) that specific ribosomal proteins confer transcript-specific translational control, thereby adding a new layer of specificity in the control of gene expression. While the first option would imply that more proteins are synthesized in newborn neurons, which has not been shown so far, the transcript-specific translational control seems to be a more appealing hypothesis. For example, RPL38 deletion in tail short (Ts) mice results in highly specific defects in the formation of the axial skeleton, suggesting that ribosomes have a regulatory role during tissue patterning by regulating the translation initiation of certain mRNA transcripts, like *Hox* mRNAs (Kondrashov et al 2011).

Also other ribosomal proteins were found to be specifically expressed and to be linked to diseases (reviewed in Xue & Barna 2012). The specific expression of some ribosomal proteins may generate “specialized ribosomes” that have substantial impact on the translation of mRNA transcripts in space and time (Shi & Barna 2015, Xue & Barna 2012).

Altogether, these data suggest that there are numerous levels of gene regulation, of which ribosome-mediated control is emerging as one of them by directing how the genome is translated in space and time.

Other interesting enriched GO terms at 12 dpi were cytoskeleton organization and small GTPase mediated signal transduction, most likely reflecting the rapid dendrite and axonal outgrowth and their regulation. It is well known that the cytoskeleton of newborn neurons undergoes extensive and dynamic remodeling to facilitate neurogenesis, cell migration and terminal differentiation (Heng et al 2010). Indeed, newborn neurons from the adult

hippocampus undergo an initial phase of rapid axonal and dendritic growth between 7 and 21 days (Sun et al 2013), which is likely accompanied by cytoskeleton reorganization. Many of the cytoskeletal organization associated genes being highest expressed at 12 dpi and becoming downregulated at 28 dpi belonged to the microtubule binding family (Tubb3, Tubb5, Tuba1a, Tuba1b, Tuba1c) or the actin binding family (Actg1, Fmn1, Marcks1, Coro1a, Mllt4, Csrp1) (Fig 14A' and 14B).

Another interesting enriched GO term at this time point was mitochondrial energy supply. Especially genes belonging to the complex I (Ndufa3, Ndufa13, Ndubf6, Ndufs2, Ndufs7) and complex V (Atp5d, Atp5e, Atp5o) of the electron transport chain were upregulated at 12 dpi (Fig 14A''). High average expression levels of mitochondrial complex I genes were detected in periods of intense neurogenesis during mouse embryonic brain development (Wirtz & Schuelke 2011). Notably, the highest increase of mitochondrial genes was observed postnatally in hippocampal CA1/CA3 pyramidal neurons, a phase of synaptogenesis and maturation, suggesting that neurogenesis and synaptogenesis strongly depend on proper mitochondrial function.

Consistent with this finding, a single-cell RNA-Seq screen recently suggested a shift in energy metabolism from glycolysis to oxidative phosphorylation around the transition from Nestin-positive activated neural stem cells to Nestin-positive early neuroblasts (Shin et al 2015). Together with this dataset it is suggestive that oxidative phosphorylation remains the main energy source for immature granule neurons. The increasing expression of genes associated with oxidative phosphorylation may reflect the increasing reliance of the maturing neuron on mitochondria dependent oxidative phosphorylation for energy supply.

Notably, misregulations of some of these genes leading to mitochondrial dysfunction (Ndufa3, Ndubf6, Uqcr11, Cox6c, Atp5d, Atp5e, Atp5o) have been associated with several neurological diseases (Johri & Beal 2012), which show impaired adult neurogenesis (Winner & Winkler 2015), again suggesting that coordinated mitochondrial gene expression is important for proper development and function of neurons.

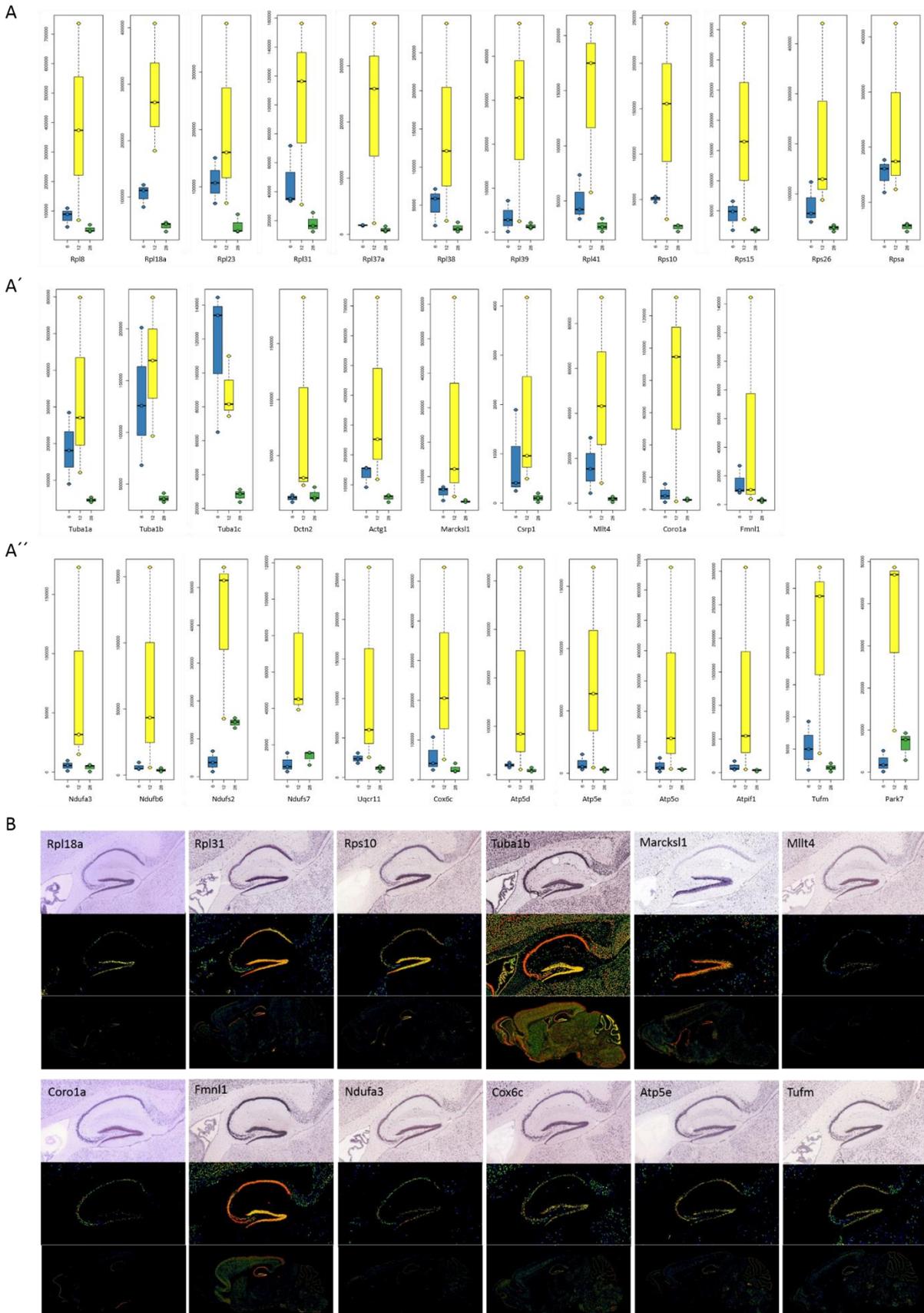


Fig 14. Highly expressed genes at 12 dpi.

A-A'') Genes belonging to the GO cluster “cytoplasmic ribosomal subunits” (A), “cytoskeleton organization” (A') and “mitochondrial energy supply” (A''). Count expression is depicted for 6 dpi samples (blue), 12 dpi samples (yellow) and 28 dpi samples (green). **B)** *In situ* hybridization expression data for selected genes in the hippocampus and in the total mouse brain taken from the Allen Brain Atlas (Lein et al 2007).

4.4.3 Genes and cellular processes enriched at 28 dpi

Zinc finger proteins were the highest enriched GO term at 28 dpi and were also highly enriched in the cluster 12 + 28 dpi. Zinc finger proteins are a highly heterogeneous group of proteins that impact on gene expression through the regulation of transcription, translation, mRNA trafficking, and chromatin remodeling (Laity et al 2001). Among zinc finger proteins with a high expression in 28 dpi neurons was the Krüppel-like transcription factor Klf9, which was previously assigned a role in neuronal maturation during adult hippocampal neurogenesis (Scobie et al 2009).

Another interesting zinc finger gene Zfp365 (also called DBZ (DISC1-binding zinc finger protein)) was specifically found in 28 dpi samples, but not yet in 12 dpi or 6 dpi samples (Fig 15A), suggesting that Zfp365 (DBZ) plays a role in late-phase maturation of adult-born hippocampal neurons. In favour of this assumption, DBZ was shown to regulate neurite outgrowth *in vitro* (Hattori et al 2007), and DBZ deficiency *in vivo* impaired dendritic development and branching in interneurons of the somatosensory cortex (Koyama et al 2013) and in cortical pyramidal neurons (Koyama et al 2015). Besides a decreased dendritic arborization, also the morphogenesis of spines was impaired (Koyama et al 2015), indicating that DBZ is important for the proper development and integration of neurons. At the mRNA level DBZ (Zfp365) is specifically expressed in the brain, with highest expression in the cortex, hippocampus and striatum (Hattori et al 2007) (Fig 15B).

Several other differentially expressed zinc finger protein-encoding genes displayed a particularly high expression in both adult neurogenic regions, especially in the hippocampus (Zfp236, Zfp277, Zfp322a, Zfp451, Zfp629, Zscan6) (Fig 15B), raising the possibility that these zinc finger proteins may be regulators of differentiation and maturation of different neuronal subtypes.

Another intriguing functional term at 28 dpi was the regulation of RNA transport and nuclear pore complex (Fig 15A'). The nuclear pore complex (NPC) is a multiprotein assembly composed of about 30 different proteins, which is embedded in the nuclear envelope and forms transport channels whose primary function is the exchange of molecules between the cytoplasm and the nucleus. Intriguingly, besides a role in nucleo-cytoplasmic transport, recently nucleoporins could have been linked to differentiation processes. It was shown that a change in the composition of the nuclear pore complex is critical for both myogenic and neuronal differentiation (D'Angelo et al 2012). In addition, Nup133 was found to play an important role during embryonic neurogenesis, as disruption of Nup133 impaired differentiation along the neural lineage (Lupu et al 2008). The question now is i) whether the composition of the nuclear pore complex mediates the redistribution of key molecules involved in cell differentiation or ii) whether nucleoporins are directly involved in the regulation of transcription (Burns & Wenthe 2014).

The simplistic view of the nuclear pore complex as static structures has changed a lot since the discovery of their dynamic nature, their variable composition and tissue-specific roles (Raices & D'Angelo 2012). It is now becoming clear that the NPC is a complex platform that can exert various nuclear functions, including gene transcription and the organization of chromatin structure by interacting with components of the transcriptional machinery (e.g., activator and repressor proteins), regulators of chromatin structure (e.g., chromatin remodeling complexes) and protein modifiers (e.g., desumoylases) (Ptak et al 2014).

For example, various nucleoporins have been detected in association with transcriptionally active genes inside the nucleoplasm in *Drosophila* (Capelson et al 2010, Kalverda et al 2010) and in human cells (Liang et al 2013) (and not outside at the nuclear envelope). Notably, these nucleoporin-bound genes were particularly involved in developmental regulation and the cell cycle (Kalverda et al 2010). Genes strongly interacting with nucleoplasmic Nup98 were downregulated upon Nup98 depletion and activated upon overexpression (Kalverda et al 2010). Similarly, it was shown that Nup98 and Nup133 directly bind to neurogenic genes in ESC-derived neural progenitor cells, and that overexpression and downregulation of Nup98 changed the expression levels of Nup98-bound genes, suggesting that nucleoporins may functionally interact with the genome in a dynamic manner during cell differentiation to regulate gene expression (Liang et al 2013). Two modes of gene association have been

postulated: i) several developmentally regulated genes relocate to the nuclear pore at the initial stage of transcriptional induction associated with neurogenesis (Liang et al 2013, Williams et al 2005) and astrocyte differentiation (Takizawa et al 2008) (“gene to pore” model), and ii) nucleoporins can also detach from the NPC and relocate into the nuclear interior to bind chromatin and regulate gene expression (“Nup to gene” model) (Liang et al 2013). One mechanism how nucleoporins regulate gene expression is by binding epigenetic marks or specific chromatin remodelers, a mechanism that should be analyzed by NPC-chromatin interactions (Ibarra & Hetzer 2015). Moreover, the mechanisms underlying the recruitment of nucleoporins to intranuclear regions are still uncharacterized (Ibarra & Hetzer 2015).

Taken together, the present data indicate that the composition and the architecture of nuclear pore complexes are altered during maturation, raising the interesting possibility that nuclear pore components control the neuronal maturation and functional integration of adult-born hippocampal neurons. The high and specific expression in the hippocampus (Fig 15B) further suggests that nucleoporins play a role in proper hippocampal function.

Within the cluster of 12 + 28 dpi high genes a significantly enriched GO term was regulation of translation. An increase in genes regulating the translation initiation process at 12 and 28 dpi (e.g., *Ngdn*, *Eif4g1*, *Eif4ebp3*) (Fig 15A'') might point to the important regulation of translation upon synaptic activity (Jackson et al 2010, Jung et al 2006, Kindler & Kreienkamp 2012, Sonenberg & Hinnebusch 2009). Regulation through activity-dependent local translation would – in contrast to transcriptional regulation – allow temporo-spatial control of synaptogenesis by neuronal activity and thus information-dependent synaptic integration.

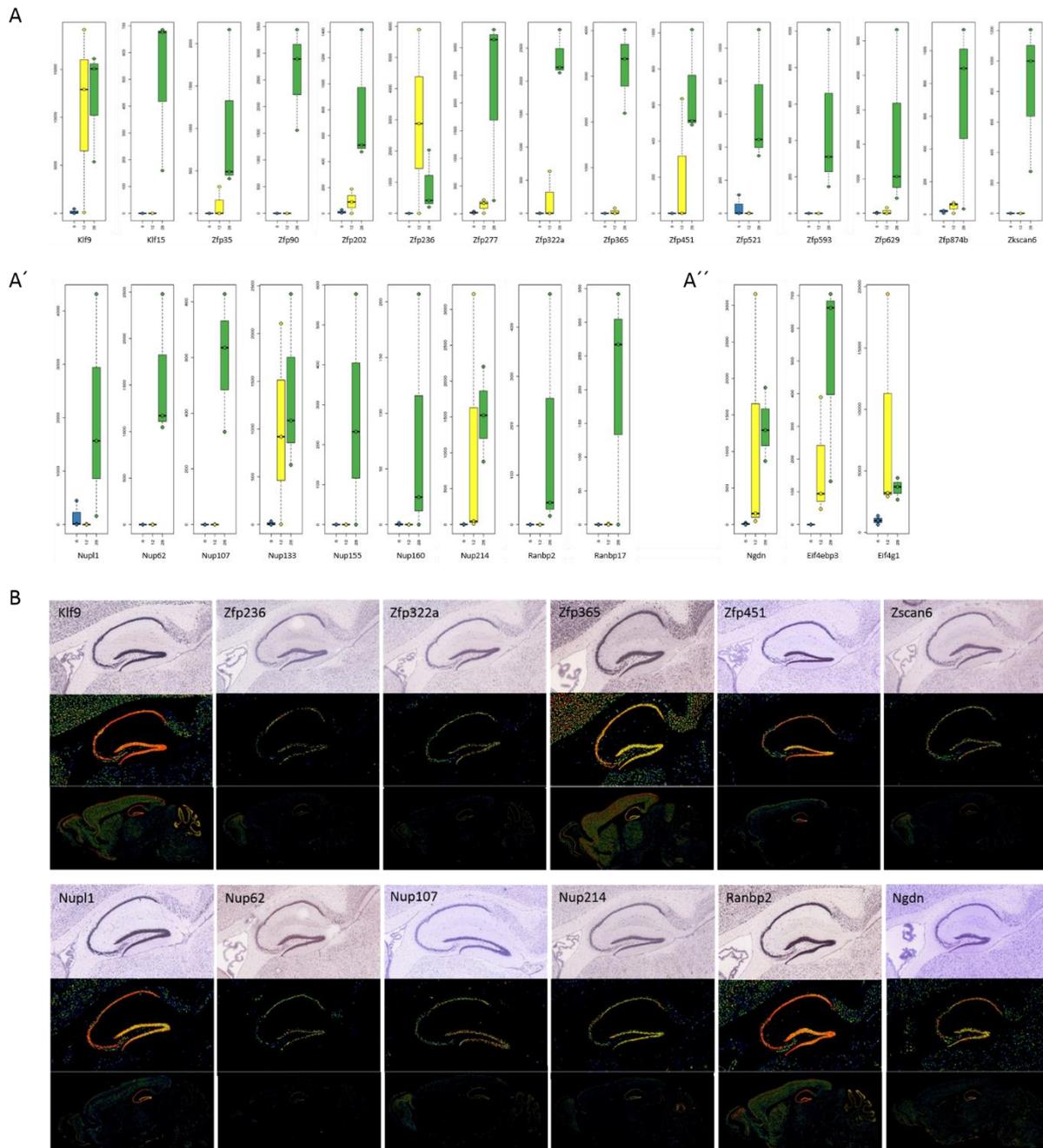


Fig 15. Highly expressed genes at 28 dpi.

A-A'') Genes belonging to the GO cluster “zinc finger” (A), “nuclear pore complex” (A') and “regulation of translation” (A''). Count expression is depicted for 6 dpi samples (blue), 12 dpi samples (yellow) and 28 dpi samples (green). **B**) *In situ* hybridization expression data for selected genes in the hippocampus and in the total mouse brain taken from the Allen Brain Atlas (Lein et al 2007).

4.5 Dynamic expression of RefSeq annotated long non-coding RNAs

In addition to protein-coding genes, 318 annotated ncRNA RefSeq genes were identified to be differentially expressed between timepoints (6 vs 12: 121; 6 vs 28: 215; 12 vs 28: 193) (Fig 12A). Approximately 60% of these annotated lncRNAs belonged to the class of long intergenic non-coding RNAs (lincRNAs), a fact that is in accordance with earlier estimates in humans (Derrien et al 2012). The remaining lncRNAs intersected protein-coding gene loci (Derrien et al 2012) (Fig 16A and 16B). In detail, 25% overlapped exonic regions in antisense, 2% percent overlapped exonic regions in sense; 2% overlapped intronic regions in sense and antisense, respectively; 3% and 6 % contained protein-coding genes in sense and antisense, respectively (Fig 16B). About 20% of these differentially expressed RefSeq ncRNAs belonged to the class of small ncRNAs. However, the number of differentially expressed small ncRNAs will likely be an underestimation, as the amplification protocol enriched for long (> 200 nt) and polyadenylated transcripts (Fig 6A). The subsequent analysis only focused on long ncRNAs.

In order to gain insight into the potential function of these differentially expressed lncRNAs, the closest protein-coding gene was analyzed, since it is suggested that many lncRNAs are linked to the activity of neighbouring genes (“guilt-by-association”) (Cabili et al 2011, Guttman et al 2009, Hu et al 2014, Rinn & Chang 2012, Sigova et al 2013, Zhang et al 2014).

Interestingly, the overall second highest enriched GO term of these protein-coding neighbours was related to neurogenesis and neuron differentiation (Fig 16C). Within this cluster, a well-known gene pair could be identified, i.e., *Emx2/Emx2os*.

The transcription factor *Emx2* is essential for dentate gyrus development (Pellegrini et al 1996) and specifically expressed in the subgranular zone of the dentate gyrus (Fig 16D). *Emx2* is expressed at 6dpi, peaks at 12 dpi, and becomes completely downregulated at 28 dpi (Fig 16D'' and 16D'''). The *Emx2os* transcript is divergently transcribed on the opposite strand of a bidirectional promoter (Fig 16D'). It shows discordant expression and increases over time and peaks at 28 dpi (Fig 16D''), raising the possibility that *Emx2os* - in analogy to its function in developing cortico-cerebral neurons - contributes to downregulation of *Emx2* in maturing DG neurons (Spigoni et al 2010).

Several other interesting lincRNA-mRNA expression pairs were identified with various kinds of expression patterns, suggesting different forms of putative regulations. Examples are given for exonic antisense transcripts (Fig 16E), and lincRNAs with distances below 10 kb (Fig 16F) or above 10 kb (Fig 16F') with respect to their closest protein-coding neighbour gene. Correlations were calculated based on the expression patterns of the two neighbouring ncRNA-mRNA genes. While antisense transcripts showed a trend towards more positive correlations (Fig 16G) (mean correlation: 0.313), lincRNAs did not show preferential positive or negative correlations with their closest protein-coding neighbouring gene (Fig 16H) (mean correlation: 0.065). However, when plotting the expression of several interesting lincRNA-mRNA gene pairs, I noticed that lincRNAs in close proximity to their protein-coding neighbouring gene (< 10 kb) displayed higher positive correlations (Fig 16F), while lincRNAs more distant to their protein-coding neighbouring gene (> 10 kb) had more often negative correlations (Fig 16F'). Indeed, when plotting the correlations of all lincRNA-mRNA gene pairs with respect to the distance, expression pairs in close proximity (< 10 kb) were more positively correlated (mean correlation: 0.165), while more distant gene pairs (> 10 kb) were more often negatively correlated (mean correlation: -0.170). A comparison between the two distributions showed a significant difference ($p = 0.018$, t-test) (Fig 16H').

Consistent with these observations, an overall trend towards more positive correlations for both antisense lincRNAs and lincRNAs with their closest protein-coding gene within 4 vs 10 kb has been reported previously (Hu et al 2014, Zhang et al 2014). Interestingly, directional RNA-Seq in the human prefrontal cortex revealed that antisense transcripts located upstream of protein-coding genes display a significant positive correlation with their protein-coding neighbours, which are themselves involved in neuronal functions (Hu et al 2014).

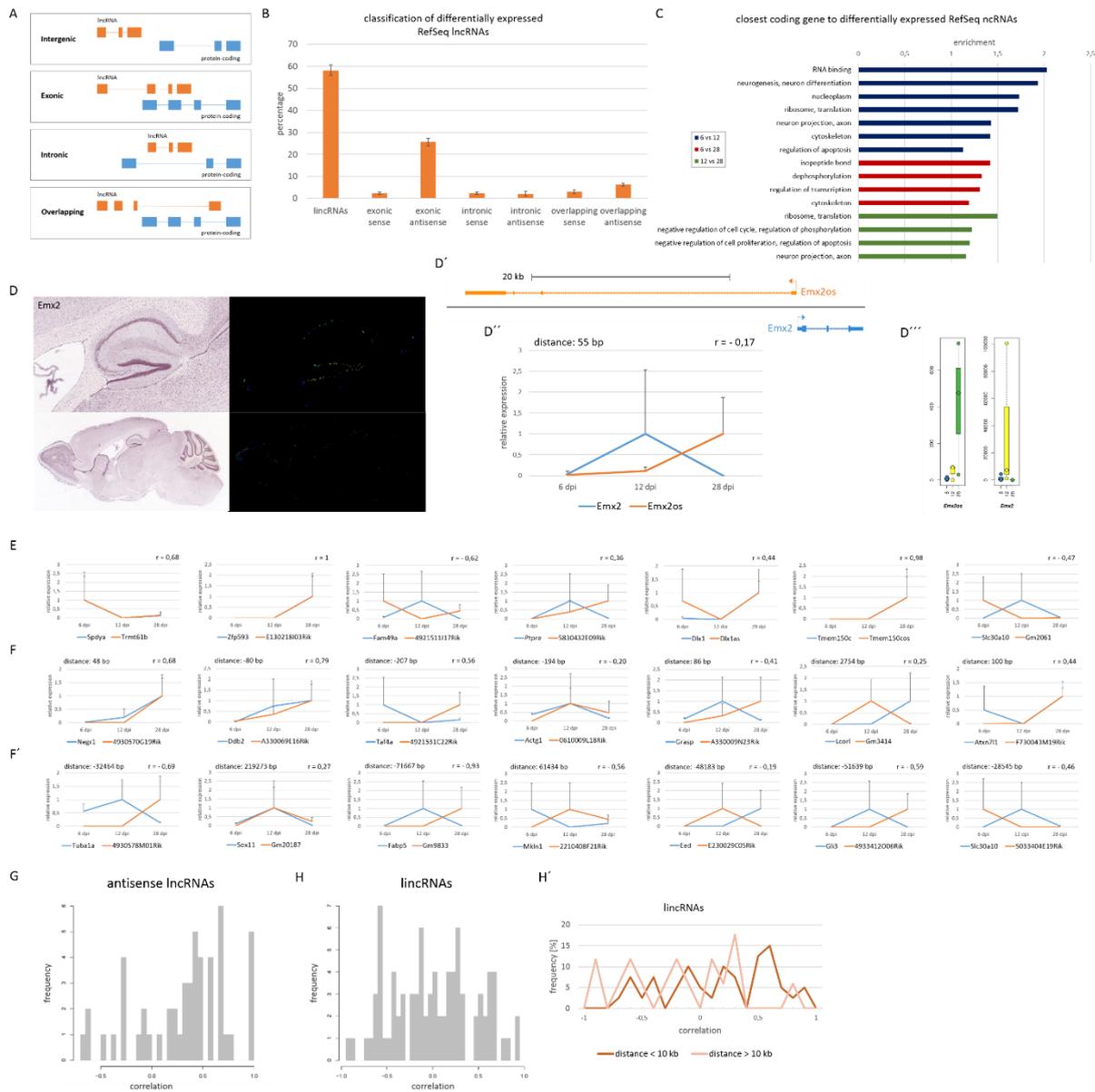


Fig 16. Dynamic expression of differentially expressed RefSeq lncRNAs.

A) Schematic illustration of different classes of lncRNAs (orange) with respect to their closest protein-coding gene (blue): intergenic, exonic, intronic, overlapping. In addition to the location, the orientation in sense or antisense can be considered. **B)** Percentages of different classes of differentially expressed RefSeq lncRNAs. **C)** DAVID-functional annotation analysis of closest protein-coding neighbouring genes to differentially expressed annotated RefSeq lncRNAs for each timepoint comparison. **D-D''')** Emx2-Emx2os gene pair. **D)** *In situ* hybridization expression data for Emx2 in the hippocampus and in the total mouse brain taken from the Allen Brain Atlas (Lein et al 2007). **D')** Genomic location of Emx2 (blue) and Emx2os (orange). Both transcripts are transcribed from a bidirectional promoter on opposite strands. **D'')** Relative expression of Emx2 (blue) and Emx2os (orange) over the three timepoints. For each transcript, the highest expression value was arbitrarily set to 1 and the relative expression calculated for the other 2 timepoints. Bars = s.d. **D''')** Absolute expression of Emx2 and Emx2os. Count expression is depicted for 6 dpi samples (blue), 12 dpi samples (yellow) and 28 dpi

samples (green). **E-F'**) Relative expression of differentially expressed RefSeq lncRNAs (orange) and their closest protein-coding neighbour (blue). LncRNA-mRNA expression pairs are shown for exonic antisense transcripts (E), lincRNAs with distances < 10 kb (F) or > 10 kb (F') to their corresponding protein-coding neighbour gene. Distances are shown for each plot at the top left and correlations at the top right. Bars = s.d. **G-H')** Distribution of Pearson correlation coefficients calculated based on the expression of differentially expressed antisense lncRNAs (G) or lincRNAs (H) and their closest protein-coding mRNA gene. H') Correlations of lincRNAs to their closest mRNA gene were plotted again for distances < 10 kb (dark red) and > 10 kb (light red).

Figure 17 displays a plot of all differentially expressed annotated RefSeq lncRNAs with a differentially expressed mRNA neighbour as pairs (Fig 17 A-A').

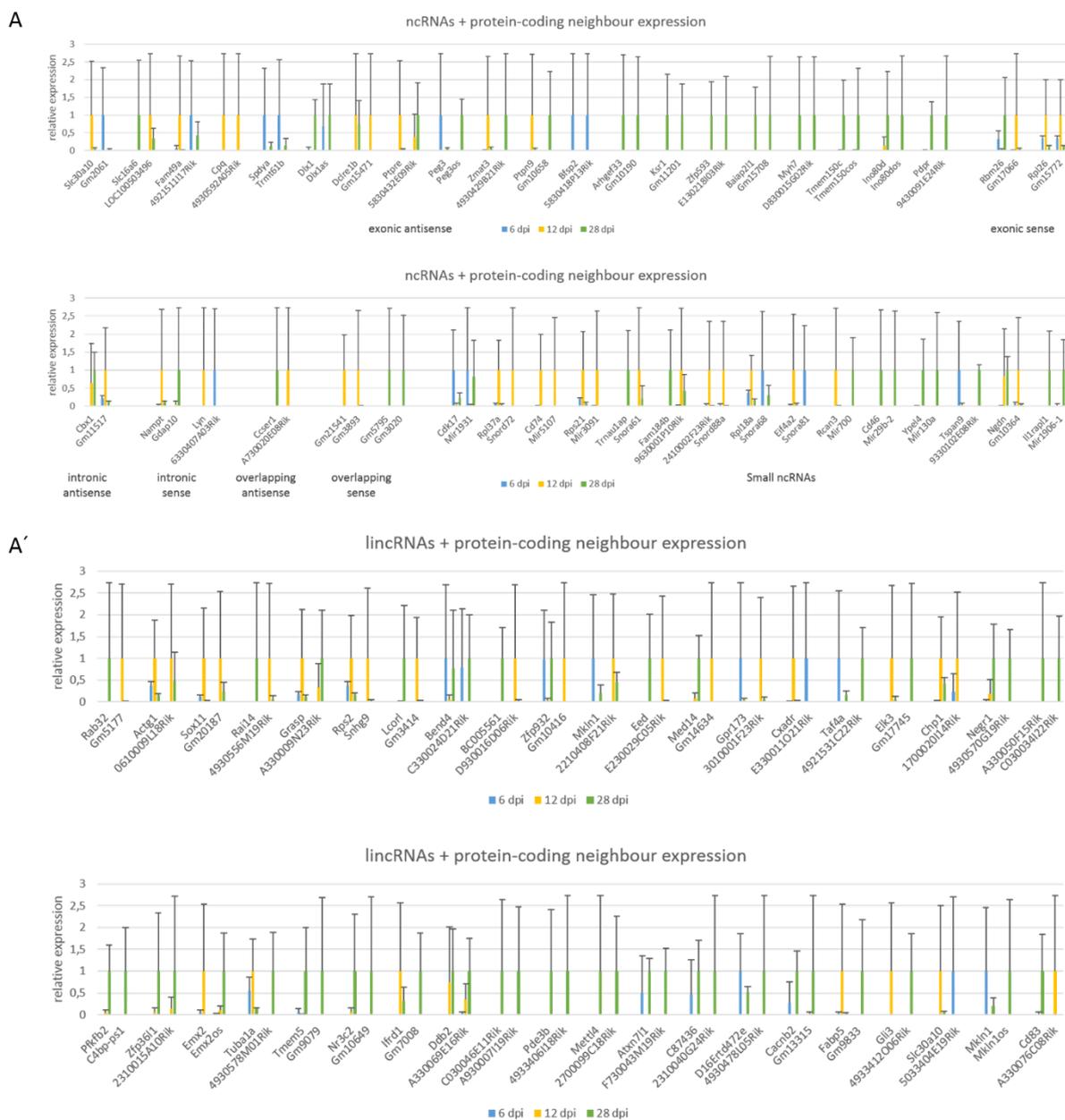


Fig 17. LncRNA-mRNA neighbouring pairs.

A-A') Relative expression of differentially expressed lncRNAs with a differentially expressed protein-coding neighbour. Relative count expression is depicted for 6 dpi samples (blue), 12 dpi samples (yellow) and 28 dpi samples (green). Expression pairs are shown for different genic lncRNA classes (exonic antisense, exonic sense, intronic antisense, intronic sense, overlapping antisense, overlapping sense) and small ncRNAs (A) and intergenic lncRNAs (A'). Bars = s.d.

4.6 Antisense expression

The usage of a strand-specific protocol allowed us to gain deeper insight into the expression of antisense transcripts. Notably, among the 14136 expressed RefSeq coding genes, 4374 (30.94 %) were found to have expression on the antisense strand as well (Fig 18A), a fact that is in accordance with earlier estimates using Direct RNA Sequencing (Ozsolak et al 2010). Gene ontology (GO) analysis revealed that genes having both sense and antisense expression are enriched for the categories of splicing, transport (protein transport, vesicle-mediated transport, nuclear pore complex), RNA binding, neuron projection development (axon guidance, dendrite, synapse), zinc finger binding, cytoskeleton, chromatin organization and modification, transcription, metabolic processes, apoptosis, cell cycle, signaling (MAPK signaling, Neurotrophin signaling) and small GTPase mediated signal transduction (Fig 18B). Notably, several of these GO terms were found to be enriched also at 6 dpi, 12 dpi and 28 dpi (Fig 12C), suggesting that sense and antisense expression are tightly regulated during adult hippocampal neurogenesis.

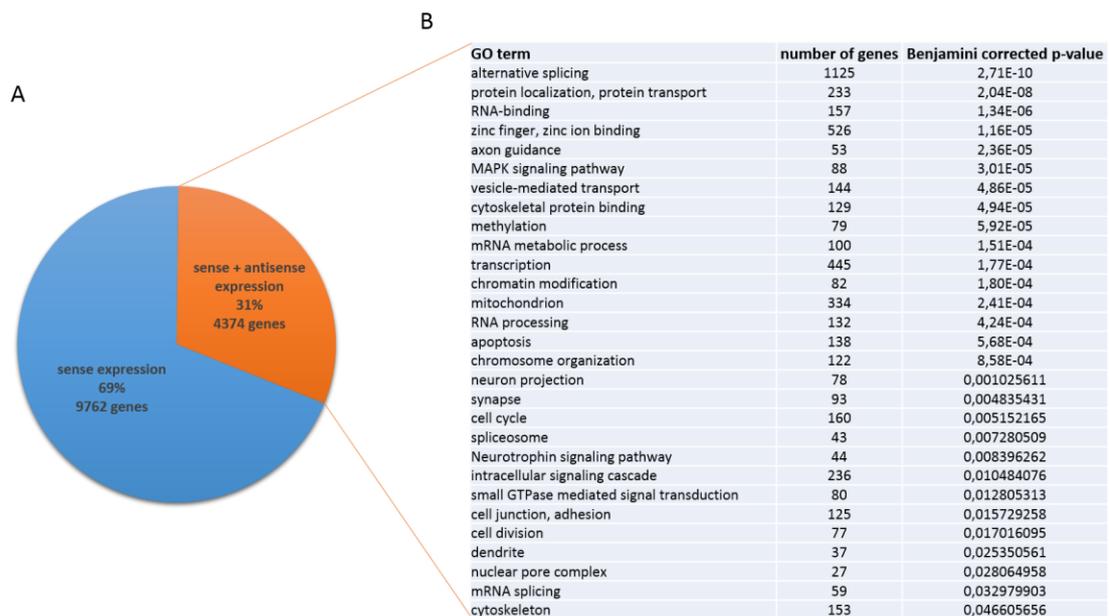


Fig 18. Antisense transcripts.

A) Number and percentage of RefSeq coding genes having expression on the sense strand only (blue) or expression on both the sense and antisense strand (orange). **B)** DAVID-functional annotation analysis of enriched GO categories of genes having both sense and antisense expression. P-values are Benjamini corrected for multiple comparisons.

4.7 Novel lincRNA transcripts

The present analysis identified an additional 21229 novel transcripts that did not map to annotated RefSeq genes and fell into the Cufflinks class of “unknown intergenic transcripts”. Looking into the NONCODE v4.0 database (Xie et al 2014), entries for 4016 transcripts (18.92 %) were found.

Besides their relatively low expression (Fig 9C), these 21229 novel intergenic transcripts displayed low exon number and a relatively short transcript length compared to known RefSeq protein-coding genes (Table 6 and Fig 19A). While known protein-coding transcripts have 6 exons on average, most exonic antisense and lincRNAs are single-exon transcripts. Among the 21229 novel transcripts, only 892 (4.2 %) were found to be multi-exonic. The lower number of exons goes along with an overall short transcript length. While for known protein-coding RefSeq genes the median length was 3905 nt, for novel transcripts it was only 315 nt (Wilcoxon $p < 2.2 \times 10^{-16}$) (Table 6 and Fig 19A). My findings regarding relatively short

transcript length, exon number, and low expression level are consistent with recent descriptions of characteristics of mouse and human lncRNA (Zhang et al 2014). The lower number in transcript length and exon number compared to Derrien et al (2012) and Zhang et al (2014) (Table 6) could be also affected by the incomplete gene-body coverage with a strong 3' end bias (Fig 9A), but collectively it shows that novel lncRNAs are shorter in length and have a lower exon number. Consistently, the NONCODE database (www.noncode.org/), a database consisting of 56018 and 46475 literature documented lncRNA genes for human and mouse, respectively, documented that the majority of these lncRNAs are 200 – 1000 nt in length with the majority consisting of 2 and 1 exon for human and mouse, respectively (Xie et al 2014).

In order to gain some possible functional insights into these novel lncRNAs, differential expression between timepoints was assessed, and the closest protein-coding gene to these differentially expressed novel transcripts was examined. Among those 21229 novel transcripts, 3087 (14.54 %) were found to be differentially expressed between timepoints.

Similar to annotated lncRNAs, a trend towards more positive correlations (mean correlation: 0.177) with their closest protein-coding gene was found (Fig 19B), irrespective of the distance between the two genes, suggesting that lncRNA-mRNA neighbouring genes are co-regulated (Cabili et al 2011, Hu et al 2014, Zhang et al 2014). Intriguingly, protein-coding neighbours to these differentially expressed novel lncRNAs were found to be enriched for GO categories involved in neurogenesis (neuronal differentiation, neuron projection development, cell fate specification), neural morphogenesis (dorsal/ventral and anterior/posterior pattern formation, neural tube development, embryonic morphogenesis, sensory organ development, cell adhesion, EGF), neuronal function (synaptic transmission, antiporter activity), indicating that these differentially expressed novel transcripts may indeed be relevant players in the regulation of neuronal maturation (Fig 19B').

In addition, I compared the present dataset of novel lncRNAs with the dataset of 8992 lncRNAs derived from the SVZ (Ramos et al 2013). 1665 lncRNAs (18.5 %) were identified also in our dataset, with 337 lncRNAs being differentially expressed between timepoints.

Similar to annotated and novel lncRNAs, a strong trend towards positive correlations between the 337 differentially expressed lncRNAs with their closest protein-coding gene was found (mean correlation: 0.311) (Fig 19C). GO analysis revealed that these neighbouring genes contribute to processes related to neurogenesis (regulation of cell proliferation, cell fate commitment, neuron development, forebrain development, synapse), and regulation of transcription (Fig 19C').

Similar to annotated lncRNAs, the finding that differentially expressed lncRNAs are co-regulated with adjacent protein-coding genes, which are themselves implicated in fundamental processes regulating neurogenesis and differentiation, makes it conceivable that these lncRNAs fulfill a function in the regulation of adult neurogenesis.

By analyzing the expression of some of these differentially expressed lncRNAs, one gene immediately attracted our attention: TCONS_00007941, also known as Gm26735 in Ensembl. It is located close to the transcription factor Sox4, transcribed convergently ("tail-to-tail") on the opposite strand and overlapping parts of the 3' UTR of Sox4 (Fig 19D').

Our group has shown recently that Sox4 and Sox11 are required for neuronal differentiation in adult hippocampal neurogenesis (Mu et al 2012). Identifying a specifically expressed gene adjacent to Sox4 tempted us to take a closer look to its expression.

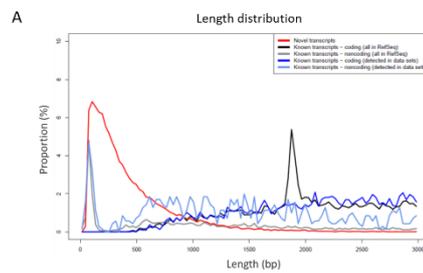
Ramos and colleagues have shown that TCONS_00007941 is specifically expressed within neural stem cells and adult-born neurons of the SVZ, while it is not expressed in niche astrocytes. Highest expression was found in transit amplifying (TA) cells of FACsorted SVZ neurons, similar as Sox4 (Ramos et al 2013). In the hippocampal SGZ, TCONS_00007941 is highest expressed at 6 dpi, decreases with time and is hardly detectable in 28 dpi neurons (Fig 19D and 19D''). The expression of Sox4 is highly correlated, also being highly expressed in immature neurons and decreased at 28 dpi, suggesting that both Sox4 and TCONS_00007941 are important regulators in adult hippocampal neurogenesis (Fig 19D and 19D''). The high level of expression and the high level of mammalian conservation further supports the functionality of this novel lncRNA (Fig 19D).

Taken together, I speculate that many of these novel transcripts may indeed be functionally involved in the regulation of adult neurogenesis.

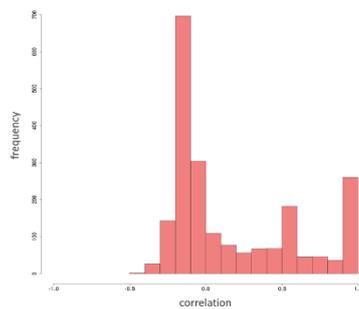
Table 6
Length distribution and exon number of mouse and human novel lncRNAs and annotated mRNAs

	novel lncRNA	RefSeq coding
median length [nt]*	315	3905
mean length [nt] (Zhang et al 2014)*	550	3162
median length [nt] (Derrien et al 2012) †	592	2453
mean exon number*	1,1	6,3
mean exon number (Zhang et al 2014)*	3,7	11
median exon number (Derrien et al 2012) †	3	8

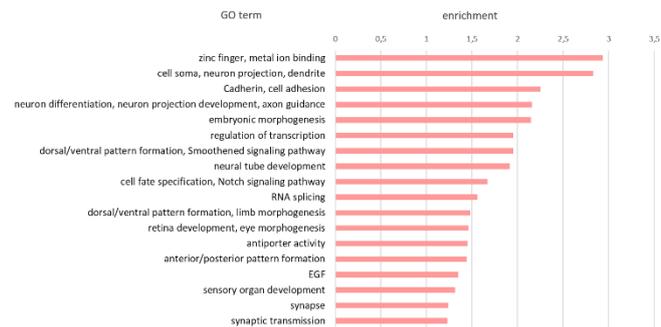
* mouse transcripts
† human transcripts



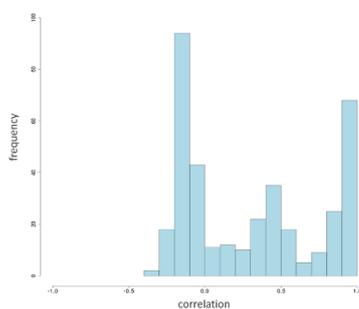
B Correlation differentially expressed novel lncRNAs - closest coding gene



B' closest coding gene to differentially expressed novel lncRNAs



C Correlation differentially expressed lncRNAs (Ramos et al 2013) - closest coding gene



C' closest coding gene to differentially expressed TCONS (Ramos et al 2013)

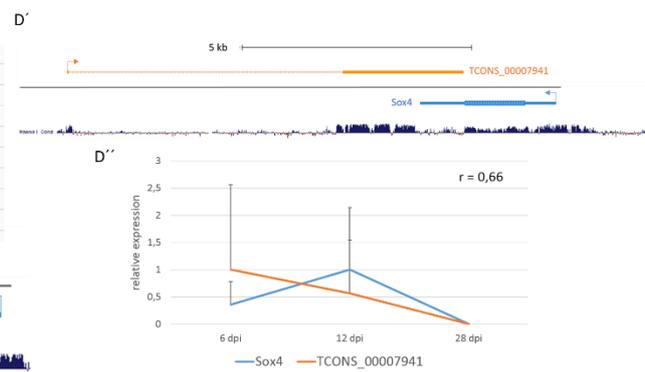
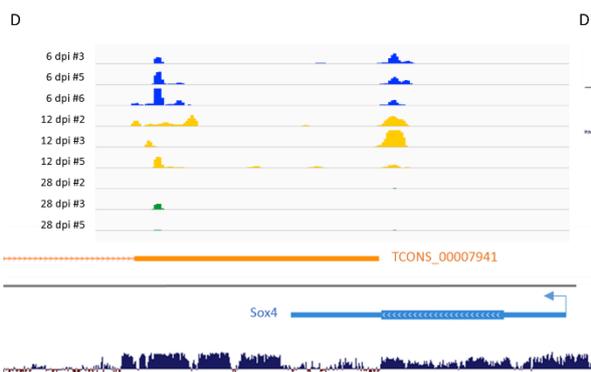
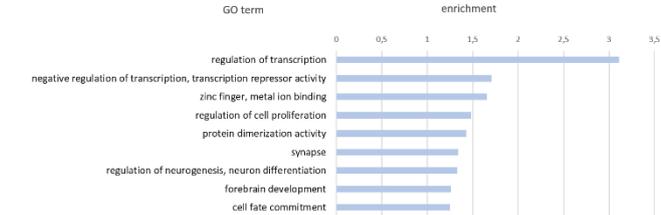


Table 6 and Fig 19. Novel lncRNAs, supposedly involved in the regulation of adult hippocampal neurogenesis.

Table 6. Comparison between novel lncRNAs and annotated mRNAs derived from this dataset and by Zhang et al 2014 and Derrien et al 2012 with regards to transcript length and exon number.

Fig 19 A) Length distribution of novel lncRNAs (red), RefSeq coding genes total (black) and RefSeq coding genes from this dataset (dark blue), and RefSeq non-coding genes total (grey) and RefSeq non-coding genes from this

dataset (light blue). **B-B')** Analysis of differentially expressed novel lncRNAs. **C-C')** Analysis of differentially expressed lncRNAs derived from the dataset by Ramos et al 2013. B and C) Distribution of Pearson correlation coefficients calculated based on the expression of differentially expressed lncRNAs and their closest protein-coding mRNA gene. B' and C') DAVID-functional annotation analysis of closest protein-coding neighbouring genes associated with differentially expressed lncRNAs. **D-D')** Expression analysis of a novel gene pair likely involved in the regulation of adult neurogenesis. D) Read coverage of individual samples along the Sox4 locus (blue) and parts of the novel gene locus TCONS_00007941 (orange). 6 dpi samples are marked in blue, 12 dpi samples in yellow, 28 dpi samples in green. The dark blue line at the bottom indicates mammalian conservation. D') Genomic location of Sox4 (blue) and TCONS_00007941 (orange). D'') Relative expression of Sox4 (blue) and TCONS_00007941 (orange) over the three timepoints. For each transcript, the highest expression value was arbitrarily set to 1 and the relative expression calculated for the other 2 timepoints. Bars = s.d. The correlation is indicated at the top right.

5 DISCUSSION

5.1 Cellular processes involved in the maturation of adult-born neurons of the adult hippocampus

RNA-Seq of retrovirally labeled neurons made it possible to examine the gene expression profile of newborn neurons within the adult hippocampus with high temporal resolution. The present study provides an in depth characterization of the dynamic changes in gene expression from early immature neuron stage right after neuronal fate determination and cell cycle exit (6 dpi), over early maturation phase (12 dpi), to late maturation phase (28 dpi) in adult hippocampal neurogenesis from the *in vivo* niche (Fig 20A). Very importantly, important cellular processes were unraveled that are involved in the proper maturation, differentiation and integration of newborn neurons during adult hippocampal neurogenesis (Fig 20A').

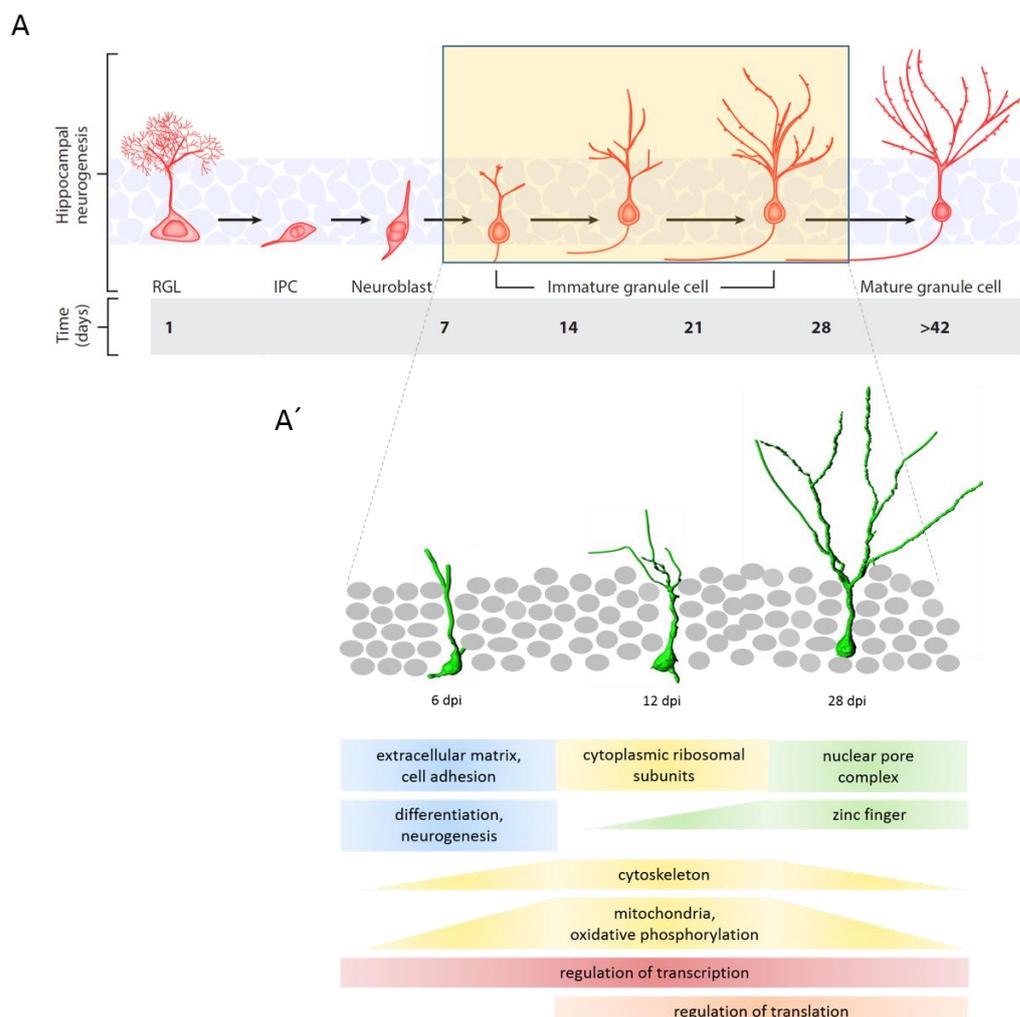


Fig 20. Processes involved in the regulation of adult hippocampal neurogenesis.

A) Summary of the developmental maturation of newborn neurons within the adult hippocampus. Newborn neurons from the immature granule cell stage have been examined (highlighted box). A') Summary of the cellular processes involved in the proper maturation, differentiation and integration of newborn neurons.

Extracellular matrix (ECM)/ cell adhesion and differentiation/ neurogenesis were among the most enriched cellular processes at 6 dpi, underlining the importance of these processes for the proper maturation during the early maturation phase. The ECM is known as a major component of the neurogenic niche, controlling stem cell maintenance, proliferation and differentiation (Faissner & Reinhard 2015). The high expression of ECM related genes in the 6dpi population implies that the neurogenic lineage itself significantly contributes to the creation of a neurogenic environment.

At 12 dpi the highest enriched GO term belonged to the category of ribosomal subunits. Strikingly, an upregulation of genes encoding ribosomal subunits has recently been demonstrated in activated neural stem cells and intermediate progenitor cells in the adult hippocampus (Shin et al 2015), as well as in the adult SVZ niche (Llorens-Bobadilla et al 2015), suggesting that the translation machinery is upregulated in newborn neurons in the adult brain. The specific expression of some ribosomal genes within adult neurogenic regions (Fig 14B) further suggests that some ribosomal genes might be involved in transcript-specific translational control, thereby regulating how transcripts are translated in space and time (Kondrashov et al 2011, Shi & Barna 2015, Xue & Barna 2012).

Cytoskeletal genes and mitochondrial genes seem to be important at all maturation stages of adult hippocampal neurogenesis, with a peak in 12 dpi neurons, most likely reflecting the rapid axonal and dendritic growth between 7 and 21 days, followed by modest growth (Sun et al 2013), going along with a higher demand of energy supply during these maturation stages. Intriguingly, an upregulation of oxidative phosphorylation-related genes was also shown in early neuroblasts of the adult SGZ (Shin et al 2015), indicating that newborn neurons primarily rely on mitochondria-dependent metabolic pathways such as oxidative phosphorylation.

Zinc finger genes have been highly enriched at 28 dpi and also in the cluster 12 + 28 dpi, suggesting that they are involved in intermediate and late phase maturation. The zinc finger

Krüppel-like transcription factor Klf9 was previously assigned a role in late-phase neuronal maturation during adult hippocampal neurogenesis, as adult-born neurons of adult Klf9-null mice exhibit impaired differentiation and decreased synaptic plasticity (Scobie et al 2009). Another zinc finger gene called Zfp365 or DBZ (DISC1-binding zinc finger protein) recently was shown to be involved in dendritic branching and spine morphogenesis of cortical pyramidal neurons (Koyama et al 2015), further indicating that some zinc finger genes are involved in proper development and integration of neurons. The high and specific expression within the hippocampus, further suggests that some zinc finger genes may be involved in the regulation of adult hippocampal neurogenesis.

Genes involved in RNA transport and belonging to the nuclear pore complex have been specifically enriched during late-phase maturation. Besides a role in nucleo-cytoplasmic transport, nucleoporins recently could have been linked to differentiation processes, as a change in the composition of the nuclear pore complex was critical for both myogenic and neuronal differentiation (D'Angelo et al 2012). Mechanistically, nucleoporins were shown to functionally interact with the genome in a dynamic manner during cell differentiation to regulate gene expression (Liang et al 2013). The simplistic view of the nuclear pore complex as static structures has also changed since the discovery of their dynamic nature, their variable composition and tissue-specific roles (Raices & D'Angelo 2012). The present data raises the interesting possibility that nuclear pore components also control the neuronal maturation and functional integration of adult-born hippocampal neurons.

The regulation of transcription was enriched at all timepoints examined, further supporting the view that neuronal maturation, differentiation and integration are tightly regulated by transcription factors (Beckervordersandforth et al 2015). Moreover, the regulation of translation was enriched during mid- and late-phase maturation, indicating that as neuronal development progresses, transcriptional regulation is complemented by post-transcriptional regulatory mechanisms. Indeed, many key postsynaptic glutamatergic genes were found to be equally expressed between different maturation timepoints. The early expression of postsynaptic genes is in stark contrast to the fact that of the three developmental time points examined, the early ones, i.e., 6 dpi and 12 dpi neurons, lack synaptic glutamatergic input (Chancey et al 2014, Espósito et al 2005, Ge et al 2006, Zhao et al 2006) and that only 28 dpi neurons display high levels of spinegenesis (Toni & Schinder

2016, Toni et al 2007, Zhao et al 2006), suggesting that the timing and activity of synaptogenesis are regulated by post-transcriptional mechanisms. Regulation through, e.g., RNA-transport and activity-dependent local translation would allow rapid temporo-spatial control of synaptogenesis by neuronal activity and thus information-dependent synaptic integration.

In addition to these processes, a number of candidates for new markers for specific phases in neurogenesis could be identified within this dataset. An interesting example is *Cacnb4*. *Cacnb4* is a voltage gated calcium channel, which was highest expressed in 28 dpi neurons (Fig 11). Notably, it was shown before that *Cacnb4* is highly expressed in the granule cell layer of the adult dentate gyrus (similar to NeuN) and its expression increases in parallel with axon outgrowth and synapse formation (Ferrández-Huertas et al 2012, Tadmouri et al 2012), suggesting an involvement of *Cacnb4* in late-phase differentiating neurons. Subcellularly, it was found to be expressed both pre- and post-synaptically in dendritic spines and shafts, which suggests an involvement in pre- and postsynaptic aspects of neurotransmission (Ferrández-Huertas et al 2012). In addition, *Cacnb4* expression was also observed in the nucleus in a differentiation-dependent progressive manner (Tadmouri et al 2012), with a predominant nuclear localization in mature neurons. Mechanistically, it was shown that upon electrical activity and Ca^{2+} entry, *Cacnb4* is targeted to the nucleus where it promotes the formation of a new nuclear complex regulating gene expression, thereby coupling neuronal excitability to transcription (Tadmouri et al 2012).

In the future, gain- and loss of function experiments should shed light on the involvement of candidate genes and processes in the proper maturation, differentiation and integration of newborn neurons in the adult hippocampus.

5.2 Long ncRNAs involved in the maturation of adult-born neurons of the adult hippocampus

Besides the identification of differentially expressed annotated RefSeq long ncRNAs (lncRNAs), the directional RNA-Seq protocol also allowed the identification of antisense transcripts and novel, not yet annotated lncRNAs that were associated with maturation, differentiation and integration of newborn neurons.

Almost 31% of expressed sense transcripts revealed an expression on the antisense strand as well (Fig 18A), which is consistent with numbers obtained by Direct RNA-Sequencing (Ozsolak et al 2010). Notably, sense coding genes with an overlapping antisense counterpart were enriched in GO terms similar to those enriched at 6 dpi, 12 dpi and 28 dpi (Fig 12C), suggesting that the interplay of sense *and* antisense transcripts may be crucial in the regulation of adult hippocampal neurogenesis.

Recent work identified thousands of novel lncRNAs associated with neuronal fate determination and early differentiation of adult neural stem cells within the SVZ niche (Ramos et al 2013). Importantly, gain- and loss-of-function analyses have identified lncRNAs as essential components of gene regulatory networks controlling early steps in neurogenesis (Ramos et al 2013). The present study complements available databases and provides a comprehensive description of lncRNAs including more than 20000 novel lncRNAs expressed during neuronal maturation in the adult hippocampus. Around 15% of novel lncRNAs showed differential expression between maturation timepoints.

5.2.1 Which novel lncRNAs are functional?

The transcriptome may not consist exclusively of functional RNAs. It has been proposed that regions with higher expression signals generally exhibit higher levels of evolutionary conservation and thus, are more likely to exert functional roles (Kellis et al 2014).

One example for a lncRNA with high expression levels and high conservation is TCONS_00007941 adjacent to Sox4. TCONS_00007941 is highly expressed (being among the 60 highest expressed genes out of 1665 expressed genes at 6 dpi) and highly conserved among mammals, thereby supporting the functionality of this novel lncRNA (Fig 19D). It is

unlikely that this transcript represents a “transcriptional byproduct” of Sox4 as the promoters for Sox4 and for TCONS_00007941 are approximately 10 kb apart from each other and no other gene expressed within 50 kb.

For the large proportion of the genome with reproducible but low expression levels and little evolutionary conservation it is challenging to predict whether their expression is of functional relevance or constitutes transcriptional noise.

LincRNAs are the lowest expressed lncRNA class in terms of expression levels according to our analysis (Fig 9C) and others (Grindberg et al 2013, Hebenstreit et al 2011). Yet, a number of lincRNAs have been found to have functional roles and tend to be the most studied class of lncRNAs (Cabili et al 2011, Guttman et al 2009, Khalil et al 2009, Rinn & Chang 2012, Sauvageau et al 2013, Ulitsky & Bartel 2013). As the functionality of lncRNAs cannot be predicted from their expression level, the functionality of each individual lncRNA needs to be tested experimentally.

It is also important to mention that low expression levels in our dataset may be the consequence of technical restraints. One transcript per cell is the minimum nonzero value possible. Values below 1 count or rpkm practically can arise from two scenarios, i) mapping artifacts (due to high sequence homology of paralogs) or ii) represent inefficiently amplified and fragmented transcripts. The reverse transcriptase is estimated to have an efficiency of only 5 - 25% and does not reverse transcribe the full transcript if it is very long (Islam et al 2012, Shapiro et al 2013). In addition, PCR amplification can introduce some bias, and RNA fragmentation before library preparation can lead to some RNA loss. Thus, RNAs that are present at low levels to begin with, thereafter inefficiently reverse transcribed and amplified, highly fragmented, and with a read coverage of the 3' end only, can result in read values that are far below 1.

Moreover, we have analyzed 30 to 100 cells per sample. If a very lowly expressed transcript was just present in one cell but absent in the others, the rpkm level is likely to be decreased. Such a scenario may be explained by high cellular heterogeneity and dynamic expression (discussed in 5.3.3).

5.2.2 Linking lncRNA to function

The identification of ncRNA function is not trivial, because in most cases it is unclear which phenotype to investigate. One frequently taken path to generate hypotheses regarding the function of a specific ncRNA is the “guilt-by-association” approach (Guttman et al 2009, Rinn & Chang 2012). This approach associates ncRNAs with biological processes based on a common expression pattern across cell types and tissues and their co-expression with protein-coding genes.

It has been shown that a number of lncRNAs are linked to the activity of neighbouring genes (Sigova et al 2013) and that the expression of lncRNAs is positively correlated with the expression of the closest protein-coding gene (Cabili et al 2011, Hu et al 2014, Zhang et al 2014). This sign of co-regulation suggests that both the mRNA as well as the adjacent lncRNA transcript are likely involved in the same functional processes. Even though these correlations do not prove that ncRNAs have a function in these processes, they do provide a hypothesis for targeted gain- or loss-of-function experiments.

In the present study, differentially expressed RefSeq annotated lncRNAs as well as novel differentially expressed lncRNAs displayed positive correlations with the expression of protein-coding neighbouring genes (Fig 16G, 16H', 19B, 19C), suggesting that lncRNAs are co-regulated with protein-coding neighbouring genes. As a word of caution, it was noticed in several studies that the transcription of neighbouring mRNA gene pairs is also slightly positively correlated (Cabili et al 2011, Zhang et al 2014), questioning whether lncRNAs indeed have an effect in *cis* or whether positive coordinated transcription between coding RNAs and lncRNAs is merely a result of proximal transcriptional activity in the surrounding open chromatin (Ebisuya et al 2008). Still, the correlation between neighbouring lncRNA:mRNA gene pairs is higher than for mRNA:mRNA gene pairs (Zhang et al 2014). Moreover, the highly specific expression of lncRNAs within tissues, cells and developmental stages - a hallmark of lncRNAs - argues against being a noisy byproduct of mRNA expression (Gloss & Dinger 2016).

GO analysis of the closest protein-coding neighbor to differentially expressed lncRNAs showed a striking enrichment in functional categories directly relevant to neuronal maturation processes including neuronal differentiation, cell fate specification, neuron

projection development, dorsal/ventral and anterior/posterior pattern formation, synapse and synaptic transmission (Fig 19B' and 19C').

Thus, even though many novel transcripts are lowly expressed, the specific regulation of their expression together with specific functions in neurogenesis of their protein-coding neighbouring genes implies that they may serve functional roles in newborn hippocampal granule neurons. In conclusion, the data of the present study provides a rich database to identify candidate lncRNAs critical for the maturation and synaptic integration of neurons.

5.3 Strengths, limitations and perspectives

The present directional single-cell RNA-Seq protocol allowed a comprehensive description of both coding and non-coding transcripts in a directional manner from a limited number of cells (< 100 cells) isolated *in vivo*. While existing directional single-cell RNA-Seq protocols were not able to describe lowly expressed transcripts or provide a comprehensive description of non-coding transcripts (Hashimshony et al 2012, Islam et al 2011), these limitations have been overcome by a higher sequencing depth and an amplification strategy with a better gene body coverage (as compared to sequencing the very 3' (Hashimshony et al 2012) or 5' end (Islam et al 2012, Islam et al 2011) only).

There are several lines of evidence that both the qualitative and quantitative genomic characteristics have been preserved by this amplification protocol.

First, the overall reads distribution among different genomic classes has been roughly the same as in non-amplified samples from large number of cells. In poly-A⁺ samples of mouse cerebral bulk cells 59%, 15% and 23% mapped to exonic, intronic and intergenic regions, respectively (Cui et al 2010). In our dataset of polyadenylated transcripts 62%, 14% and 17% of reads mapped to exonic, intronic and intergenic regions (Table 5, Fig 9B). The remaining 7% of reads mapped antisense to annotated regions. Similar numbers of antisense reads have been obtained by direct RNA-Sequencing in human liver (Ozsolak et al 2010).

Second, strand-specific assignment has not been distorted by amplification, as more than 94% of reads could be specifically assigned to either strand. Moreover, the percentage of

antisense transcripts overlapping sense coding genes is highly similar to the percentage of antisense transcripts identified by Direct RNA-Sequencing, which works without cDNA synthesis or amplification (Ozsolak et al 2010). In human liver, 30.2 % of annotated transcripts were found to have antisense transcription (Ozsolak et al 2010), while we identified 30.9 % of annotated genes having antisense transcription (Fig 18A), thus strengthening the directionality of this protocol even after amplification. Overall, due to the low expression of antisense transcripts, the total number of antisense reads was shown to be rather small, only comprising 8% of all reads (Ozsolak et al 2010). In our dataset, 6.76% of all reads were antisense reads (Table 5, Fig 9B), again demonstrating that the overall distribution of reads is not distorted by this directional amplification protocol.

Third, also the quantitative expression has been preserved after amplification, as the expression levels of different genomic classes are very similar to the numbers provided by the ENCODE consortium (Kellis et al 2014) and single-cell studies (Marinov et al 2014, Suzuki et al 2015), with the peak expression of protein-coding genes being between 1 and 50 rpkm (Fig 9C). The ENCODE consortium has plotted all expressed human genomic regions dependent on the expression level. Overall, in the human genome 70% of all polyadenylated transcripts are expressed below 1 transcript per cell and 50% below 0.1 fpkm (Kellis et al 2014). Consistently, in the present dataset 68% are expressed below 1 rpkm and 48% below 0.1 rpkm (Fig 9C').

5.3.1 Sensitivity, expression levels and sequencing depth

The sequencing depth and the number of biological replicates have to be considered carefully at the beginning of an experiment (Conesa et al 2016, Kukurba & Montgomery 2015). While a very high sequencing depth allows for the detection of lowly expressed transcripts and rare splice events, the inclusion of more biological replicates per condition (e.g., by multiplexing) allows more accurate estimates of biological variability and contributes to a more robust analysis of differential expression (Ching et al 2014, Liu et al 2014b, Sims et al 2014).

If the goal is to analyze coding genes only, most cells can be detected reliably using a relatively small sequencing depth, thus high multiplexing is favourable. It has been estimated that a sequencing depth of 2 million reads is sufficient, as 90% of genes detected at 30 million reads have been already detected at a sequencing depth of 2 million reads (Wu et al 2014). Still, it is recommended to sequence 30-40 million reads in order to accurately quantify transcripts of moderate to high abundance (Kukurba & Montgomery 2015, Zeng & Mortazavi 2012). Caution is recommended using very high multiplexing as it goes at the expense of sequencing depth and sensitivity. In an extreme example of multiplexing, 1536 cells have been multiplexed in one sequencing lane, a method called massively parallel single-cell RNA-Seq (MARS-Seq) (Jaitin et al 2014). By multiplexing 1536 single cells from mouse spleen, on average 22,000 reads could be aligned per cell, leading to the identification of 200 to 1500 distinct RNA molecules from each cell (Jaitin et al 2014). Knowing that thousands of genes are expressed in a single cell (Marinov et al 2014), this information provides only a first glimpse that can be used to identify different subtypes but can not describe the qualitative transcriptome of single cells.

If the goal is to detect also lowly expressed non-coding RNAs, a higher sequencing depth is a must. It was shown that in human H1 embryonic stem cells, 36 million reads are sufficient to accurately quantify 80% of transcripts that are expressed at > 10 FPKM. However, genes that are expressed at low levels (i.e., fewer than 10 FPKM) could only be accurately quantified with 80 million mapped reads (ENCODE 2011). Thus, if also low abundant transcripts like lncRNAs should be accurately quantified, samples should be sequenced at high depth or RNA-Capture Seq techniques should be used to enrich for low-abundant

transcripts (Clark et al 2015, Fu et al 2014, Mamanova et al 2010, Mercer et al 2011, Ramos et al 2013). For new transcript discovery it is recommended to sequence at least 100 million reads per sample (Zeng & Mortazavi 2012). In agreement with this number I found that the error rate of reads gets saturated at approximately 100 million reads (Fig 8B'). Thus, the reliable detection of lowly expressed transcripts is only allowed above 100 million reads.

In the present study using very deep sequencing (> 100 million mapped reads), 8000 – 11000 coding genes could be identified (Fig 10C). In addition, several thousand non-coding transcripts of very low abundance could be detected (Fig 9C and 9C').

Other single-cell amplification methods, ending up with approximately 1 million reads per sample, were able to detect 5000 – 6000 coding genes and did not provide information on non-coding RNA transcripts (Hashimshony et al 2012, Nakamura et al 2015, Zeisel et al 2015). It was shown that such high multiplexing can only reliably detect highly and medium-expressed genes of more than 10 copies per cell (Hashimshony et al 2012, Islam et al 2011, Nakamura et al 2015). Given that 50 % of protein-coding genes are expressed below 5 fpkm (Marinov et al 2014) (Fig 9C), it is obvious that many single-cell amplification methods combined with low sequencing do not only miss a large fraction of protein-coding transcripts, but also most non-protein-coding transcripts. The ENCODE consortium has plotted all expressed human genomic regions dependent on the expression level and found that among 75% of transcribed genomic regions, 50% are expressed below 0.1 fpkm, and 70% of polyadenylated transcripts are expressed below 1 transcript per cell (Kellis et al 2014). Consistent with these numbers, I found that 48.2 % of polyadenylated transcripts from this dataset are expressed below 0.1 rpkm and 68.1% below 1 rpkm.

Collectively, these numbers demonstrate that a high sequencing depth is needed in order to be able to detect also lowly expressed transcripts. Moreover, these numbers show that the present amplification protocol from a limited number of cells combined with high sequencing depth is capable of detecting lowly expressed transcripts at similar levels as sequencing datasets from high cell samples.

5.3.2 Why are more genes detected at 28 dpi compared to 6 and 12 dpi samples?

While 6 and 12 dpi samples express more than 8000 RefSeq genes (8680 vs 8218, respectively), the number is strikingly higher in 28 dpi samples (11682 detected RefSeq coding genes) (Fig 10C and 10C'). Since all the samples were amplified in parallel, this difference unlikely represents technical variability. One explanation could simply be the size of the cell soma. While the average square size of LCM isolated 6 dpi immature dentate granule cells was 75 – 80 μm^2 , it increased to 80 – 85 μm^2 at 12 dpi and reached around 100 μm^2 at 28 dpi. Consistent with the knowledge that the cellular numbers of rRNAs, mRNAs and ribosomes scale with the cell size in order to maintain the appropriate concentrations within a cell (Marguerat & Bähler 2012), it could likely be that the increased number of expressed genes can be partly explained by the increased soma size due to differentiation. Still, this would only explain why certain genes are expressed at higher levels, but not why the total amount of different expressed genes is higher.

Another possible explanation could be that cells express more genes with increasing complexity and that the synaptic connection with other cells enhances the transcription machinery (Kandel 2001). It has been shown that single neurons from the mouse somatosensory cortex and hippocampal CA1 region contain more RNA than glia and vascular cells and a larger number of detectable genes (Zeisel et al 2015), indicating that synaptically connected neurons express more genes than other cell types. Only 28 dpi neurons (but not 6 or 12 dpi neurons) are synaptically integrated, and 4-8 week old adult-born neurons exhibit a lower threshold and larger amplitude for LTP (Ge et al 2007, Mongiat & Schinder 2011, Schmidt-Hieber et al 2004), raising the possibility that the larger synaptic input enforces the transcription of genes.

5.3.3 Reasons for sample heterogeneity

Several studies have pointed out high levels of heterogeneity in single-cell genomics (Johnson et al 2015, Kim et al 2015a, Miyamoto et al 2015, Patel et al 2014, Pollen et al 2014, Qiu et al 2012, Shalek et al 2013, Wilson et al 2015). The present dataset shows that high levels of heterogeneity also can be seen even in cell sample pools of 30 to 100 cells (Fig 10A

and 10B). The reasons for this noise are both i) technical, owing to minute amounts of input material, and ii) biological, for example, owing to inhomogeneity of adult-born neurons (Brunner et al 2014, Piatti et al 2011, Sun et al 2015b) or short bursts of RNA transcription (Cai et al 2008, Raj et al 2006).

Given that the activation rate of quiescent RGL cells decreases gradually along the septo-temporal axis of the adult mouse dentate gyrus (Sun et al 2015b), and that the septal region of the hippocampal dentate gyrus displays higher levels of basal network activity, and newborn dentate granule neurons within this region mature faster than cells in the temporal region (Piatti et al 2011), it could likely be that newborn neurons from different regions (septal vs. temporal) express different marker genes, both due to regional differences and network-specific differences. Indeed, it was suggested recently that newborn neurons along the septo-temporal axis are specialized for distinct mnemonic and mood-related behavioral tasks (Tanti & Belzung 2013, Wu et al 2015). Since in the present study the visualization and isolation of newborn neurons was dependent on the injection site, it can-not be guaranteed that samples contained an equal amount of cells derived from both parts throughout the hippocampus. Thus, some heterogeneity between samples from the same timepoint may have arise from this biological region-specific difference.

Moreover, adult neurogenesis is a highly plastic and activity-dependent process. Each behavioral change of an individual (e.g., running, learning, depression) is translated into changes at the systems and network level, which in turn affects the local circuitry and signaling systems that affect the regulation of gene expression during each step of maturation (Kempermann et al 2015). Thus, mice with different behavioral outputs (physical or cognitive) are likely to show considerable changes in gene expression. In addition, it was shown recently that adult-born granule cells from the same developmental stage can contribute to the network with fundamentally different functions (Brunner et al 2014), thereby suggesting high levels of functional heterogeneity within adult-born granule neurons from the same developmental stage.

Measurements using fluorescence in situ hybridization (FISH) have indicated that levels of specific transcripts can vary as much as 1000-fold between presumably equivalent cells (Raj et al 2006) and that the chromosome structure of individual cells is highly variable as shown by single cell Hi-C (Nagano et al 2013). By using transgenic cell lines expressing a short-lived

luciferase protein from an unstable mRNA, it was possible to record bioluminescence in real time in single cells and thus, study the bursting kinetics of different genes (Suter et al 2011a). Interestingly, bursting kinetics are highly gene-specific (Suter et al 2011a), and may be a general control strategy to coordinate multi-gene responses to external stimuli (Cai et al 2008). A potential advantage of producing a high number of mRNAs within a short time period is to provide both a fast response to a stimulus and a large dynamic range (Suter et al 2011b). Indeed, the stochastic bursting was shown to be important for the priming of multipotent progenitor cells in cell fate decision (Chang et al 2008). However, it has to be noted that most of these transcriptional bursts are buffered at the protein level due to the high stability of mRNAs and proteins in higher eukaryotes (Suter et al 2011b).

Bioinformatic techniques for assembling and analyzing single-cell genome data are also rapidly advancing (Stegle et al 2015), and several noise models have been suggested to correct for technical variability in single-cell RNA-Seq (Brennecke et al 2013, Grun et al 2014, Kim et al 2015b). Very importantly, it has been proposed to use spike-in quantification standards of known abundance in order to explicitly calculate the true gene expression levels and thus to reduce technical noise resulting from small amounts of input material. Moreover, they can provide insight into the relative amplification efficiency and detection limits (Ding et al 2015, Islam et al 2014, Kivioja et al 2012, Lovén et al 2012). By using spike-in RNA and RNA extracted from bulk cells (pool-and-split control) it is possible to calculate a conversion factor to adjust for technical variability (Grun et al 2014). Moreover, specific differential gene expression models have been developed for single-cell RNA-Seq studies that account for single-cell specific noise, such as dropout events and amplification biases (Kharchenko et al 2014). Despite this rapidly advancing progress, still several computational and analytical challenges need to be overcome (Stegle et al 2015). Moreover, new computational tools are being developed for the assessment of cell state hierarchies (Juliá et al 2015) and new tools are needed to decipher complex networks in single cells in order to be able to understand the biology behind the complex dataset generated by RNA-Seq of a low number of cells.

Conclusively, the study of a large number of single-cells is necessary in order to be able to account for both technical and biological heterogeneity. Calculations suggest that even

hundreds or thousands of single cells have to be analysed to answer targeted questions in single tissues (Shapiro et al 2013).

5.3.3.1 Future strategy to study transcriptional dynamics during adult neurogenesis

The present dataset already showed that there are remarkable differences between the differentiation stages from 6 to 12 and from 12 to 28 dpi, however, it does not provide information on transcriptional dynamics and how many different subtypes of cells are generated both in a temporal but also in a region-specific manner (septal vs. temporal).

To overcome this limitation, new unbiased clustering methods (Shin et al 2015, Trapnell et al 2014) and self-organizing maps (Kim et al 2015a) have been established recently that allow the pseudotemporal ordering of a mixture of single cells in an unbiased manner. The advantage of this method is that it can be applied on any kind of tissue harbouring different cell types (Moignard et al 2015, Shin et al 2015, Trapnell et al 2014, Treutlein et al 2014, Usoskin et al 2015, Zeisel et al 2015). Since these clustering approaches do cluster subtypes of genes without *a priori* selection based on distinct cellular markers, which are often not available, this approach also resolves the problem of isolating highly specific cell types, e.g., immature neurons at specific maturation stages. The only requirement is that a sufficient number of single cells is being analyzed in order to generate and identify reliable clusters.

Recently, a new bioinformatic approach named “Waterfall” has been used successfully to reconstruct neuronal stem cell dynamics with unprecedented temporal resolution (Shin et al 2015), thereby identifying 5 different subclusters during the transition from Nestin-positive activated neural stem cells to Nestin-positive intermediate progenitor cells within the adult hippocampal neurogenesis niche.

A similar clustering approach could be applied to resolve the identity of neuronal subtypes during later stages of adult neurogenesis. In a first step, immature neuronal cells can be pre-defined and isolated from transgenic mouse models expressing an immature neuronal marker under the control of a fluorescent reporter, e.g., Tbr2 (Hodge et al 2008, Hodge et al 2012), NeuroD1 (Gao et al 2009) or DCX (Couillard-Despres et al 2006), in order to study transitional stages from fate determination over differentiation to final maturation stages.

The isolation of DCX-positive neurons from the adult hippocampus via FACS was already successfully applied (Fig 4A) and is easier and faster than highly laborious retroviral labeling and LCM isolation techniques. In a second step, transcriptome profiling followed by subsequent clustering will generate different neuronal subclusters both in a temporal but also in a region-specific manner.

Knowing how many and which different subtypes are involved in the generation of adult-born granule neurons from neural stem cells to intermediate progenitor cells over immature granule neurons and finally functionally integrated mature neurons, will change the current classification schemes and revolutionize the characterization of neuronal subtypes during the course of adult hippocampal neurogenesis.

5.3.4 Limitations of the protocol

Although this amplification method has overcome the limitation of strandness and allows an in depth-characterization of even lowly abundant transcripts like lncRNAs, several issues need to be overcome in the future.

First, currently the amplification has a 3' end bias (Fig 9A) (amplified PCR products being ~ 3 kb, but IVT products being only 0.5 to 1 kb), which does not allow the reliable analysis of splice isoforms. Methods that allow for the usage of random primers for reverse transcription will make it possible to recover full-length cDNAs. However, for that, rRNA that makes up approximately 90 % of all RNA within a cell needs to be removed before priming with random hexamers.

Second, the present protocol only detects polyadenylated transcripts because of the usage of poly(dT) primers, while excluding most of the small ncRNAs and a significant proportion of non-polyadenylated lncRNAs (Cui et al 2010, Yang et al 2011). A comparison between ribo-minus RNA-Seq (containing both polyadenylated and non-polyadenylated transcripts) (with a left-over of 10% of ribosomal sequences) and poly-A⁺ selected RNA-Seq (containing only polyadenylated transcripts) from mouse cerebrum using bulk cells (1 µg) showed that using ribo-minus RNA-Seq a higher proportion of reads, 44% and 25%, map to intergenic and intronic regions, respectively, as compared to 23% and 15% from the polyA⁺ RNA-Seq dataset (Cui et al 2010). This result indicates that only half of the intronic and intergenic

transcripts are polyadenylated and that consequently almost 50% will be missed when using polyA⁺-RNA-Seq only (Cui et al 2010). A fifty percent loss has also been estimated by the ENCODE consortium, where they analyzed the expression of annotated and novel lncRNA transcripts in 15 human cell lines after fractionation into poly(A)⁺ and poly(A)⁻ samples, thereby identifying only 44% of novel intergenic and antisense transcripts in the polyA⁺ samples (Djebali et al 2012). Moreover, it was shown that while 75% of the human genome is transcribed, about 40% generate polyadenylated transcripts (Kellis et al 2014). Collectively, these numbers imply that the detection of novel, not yet annotated lncRNAs will be incomplete and the number will likely be even twice as high.

In order to be able to analyze both polyadenylated but also non-polyadenylated transcripts, it will be important to deplete abundant rRNA and tRNA, while preserving all other RNA transcripts. After rRNA and tRNA depletion, it would be possible to use random primers for reverse transcription that will prime all other classes of RNA, including non-polyadenylated long and small ncRNAs. Currently, there are several methods available for the removal of ribosomal RNA, e.g., probe-based approaches that are directed against rRNA species such as Ribo-Minus (Life Technologies), Ribo-Zero (Illumina), selective depletion of abundant RNA (Morlan et al 2012) or duplex-specific nuclease (DSN) treatment (Yi et al 2011), however, all these methods require a minimum of 50 – 100 ng RNA as input. It would be worth testing whether these methods work also after amplification of RNA. In such a case, genomic DNA needs to be removed before cDNA synthesis, which has been already tested to not to influence the cDNA amplification in single cells (data not shown). After priming with random primers and cDNA amplification, rRNA species could be removed by probe-based approaches or DSN treatment. It further needs to be tested i) whether rRNA depletion after amplification is efficient and ii) whether the amplification of excessive amounts of rRNA does not influence the amplification of lowly abundant transcripts or distort relative abundances.

Third, the Epicentre directional library preparation technique permits slightly higher directional sequence reads (Pease & Sooknanan 2012), i.e., >98%, while the present directional amplification followed by Epicentre directional library preparation resulted in 94% directional sequence reads. Some misdirection may have been introduced at the cDNA synthesis step due to self-priming of the reverse transcriptase (Beiter et al 2007, Perocchi et

al 2007, Tzadok et al 2013) or during the addition of the universal primer (UP) sequences before PCR amplification. Since both UP1 and UP2 primers are oligo-dT primers, an inefficient degradation of UP1 by exonuclease treatment might have led to the aberrant addition of UP1 to the 5' end. Such confounding factor might be avoided in the future by adding for example dCTPs (instead of dATPs) to the 5' end by Terminal deoxynucleotidyl transferase (TdT), which can then be primed by an oligo-dG primer. Consequently, the oligo-dT UP1 primer could not bind to the 5' end, even after inefficient elimination.

In order to reduce self-priming artefacts it is either recommended to increase RT temperature (Haddad et al 2007, Moison et al 2011, Tzadok et al 2013) or to add substances that promote the unfolding of any RNA secondary structure and the synthesis of longer cDNA products, such as Betaine (Picelli et al 2014), Actinomycin D (Perocchi et al 2007) or Periodate (Tzadok et al 2013).

Fourth, improvements by Quartz-Seq (Sasagawa et al 2013) should be implemented in the present amplification protocol in order to eliminate nonspecific products during PCR amplification and to reduce sample loss during gel purification. Sasagawa and colleagues were able to reduce the production of byproducts by the usage of a minimum primer concentration (42 pmol/L vs 12.5 nM/L for first strand cDNA synthesis) and suppression PCR. In suppression PCR, unreacted, complementary primer sequences bind to each other, so the PCR primer can not bind to the template DNA and is not being amplified. Suppression PCR is very effective in suppressing amplification of small size DNA that contains complementary sequences. Moreover, the addition of a suppression primer effectively suppressed the synthesis of byproducts (Sasagawa et al 2013). Ultimately, gel purification is not needed anymore.

Fifth, it has been proposed to use spike-in quantification standards of known abundance in order to explicitly calculate the true gene expression levels and thus to reduce technical noise resulting from small amounts of input material. Moreover, they can provide insight into the relative amplification efficiency and detection limits (Ding et al 2015, Islam et al 2014, Kivioja et al 2012, Lovén et al 2012). Even though we have also used 4 different spike-

in RNAs derived from *Bacillus subtilis* at defined concentrations (Affymetrix eukaryotic poly-A RNA control kit) in order to have an internal validation of the amplification procedure, a calculation based on just 4 transcripts is not enough for normalization purposes. Currently, a set of 92 universal RNA synthetic spike-in standards generated from random unique sequences ranging 6 orders of magnitude has been developed by the External RNA Controls Consortium (ERCC) for microarray and RNA-Seq experiments (Jiang et al 2011), which are nowadays widely used for normalization in single-cell RNA-Seq experiments (Ambion ERCC RNA spike-in control mixes).

5.3.5 Future technological advancements

The present study provides important insights into the transcriptional dynamics of immature neurons during the course of maturation. However, to get a complete picture of cellular regulation, also genomic, epigenomic and proteomic aspects need to be resolved in the future.

Besides single-cell transcriptome analysis, important advancements have been made in single-cell genome-, single-cell epigenome- and single-cell proteome-analysis as well (Kolodziejczyk et al 2015, Macaulay & Voet 2014). Whole-genome amplification (WGA) that allows the genomic analysis of single cells has been developed already years ago (Dean et al 2002) and WGA kits became available (Huang et al 2015). Recently, efforts have been made in single-cell epigenomic analysis and single cell proteomic analysis as well. Reduced representation bisulfite sequencing (RRBS) (Guo et al 2013), Hi-C (Nagano et al 2013) and DNase-seq (Jin et al 2015) were developed for single-cells in order to determine DNA methylations, 3D DNA interactions and the presence of transcriptional regulatory elements in single cells respectively. Single-cell CHIP-Seq and FAIRE-seq are still awaited for the analysis of histone modifications and for profiling active regulatory DNA elements, in order to unravel the entire epigenomic landscape in single-cells (Shin et al 2014). Recently, single cell mass cytometry was developed, which allows simultaneous measurements of surface markers, internal functional proteins and their modifications in single cells (Bendall et al 2011). Combined with computational tools that allow the mapping of high dimensional cytometry data onto two dimensions (Amir et al 2013) or determining the timing and order

of key molecular and cellular events during development (Bendall et al 2014), it is possible to establish trajectories and networks of protein expression in single cells.

Notably, there have been first attempts to perform genomic and transcriptomic analysis in parallel from single cells (Dey et al 2015, Macaulay et al 2015). Just recently, a new method was developed that allows the analysis of copy-number variations, DNA methylome and transcriptome from an individual mammalian cell simultaneously (Hou et al 2016). In the future, it will be gold standard to integrate single-cell genomic, epigenomic, transcriptomic and proteomic analysis simultaneously of an individual cell, in order to get a complete picture of functional states of individual cells (Kolodziejczyk et al 2015, Liu et al 2014a, Shapiro et al 2013).

Moreover, the development of highly sensitive third-generation sequencing (TGS) technologies allow the direct sequencing of DNA or RNA molecules without the need of amplification or reverse transcription. This avoids the introduction of artefacts and biases from PCR or reverse transcription steps, and will allow direct RNA sequencing of single cells, ultimately allowing direct counting of intact transcripts (Ozsolak et al 2009). Until now, RNA input requirements are still not down to single cells (approximately 300 pg polyadenylated RNA) and sequencing error rates are still quite high (approximately 4 %), but TGS technologies hold promise for such radical improvements in the near future (reviewed in Morey et al 2013, Ozsolak 2012, Ozsolak & Milos 2011a, Ozsolak & Milos 2011b, Pickrell et al 2012, Sanchez-Flores & Abreu-Goodger 2014).

Future studies that couple technological advances in experimental preparation and sequencing with new computational approaches will enable the exact definition of cell types and even cell states, the reconstruction of intracellular circuits and will thus, revolutionize our understanding of whole organisms, and have important clinical implications for the development of new therapies in disease.

5.3.6 Outlook and significance

Adult neurogenesis confers a unique mode of plasticity to the mature mammalian brain. The precise control of maturation of new neurons is central to hippocampal plasticity and there is growing evidence that dysregulation of the maturation process significantly contributes to the pathophysiology of neuropsychiatric and cognitive diseases (reviewed in Apple et al 2016, Kang et al 2016). Loss of adult hippocampal neurogenesis is associated with hippocampal learning and memory deficits and increased anxiety behavior (reviewed in Abrous & Wojtowicz 2015). Moreover, analysis of rodent models demonstrated a) that age-associated cognitive deficits correlate with the rate of decline in hippocampal neurogenesis; b) that adult neurogenesis is perturbed in animal models for neurodegenerative diseases such as Parkinson's Disease, and Alzheimer's Disease, and c) that the action of certain classes of antidepressants requires intact hippocampal neurogenesis (reviewed in Apple et al 2016, Kang et al 2016, Winner & Winkler 2015, Zhao et al 2008).

One overarching goal of adult neurogenesis research is to use neural stem cells and their progeny therapeutically as a regenerative source for neural repair of neurodegenerative and cognitive diseases. Understanding the molecular mechanisms which are essential for the maintenance of proper hippocampal function is the prerequisite to influence these mechanisms and thereby prevent or slow down age-related or disease-associated cognitive decline.

The present study generated important knowledge on the genetic networks controlling the development of new neurons in their natural environment and provided new insight into the function of lncRNAs in stem cell regulation. Such knowledge is expected to i) support the identification of mechanisms contributing to the pathogenesis of neurodevelopmental and neuropsychiatric disorders and ii) promote the development of strategies that target hippocampal neurogenesis for repair of the diseased nervous system.

Finally, the methodology developed will be highly valuable to decipher complex genetic programs associated with cellular function, development and disease of many biological systems.

6 MATERIALS and METHODS

6.1 *Materials*

All chemicals used in this work were, if not stated otherwise, purchased from Sigma-Aldrich (Deisenhofen, Germany), Biomol (Hamburg, Germany), Biorad (Munich, Germany), Fluka (by Sigma-Aldrich, Deisenhofen, Germany), Invitrogen (Karlsruhe, Germany), Kodak (Stuttgart, Germany), Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany), Riedel de Haen (Seelze, Germany), Serva (Heidelberg, Germany). Reagents for molecular biology were purchased from Applied Biosystems (Darmstadt, Germany), New England Biolabs (Frankfurt am Main, Germany), PeproTech (Hamburg, Germany), Promega (Mannheim, Germany), Roche (Mannheim, Germany) and Waters (Germany). MilliQ water was used for the generation of solutions (Millipore, Schwalbach, Germany). All restriction enzymes and their respective buffers were purchased from Fermentas (St. Leon-Rot, Germany), Roche (Mannheim, Germany) and New England Biolabs (NEB) (Frankfurt am Main, Germany).

Master-Mixes for PCR and quantitative real-time PCR were purchased from Eppendorf (Hamburg, Germany) and Agilent (Boeblingen, Germany).

6.1.1 Reagents for RNA amplification

Agencourt AMPure XP system (60 ml) (BeckmanCoulter)

Agilent RNA 6000 Pico kit (Agilent Technologies)

Agilent High Sensitivity DNA kit (Agilent Technologies)

Antarctic Phosphatase (2 U/ μ l) with 10x Phosphatase buffer (NEB)

ATP (10 mM)

dATP (100 mM) (GE Healthcare)

Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies)

Exonuclease I (5 U/ μ l) with 10x Exonuclease I buffer (NEB)

GeneAmp 10x PCR buffer II and 25 mM MgCl₂ (Applied Biosystems)

GeneChip Eukaryotic poly-A RNA control kit (Affymetrix)

MEGAscript T7 kit (Applied Biosystems)

MinElute PCR Purification Kit (Qiagen)

Nonidet P-40 (5%) (Roche)

Qiaquick PCR purification kit (Qiagen)

Qiaquick gel extraction kit (Qiagen)

RNase H (2 U/ μ l) (Invitrogen)

RNase Inhibitor (cloned) (40U/ μ l) (Applied Biosystems)

RNeasy Mini kit (Qiagen)

RNeasy MinElute Cleanup kit (Qiagen)

ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre)

Small RNA Sample Prep Kit with the v1.5 sRNA 3' Adaptor (Illumina)

SUPERase-In (20U/ μ l) (Applied Biosystems)

SuperScript III reverse transcriptase (200 U/ μ l) with 0.1 M DTT (Invitrogen)

SYBR GreenER qPCR Super Mix (ABI PRISM, Applied Biosystems)

T4 gene 32 protein (5 μ g/ μ l) (Roche)

T4 Polynucleotide kinase (PNK) (10 U/ μ l) with 10x PNK buffer (NEB)

T4 RNA Ligase 2, truncated (200 U/ μ l) with 10x T4 RNL2 truncated reaction buffer (NEB)

Takara ExTaq Hot Start Version (5 U/ μ l) with 10x ExTaq buffer and 2.5 mM each dNTP mix (Takara)

Terminal deoxynucleotidyl transferase (TdT) (15 U/ μ l) (Invitrogen)

6.1.2 Equipment and material

Agilent 2100 Bioanalyzer (Agilent Technologies)
DNA LoBind Tubes, 1.5 ml (Eppendorf)
Gelsystem Mini (Peqlab)
Laser Capture Microscopy system LMD6000 (Leica, Germany)
Low affinity pipet tips (Fisher Scientific)
Illumina HiSeq2000 instrument (Illumina)
ND-1000 spectrophotometer (Nanodrop) (Peqlab)
PCR system (Applied Biosystems)
PCR tubes, 0.5 ml thin-walled with flat cap (Applied Biosystems)
8-well 0.2 ml PCR tube band (Greiner Bio-One)
POL frame slides (Leica, Germany)
Sliding microtome Leica SM2000 R (Leica, Germany)
Speedvac vacuum concentrator (Thermo Scientific)
StepOne Real-Time PCR system (Applied Biosystems)
Stereotactic chamber (Stoelting)
Thermomixer comfort (Eppendorf)

6.1.3 Software

Bioconductor	http://www.bioconductor.org/
Cluster3	http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm#ctv
DAVID	https://david.ncifcrf.gov/
Ensembl	http://www.ensembl.org/index.html
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
IGV	https://www.broadinstitute.org/igv/
NetPrimer	http://www.premierbiosoft.com/netprimer/
PANTHER	http://pantherdb.org/
Primer3 (v. 0.4.0)	http://frodo.wi.mit.edu/
R	https://www.r-project.org/
RSeQC	http://rseqc.sourceforge.net/

StepOne Real-Time PCR	Applied Biosystems, life technologies, CA, USA
ToppGene	https://toppgene.cchmc.org/
TreeView	http://jtreeview.sourceforge.net/
UCSC genome browser	https://genome.ucsc.edu/
Vector NTI	Invitrogen, Karlsruhe, Germany
VENNY	http://bioinfogp.cnb.csic.es/tools/venny/index.html
WebGestalt	http://bioinfo.vanderbilt.edu/webgestalt/

6.1.4 Primer

6.1.4.1 Amplification primer

Amplification primers were purchased from Metabion (HPLC purified, desalted and lyophilized).

UP1: 5'-ATATCTCGAGGGCGCGCCGGATCCTTTTTTTTTTTTTTTTTTTTTTTT-3'

UP2: 5'-ATATGGATCCGGCGCGCCGTGACTTTTTTTTTTTTTTTTTTTTTTTT-3'

T7-UP2:

5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGGATATGGATCCGGCGCGCCGTGAC-3'

6.1.4.2 qPCR primer

All primers were purchased from Metabion (desalted and lyophilized).

Housekeeping genes:

NAME	SEQUENCE (5' → 3')	POSITION	AMPLICON	EFFICACY
B-ACTIN	Fwd: TTGCTGACAGGATGCAGAAG Rev: ACATCTGCTGGAAGGTGGAC	exon/exon	141 bp	nd
GAPDH	Fwd: GTGTTCTACCCCAATGTGT Rev: ATTGTCATACCAGGAAATGAGCTT	exon/exon	248 bp	nd
L27	Fwd: ACAACCACCTCATGCCACA Rev: CTGGCCTTGCGCTTCAA	exon/exon	103 bp	nd
SRP14	Fwd: CAGCGTGTTCATCACCTCAA Rev: GGCTCTCAACAGACACTTGTTTT	exon/exon	109 bp	92 %

nd, not determined

Marker genes:

NAME	SEQUENCE (5' → 3')	POSITION	AMPLICON	EFFICACY
CALBINDIN	Fwd: GGATGGCAACGGATACATAGAT Rev: AAGAGCAAGGTCTGTTCCGGT	exon/exon	162 bp	99 %
CALBINDIN	Fwd: GGTAGAAACGGCCAGAAACA Rev: CCAGAAACCCCACCCTTATT	3'UTR	148 bp	88 %
CALRETININ	Fwd: ATGGCAAATTGGGTCTCTCA Rev: GGCGTCCAGTTCATTCTCAT	exon/exon	164 bp	nd
DCX	Fwd: TGCTCAAGCCAGAGAGAACA Rev: CTGCTTTCCATCAAGGGTGT	exon	205 bp	101 %
DCX	Fwd: TATTCCATGTGGCCCTCT Rev: TTGGGTTATTTGCTCCTT	3'UTR	127 bp	nd
DCX	Fwd: TTGTGAGGCATTTGGAGACA Rev: TTAAGCATTGCCTGGGAAAC	3'UTR	174 bp	nd
DCX	Fwd: TTTCCCAGGCAATGCTTAAC Rev: CCAAACAGCCCTCATCAAAT	3'UTR	202 bp	87 %
GFAP	Fwd: AACGACTATCGCCGCAACTGC Rev: ATGGCGCTCTTCTGTTTCGC	exon/exon	105 bp	109 %
NESTIN	Fwd: CCTTTCTTCTGTGGCTCACC Rev: TCATCATTGCTGCTCCTCTG	exon	149 bp	90 %
NEUROD1	Fwd: GCTGTTTGAGATGTGATGCTG Rev: GCAACTGCATGGGAGTTTTTC	3'UTR	112 bp	92 %
PROX1	Fwd: GCTATACCGAGCCCTCAACA Rev: GGCATTGAAAACTCCCGTA	exon/exon	106 bp	nd
SOX2	Fwd: CATGAGAGCAAGTACTGGCAAG Rev: CCAACGATATCAACCTGCATGG	3'UTR	127 bp	103 %
SOX4	Fwd: CCCTTCTTAAAATTTCTTTTTCTGC Rev: AGGTCCCATGTCCATTTTC	3'UTR	227 bp	87 %
SOX4	Fwd: GGCAGTTCGGTAAAGGGTTT Rev: TTGCACAGGATGGATGGTTA	3'UTR	88 bp	91 %
SOX11	Fwd: TGGTGACCTAGACCTCAGAGC Rev: GCGTATCCAGTTTCAAACCTTC	exon	128 bp	nd
SOX11	Fw:AAGTCATGTCTTCTGGTAGAGGTTA Rev:GTATGAATTCAACCCTGAGCATT	3'UTR	237 bp	nd

nd, not determined

Spike RNA control genes:

NAME	SEQUENCE (5' → 3')
DAP	Fwd: CCAGACCGCGCCTAATAATG Rev: CGCTTCTTCCACCAAGTGCAG
LYS	Fwd: GCCATATCGGCTCGAAATC Rev: AACGAATGCCGAAACCTCCTC
PHE	Fwd: TGAGCTCTAGGCCCAAACGAC Rev: TCCGGTTTTAGTCGGACGTG
THR	Fwd: GCCGATGCCGTAAGCAAG Rev: CAGCTCAGGCACAAGCATCG

6.1.5 Buffers, solutions and media

1F8-basic cell medium	DMEM-F12	500 ml
	Glutamine (100x)	1 % (v/v)
	FBS	10 % (v/v)
1F8-cell growth medium	1F8-basic cell medium	200 ml
	Puromycin (2 mg/ml)	200 µl
	Tetracyclin (1 mg/ml)	400 µl
	G418 sulfate (100 mg/ml)	600 µl
1F8-packaging medium	DMEM, High glucose	500 ml
	FBS	10 % (v/v)
	Glutamine (100x)	1 % (v/v)
	NEAA (100x)	1 % (v/v)
	Na-Pyruvate (100x)	1 % (v/v)
HEK293T cell medium	DMEM	500 ml
	FBS	10 % (v/v)
	Antibiotic-Antimycotic (100 x)	1 % (v/v)
4% PFA	Paraformaldehyde	4 % (w/v)
	NaOH	2-3 pellets
	Dissolve in 0.1 M PO ₄ -buffer. Heat up to solve; pH 7.4	
Phosphate buffer 0.2 M	Sodium phosphate monobasic	0.552 % (w/v)
	Sodium phosphate dibasic	2.19 % (w/v)
TBS (10x)	Tris	250 mM
	NaCl	1.37 M
	KCl	26 mM
	Adjust to pH 7.5	

Viral resuspension buffer	Pyrogen-free water	400 ml
	Tris-HCl (1M) (pH 7.8)	20 ml
	NaCl (5M)	10.4 ml
	KCl (1M)	4 ml
	MgCl ₂ (1M)	2 ml
DNA loading dye (6x)	Tris/HCl, pH 7.5	10 mM
	Glycerol	50 % (v/v)
	EDTA	100 mM
	Xylencyanol	0.25 % (w/v)
	Bromphenol blue	0.25 % (w/v)
EtBr	Ethidiumbromide staining solution	1 µg/ml

6.2 *Methods*

6.2.1 **Animals**

All experiments were carried out in accordance with the European Communities Council Directive (86/609/EEC). Stereotactic injections of retroviruses into the dentate gyrus of adult mice were approved by the Government of Upper Bavaria. For all experiments, 8 weeks old female mice C57BL/6 from Charles River (Sulzfeld, Germany) were used. Mice were grouped housed in standard cages under a 12 hour light/dark cycle and had ad libitum access to food and water. Mice were kept for 1 week in a big exploratory cage including a running wheel prior to stereotactic injection in order to increase adult neurogenesis.

6.2.2 Plasmid production

6.2.2.1 Cloning of plasmids for retrovirus preparation

DNA of interest was subcloned from other plasmids. Plasmids of interest were digested with appropriate restriction endonucleases and further cloned into a shuttle-vector, which was derived from pBluescript II carrying additional restriction sites for PmeI and SfiI. Finally, cDNA was excised and ligated into a retroviral vector of choice in the correct orientation.

Buffers, enzymes and incubation times were used according to manufacturer's recommendations (NEB).

To prevent vector self-ligation and re-circularization an additional dephosphorylation step was performed subsequent to restriction digest. The vector was incubated with antarctic phosphatase (5 U/μl) at 37 °C for 30 min. 3 U of phosphatase were used for 1 μg of DNA.

Digested DNA fragments were subjected to agarose gel electrophoresis, excised from the gel and extracted using the QIAquick Gel Extraction kit (Qiagen) or the NucleoSpin Extract II kit (Macherey-Nagel) according to manufacturer's protocol.

For ligation, the amount of insert (i) was calculated relative to the amount of vector (v) (25 – 50 ng), also taking into account the length of fragments (l), and using a multiplying factor (mf) ranging between 3 and 10 according to the formula:

$$m_i = mf \times m_v \times \frac{l_i}{l_v}$$

Ligation was performed using 400 U T4-DNA-ligase (NEB) at a temperature ranging between 16 °C to 4 °C over night (o/n) or for up to three days. As negative control, digested vector without insert was used.

After transformation and plasmid isolation, positive clones were identified by restriction digest or colony PCR.

6.2.2.2 Transformation

100 µl DH5α or TOP10 competent E.coli cells were thawed on ice for 15 min, before cells were mixed with DNA and incubated on ice for 30 min. A heat-shock was applied at 42°C for 45 sec, followed by 2 min on ice. 1 ml prewarmed LB medium was added to the cells which were allowed to recover for 1 h gentle shaking (400 rpm) at 37°C. After short centrifugation at 9.000 rpm, supernatant was discarded except for 100 µl with which cells were resuspended and plated on agar plates containing 100 µg/µl ampicillin. Plates were incubated at 37 °C o/n.

6.2.2.3 Cultivation of bacteria and plasmid isolation

One colony of transformed E. coli was picked from the agar plate and transferred into 3 ml LB-medium containing 100 µg/ml ampicillin. Cells were cultured at 350 rpm and 37 °C for a minimum of 6 h or o/n.

Depending on the required amount of plasmid DNA two different procedures were followed. For purification of small amounts o/n cultures were subjected to NucleoSpin Plasmid kit (Macherey-Nagel) according to manufacturer's protocol. If higher amount of DNA was desired, approximately 100 µl (depending on cell culture density) of cell suspension was transferred to 200 ml LB-medium containing 100 µg/ml ampicillin, as previously, and again incubated at 37 °C o/n. O/n cultures were processed with PureYield Plasmid Midiprep System kit (Promega) following manufacturer's protocol. Plasmid DNA concentration and quality was measured using a Nanodrop spectrophotometer.

For further purification, DNA was purified by EtOH precipitation. 1/10 volume 3 M NaAc (pH 5.0) and 2.5 volume 100 % EtOH were added to DNA and incubated for 1 h at -20 °C. After centrifugation at 13,000 rcf for 30 min at 4 °C, DNA pellet was washed by covering with 70 % EtOH and centrifugation repeated. Finally DNA pellet was air-dried for 10 min at RT, resuspended in 50 µl MilliQ, and DNA concentration was measured using a Nanodrop spectrophotometer.

6.2.3 Retrovirus preparation

Production of replication incompetent retrovirus was performed in HEK298T cells or GPG-1F8 cells, a human-derived packaging cell line for production of high titer retrovirus, as described previously (Tashiro et al 2006). The GPG-1F8 cells constitutively express the structural and enzymatic protein genes *gag* and *pol* of the Moloney murine leukemia virus (MMLV). The VSVG envelope gene is under Tet-off control and will be rapidly induced upon withdrawal of Tetracyclin. The viral genomic RNA will be encoded by a transiently transfected retroviral expression plasmid, minimally containing the 5' and 3' LTRs of MMLV, the retroviral packaging signal *psi*, a primer binding site for the retroviral reverse transcriptase, a cPPT element, and the cDNA of interest, i.e., CAG IRES mitodsRed. The presence of a WPRE element increases RNA stability and packaging efficiency. Only the stretch of the expression plasmid between the two LTR's will integrate. Elements of the plasmid outside this region will not be part of the integrated vector.

GPG-1F8 cells were grown in growth media, and six 10-cm dishes prepared for each virus transfection. Before transfection, media was replaced by Opti-MEM media supplemented with 10% FBS.

Transfection was performed using 9 ml Opti-MEM medium (without additives) supplemented with 360 μ l Lipofectamine 2000, and 9 ml Opti-MEM medium supplemented with 150 μ g DNA in two separate tubes and left for 10 min at RT. After combining the two tubes, the tube was incubated at RT for 30 min. 3 ml of reaction mix was added to each 10-cm dish containing 7 ml of Opti-MEM + 10% FBS. Plates were incubated at 37°C o/n.

12 to 16 hours later, the transfection mix was aspirated carefully, the cells washed and 11.5 ml of packaging media added to each 10-cm dish. Plates were incubated at 37°C for 3 days.

For virus harvest, supernatant was collected from each dish and filtered through a 0.45 μ m low-protein-binding PVDF filter (Millipore) in order to get rid of cell debris. Tubes were centrifuged at 27,000 rpm for 2 h at 4°C, virus pellet gently resuspended in 2 ml cold virus resuspension buffer, filled up to 30 ml with virus resuspension buffer and again centrifuged at 27,000 rpm for 2h at 4°C. Supernatant was aspirated carefully, virus pellet dried (upside down) and 50 – 70 μ l of ice-cold virus resuspension buffer added directly to the pellet. The tube was closed with parafilm in order to avoid evaporation of such small volumina, and left

for soaking for 3 h or o/n at 4°C. The pellet was resuspended gently (without air bubbles), aliquoted (10 µl aliquots) and aliquots stored at -80°C.

Virus was harvested three to four times (3, 5, 7, and 9 d after transfection). Katrin Wassmer mainly performed retrovirus production.

Viral titers were determined by transduction of HEK293T with a serial dilution of the retrovirus. 72 hours after viral transduction, cells were fixed with 4% PFA and fluorescent cells were counted under the fluorescence microscope. Viral titer was calculated as colony-forming units (cfu) per ml, which normally ranged between 1×10^8 cfu/ml and 1×10^9 cfu/ml. The retrovirus was injected at a titer of $1-2 \times 10^8$ cfu/ml.

6.2.4 Stereotactic injection

Stereotactic injections of retroviruses into the dentate gyrus of adult mice were approved by the Government of Upper Bavaria. For all experiments, 8 weeks old female mice C57BL/6 from Charles River (Sulzfeld, Germany) were used. Mice were grouped housed in standard cages under a 12 hour light/dark cycle and had ad libitum access to food and water. One week prior to injection until four days after injection cages were equipped with running wheels in order to increase the number of transduced cells for later analysis.

Anesthetics (Sleep and Awake) were prepared as described below. Mice were anesthetized by intraperitoneal (ip) injection of 300-400 µl Sleep (depending on weight). As soon as mice were deeply narcotized they were fixed in a stereotactic chamber and the scalp was opened with a scalpel at a length of app. 0.5-1 cm. Coordinates for injections into the HC were set relative to bregma (medial/lateral = ± 1.6 ; anterior/posterior = -1.9) and holes for injections into both hemispheres were drilled into the skull. 0.9 µl virus with a titer of 2×10^8 cfu/ml (diluted with 1x PBS) was injected into the left and right dentate gyrus (coordinates dorsal/ventral from dura mater = -1.9) at a speed of 250 nl/min using a Digital Lab Standard Stereotaxic Instrument. Transferring mice onto a heating plate normalized body temperature, scalp was sutured, and 400 µl Awake was injected ip to antagonize Sleep and wake up mice.

<u>Sleep:</u>	Fentanyl (Janssen-Cilag) (0.1 mg/ml)	5 % (v/v)	0.25 ml
	Dormicum (Roche) (5 mg/ml)	10 % (v/v)	0.5 ml
	Domitor (Pfizer) (1 mg/ml)	5 % (v/v)	0.25 ml
	NaCl		4.0 ml
 <u>Awake:</u>	Temgesic (Essex Pharma) (0.3 mg/ml)	3.4 % (v/v)	0.17 ml
	Antisedan (Pfizer) (5 mg/ml)	5 % (v/v)	0.25 ml
	Anexate (Hexal) (0.1 mg/ml)	50 % (v/v)	2.5 ml
	NaCl		2.08 ml

6.2.5 Fluorescence-activated cell sorting (FACS)

Mice were killed by cervical dislocation. Brains were removed quickly, specific regions (e.g., hippocampi) taken out in cold PBS and centrifuged at 1000 rcf for 3 min in order to remove residual PBS. Tissue from one mouse was resuspended in 1 ml of enzyme mixture (10 ml/gram tissue) and incubated for a maximum of 25 min in a 37°C warm water bath, while mixing and resuspending every 5 min with a flamed glass capillary until solution got “milky” and no big tissue pieces were detected anymore. The enzyme mixture was composed of two mixes. The first mix contained 2 mg Papain (Worthington) dissolved in 5 ml HBSS supplemented with 50 µl (50 mM) EDTA and 50 µl (100 mM) L-Cysteine, which was left shaking for 30 min at 4°C in order to activate Papain. The second mix was composed of 5 mg Dispase (Roche), 2 mg DNase I (Worthington) and 167 µl (1M) MgSO₄ dissolved in 5 ml HBSS. Both mixes were mixed and sterile filtered before use.

After dissociation, the tissue-enzyme mix was passed through a 70 µm filter into a 50 ml Falcon tube. The filter was flushed with 2 ml of DMEM/F12 (Gibco) supplemented with 10 % FBS. The cell suspension was transferred into a 15 ml Falcon using another 3 ml medium to rinse the 50 ml tube. After centrifugation at 1000 rcf for 3 min, the supernatant was discarded and the pellet resuspended in 10 ml DMEM/F12 supplemented with 10 % FBS. This centrifugation step was repeated once and the cell pellet was resuspended in 5 ml DMEM/F12 supplemented with 10 % FBS and mixed with 5 ml Percoll (Amersham) (4.5 ml Percoll plus 0.5 ml 10x PBS). The mixture was transferred into an UltraClear tube for SW41

buckets and ultra-centrifuged at 20000 rcf for 30 min at RT. Lipids and proteins were aspirated very slowly and gently. The neuronal cells were just above the red blood cell layer. 40 ml PBS was added in order to dilute the Percoll and the suspension transferred to a 50 ml tube, followed by a centrifugation at 1000 rcf for 3 min. Supernatant was discarded carefully and the cell pellet resuspended twice, once with 0.5 ml DMEM/F12 medium and the second time with 0.3 ml medium, and transferred into an Eppendorf tube.

This procedure enabled the dissociation of 300.000 to 700.000 single cells from two hippocampi, without debris in the mixture, which was very suitable for FACsorting.

Right before FACsorting, cells were passed through a 40 µm strainer and sorted based on the fluorescent signal. Flow cytometry was conducted in collaboration with Dr. Wolfgang Beisker at the Institute of Molecular Toxicology and Pharmacology at the HMGU.

For characterization of FACsorted cells, cells have been centrifuged at 150 rcf for 3 min onto glass slides using a Cytospin. Cells were fixed using 4% PFA for 15 min and washed 3 times for 5 minutes with TBS, the second time including DAPI at a concentration of 1:10000. Glass slides were mounted with Aqua-Poly/Mount (Polysciences).

6.2.6 Cryosectioning and LCM of retrovirally labeled cells

Mice were killed by cervical dislocation. Brains from three groups of three mice [6, 12, and 28 days post retroviral injection (dpi)] were quickly removed, flash-frozen in Isopentane (Sigma-Aldrich, Germany) and stored at -80°C. For cryosectioning, brains were cooled to -17°C and the knife to -18°C in a cryostat. 12 µm thick coronal sections were obtained, mounted on pre-cooled POL-frame slides (Leica Microsystems, Germany), and frozen at -80 °C until further processing.

Sections were dehydrated in 100 % cold ethanol for 1 minute, dried briefly and immediately used for LCM. A Laser Capture Microscopy system (LMD6000, Leica, Germany) was used to isolate the cell body of CAG-IRES-mdsRED labeled cells. Laser power, balance and speed was adjusted for each objective. Approximately 30 – 100 cells were isolated per animal within 2-3 hours using the 40x objective.

6.2.7 RNA amplification and library preparation

RNA amplification was based on previous amplification protocols (Kurimoto et al 2007, Tang et al 2010) with modifications to make it suitable for strand-specific mRNA-Seq and cells isolated from cryofixed brain tissue.

As in the case of the Tang protocol, the shearing of double-stranded cDNA leads to the loss of sequence orientation. Thus, an in-vitro transcription step after PCR amplification has been included, followed by the fragmentation of amplified sense RNA and the preparation of directional libraries following the Script Seq v2 library preparation kit by Epicentre (2011) or the Illumina pre-release directional mRNA-Seq library preparation kit v1.0 (2009) in order to obtain strand-specific RNA-Seq libraries.

LCM-collected cells were lysed by adding 4.5 μ l of freshly prepared cell lysis buffer into the PCR tube cap.

Cell lysis mastermix:

Component	1x volume (μl)	final concentration
10x PCR buffer II (without MgCl ₂)	0.45	1x
25 mM MgCl ₂	0.27	1.5 mM
5% NP-40	0.45	0.5%
0,1 M DTT	0.225	5 mM
SUPERase-In (20 U/ μ l)	0.045	0.2 U/ μ l
RNase Inhibitor (40 U/ μ l)	0.045	0.4 U/ μ l
UP1 primer (10 ng/ μ l)	0.09	0.2 ng/ μ l (~12.5 nM)
dNTP mix (2.5 mM each)	0.09	50 μ M (each)
spike RNA mixture diluted to 1 ng	1	1 ng
Nuclease-free water	1.835	-
Total volume	4.5	

Cells were collected in the epitube by centrifugation at 7500 g for 15 sec, followed by cell lysis at 75°C for 10 minutes in a heatblock or waterbath. The cell lysis time was increased to 10 minutes in order to break up cells from cryofixed brain tissue. It is highly recommended

to assess the optimal lysis time for the tissue of interest. Especially cells isolated *in vivo* from the intact tissue may require longer lysis time than cells from cell culture.

After cell lysis, tubes were immediately placed on ice and centrifuged at 7500 g for 15 sec. For first-strand cDNA synthesis, 0.5 μ l of RT mastermix was added to each tube, mixed by gentle tapping and incubated at 50°C for 30 min, followed by heat-inactivation of the reverse transcriptase at 70°C for 15 min. Tubes were immediately placed on ice and centrifuged at 7500 g for 30 sec.

Optionally, cDNAs can be stored at -80°C over night and amplified the day after.

Reverse transcription mastermix:

Component	1x volume (μl)	final concentration
SuperScriptIII reverse transcriptase (200 U/ μ l)	0.335	133U/ μ l
RNase inhibitor (40 U/ μ l)	0.055	4.45 U/ μ l
T4 gene 32 protein (5 μ g/ μ l)	0.11	1.1 μ g/ μ l
Total volume	0.5	

For elimination of unreacted UP1 primer, 1 μ l of Exonuclease I mastermix was added to each tube and incubated in a thermal cycler at 37°C for 30 min, followed by heat-inactivation at 80°C for 25 min. Tubes were immediately placed on ice and centrifuged at 7500 g for 30 sec.

Exonuclease I mastermix:

Component	1x volume (μl)	final concentration
10x Exonuclease I buffer	0.1	1x
Nuclease-free water	0.8	-
Exonuclease I (5 U/ μ l)	0.1	0.5 U/ μ l
Total volume	1	

First strand cDNA was polyadenylated at the 3' end by adding 6 μ l of terminal deoxynucleotidyl transferase (TdT) mastermix to each tube. Reactions were mixed by gentle tapping and incubated in a thermal cycler at 37°C for 15 min, followed by heat-inactivation at 70°C for 10 min. Tubes were immediately placed on ice and centrifuged at 7500 g for 30 sec.

3' poly(A) tailing TdT mastermix:

Component	1x volume (μl)	final concentration
10x PCR buffer II (without MgCl ₂)	0.6	1x
25 mM MgCl ₂	0.36	1.5 mM
dATP (100 mM)	0.18	3 mM
Nuclease-free water	4.26	-
RNase H (2 U/ μ l)	0.3	0.1 U/ μ l
TdT (15 U/ μ l)	0.3	0.75 U/ μ l
Total volume	6	

Before second-strand cDNA synthesis, reaction tubes were split up to four tubes (3 μ l each), and 19 μ l of second-strand cDNA synthesis mastermix was added to each tube. Performing PCR amplification in independent tubes reduces the intrinsic variance associated with PCR amplification. Tubes were mixed gently, and second-strand cDNA synthesis was performed by one round of PCR as follows: 95°C for 3 min, 50°C for 2 min and 72°C for 10 min (hold at 4°C).

Second-strand cDNA synthesis mastermix:

Component	1x volume (μl)	final concentration
10x ExTaq buffer	7.6	1x
dNTP Mix (2.5 mM each)	7.6	0.25 mM (each)
UP2 primer (1 μ g/ μ l)	1.52	0.02 μ g/ μ l (~1 μ M)
Nuclease-free water	58.52	-
ExTaq Hot Start Version (5 U/ μ l)	0.76	0.05 U/ μ l
Total volume	76	

This bi-tagged double-stranded cDNA was amplified by adding 19 μ l of PCR mastermix to each tube (endvolume: 41 μ l/tube) followed by 20 cycles of PCR as follows: 95°C for 3 min, then 20 cycles of 95°C for 30 sec, 67°C for 1 min and 72°C for 6 min (+6 sec each cycle), followed by 72°C for 10 min (hold at 4°C). After combining the 4 reaction tubes, the amplified

cDNA products were purified using the Qiaquick PCR purification kit according to the manufacturer's instructions, eluting with 50 μ l EB buffer.

PCR amplification mastermix:

Component	1x volume (μl)	final concentration
10x ExTaq buffer	7.6	1x
dNTP Mix (2.5 mM each)	7.6	0.25 mM (each)
UP1 primer (1 μ g/ μ l)	1.52	0.02 μ g/ μ l (~1 μ M)
Nuclease-free water	58.52	-
<u>ExTaq Hot Start Version (5 U/μl)</u>	<u>0.76</u>	<u>0.05 U/μl</u>
Total volume	76	

After this initial PCR amplification, the amplification efficiency and purity of the samples was evaluated by qRT-PCR based on the expression of housekeeping genes, spike-RNA transcripts and genes of interest. 1 μ l of amplified cDNA was diluted 20-fold, and 1 μ l of this diluted PCR product used as a template in a 20 μ l SYBR green real-time PCR reaction (see 6.2.8).

An additional primer bearing the T7 promoter sequence (T7-UP2) was added to the 5' end of amplified cDNAs by preparing eight 50 μ l PCR reactions per sample, followed by 9 cycles of PCR as follows: 95°C for 5 min 30 s, 64°C for 1 min and 72°C for 5 min 18 s (for one cycle); 95°C for 30 s, 67°C for 1 min and 72°C for 5 min 18 s (+6 s for each cycle) (for eight cycles); followed by 72°C for 10 min (hold at 4°C). After PCR amplification, all eight PCR reactions were combined and purified using the Qiaquick PCR purification kit according to the manufacturer's instructions, eluting with 30 μ l EB buffer.

PCR mastermix for T7 promoter addition at the 5' end of cDNA:

Component	1x volume (μl)	final concentration
Purified first-round PCR products	0.7	-
10x ExTaq buffer	5	1x
dNTP Mix (2.5 mM each)	5	0.25 mM (each)
T7-UP2 primer (1 μ g/ μ l)	1	0.02 μ g/ μ l (~1 μ M)
UP1 primer (1 μ g/ μ l)	1	0.02 μ g/ μ l (~1 μ M)
Nuclease-free water	36,8	-
ExTaq Hot Start Version (5 U/ μ l)	0,5	0.05 U/ μ l
Total volume	50	

In order to eliminate primer dimer products shorter than 300 bp, PCR products were gel purified by 2% agarose gel electrophoresis (100 V, 10 min). It is not recommended to run electrophoresis for longer time, as large gel volumes result in inefficient recovery of DNA. Fractions above 300 bp were cut from the gel and purified with the QiaQuick Gel Extraction kit, according to the manufacturer's instructions, eluting with 30 μ l EB buffer. The elimination of primer dimer products by gel purification is a crucial step, as these dimer products result in high numbers of polyA/polyT reads and thus, dramatically influence the percentage of mappability of sequenced reads.

2 μ l of gel-purified cDNA was used for IVT (4 hours at 37°C) in a reaction volume of 10 μ l using the MEGAscript T7 kit (Ambion) according to manufacturer's guidelines. After IVT, samples were treated with 0.5 μ l Turbo DNase (2 U/ μ l) at 37°C for 15 min in order to eliminate the cDNA. RNA was purified using the RNeasy Mini Kit (Qiagen) according to manufacturer's guidelines, eluting with 30 μ l RNase-free water.

After this step, all cDNAs were reverse transcribed to RNA by T7 RNA polymerase, having the orientation of the original RNA transcripts prior to amplification.

RNA concentration was measured using a Nanodrop spectrophotometer, yielding several μ g of amplified RNA in total per sample. RNA size distribution was evaluated by 2% agarose gel electrophoresis and by the Bioanalyzer using an RNA Pico chip. 1-2 μ g of RNA were subjected to cDNA synthesis and transcript abundances measured by qPCR.

Amplified RNAs can be stored at -80°C , but were used for library preparation within the next weeks in order to avoid RNA degradation.

For Illumina libraries, 100 ng purified RNA was used and processed following the Illumina pre-release directional mRNA-Seq library preparation protocol v1.0 (2009). In brief, 100 ng purified RNA was fragmented using RNA fragmentation buffer (at 94°C for 8 min) and purified using the RNeasy MinElute Cleanup kit (Qiagen) according to the manufacturer's instructions. Each 1 μl of diluted and fragmented RNA was loaded on an Agilent RNA 6000 Pico chip in order to evaluate the RNA before and after fragmentation. RNA ends were repaired by Antarctic phosphatase treatment (0.1 U) at 37°C for 30 min, followed by phosphorylation via T4 Polynucleotide Kinase (PNK) treatment (0.2 U) at 37°C for 60 min. End-repaired RNA transcripts were purified using the RNeasy MinElute Cleanup kit (Qiagen) and volume dried down using a speedvac concentrator (Thermo Scientific) (30 min at RT). 3' and 5' adaptors were serially ligated to RNA ends. First, the 3' adaptor was ligated by the truncated version of T4 RNA ligase 2 (28 U) (22°C for 60 min), which specifically ligates the pre-adenylated 5' end of the v1.5 sRNA 3' adaptor to the 3' end of RNA. Next, the 5' adaptor was ligated by T4 RNA Ligase 1 at 20°C for 60 min. Unlike full length T4 RNA Ligase 1, the truncated version of T4 RNA Ligase 2 is unable to ligate the phosphorylated 5' end of RNA (adaptor) to the 3' end of RNA, which ultimately allows the ligation of both 3' and 5' adaptors subsequently without an intermediate purification step. After reverse transcription (at 44°C for 60 min), the di-tagged cDNA was amplified by 12 cycles of PCR, thereby adding the Illumina sequencing primers required for cluster generation: 98°C for 30 sec, then 12 cycles of 98°C for 10 sec, 60°C for 30 sec and 72°C for 15 sec, followed by 72°C for 10 min (hold at 4°C). Libraries were purified by two rounds of AMPure XP bead purification (BeckmanCoulter), finally eluting in 10 μl EB buffer (Qiagen). 1 μl of the ready-to-sequence library was loaded on an Agilent High Sensitivity DNA chip in order to evaluate library quality and quantity. Libraries were quantified by serial dilution qPCR.

For Epicentre libraries, 10 ng of purified RNA was fragmented and libraries have been prepared using the Script Seq v2 library preparation kit by Epicentre following the instructions by the manufacturer. In brief, 10 ng of amplified sense RNA was fragmented using the RNA Fragmentation Solution (at 85°C for 5 min). Each 1 μl of diluted and fragmented RNA was loaded on an Agilent RNA 6000 Pico chip in order to evaluate the RNA

before and after fragmentation. Fragmented RNA was reverse transcribed into cDNA using random-sequence primers containing a tagging sequence at their 5' ends (at 42°C for 20 min). The 3' end of single stranded cDNA was tagged by a terminal tagging oligo (TTO) containing a tagging sequence at its 5' end and a 3'-blocking group on the 3'-terminal nucleotide, which prevents extension of the TTO by the DNA Polymerase (at 25°C for 15 min). The di-tagged cDNA was purified using the MinElute PCR Purification Kit (Qiagen), eluting in 25 µl EB buffer or the AMPure XP system (BeckmanCoulter), eluting in 25 µl nuclease-free water. Finally, the di-tagged cDNA was amplified by 10 cycles of PCR, thereby adding the Illumina sequencing primers required for cluster generation: 95°C for 1 min, then 10 cycles of 95°C for 30 sec, 55°C for 30 sec and 68°C for 3 min, followed by 68°C for 7 min (hold at 4°C). Libraries were purified by AMPure XP bead purification (BeckmanCoulter), finally eluting in 20 µl nuclease-free water. 1 µl of the ready-to-sequence library was loaded on an Agilent High Sensitivity DNA chip in order to evaluate library quality and quantity. Libraries were quantified by serial dilution qPCR.

6.2.8 qRT-PCR

Quantitative real-time PCR (qRT-PCR) was performed on a StepOne instrument (Applied Biosystems) using Brilliant II SYBR Green qPCR Master Mix (Agilent Technologies) or SYBR GreenER qPCR Super Mix (ABI PRISM, Applied Biosystems) for detection. Each qPCR was set up in a 20 µl reaction volume according to the manufacturer's guidelines, using 200 nM primer each and 1 µl of diluted cDNA (1:10 – 1:20). Reactions were set up in replicates or triplicates. For the Agilent mastermix the following cycling conditions were used: 95°C for 2 min, then 40 cycles of 95°C for 5 sec and 60°C for 20 sec. For the ABI PRISM mastermix the following cycling conditions were used: 95°C for 10 min, then 40 cycles of 95°C for 15 sec and 60°C for 1 min.

Expression was calculated using the comparative C_T method (Schmittgen & Livak 2008).

NetPrimer and Primer3 software were used for design of primers (see 6.1.4). Primer efficiency was calculated on the basis of C_T values measured at different cDNA concentrations.

6.2.9 Bioinformatic analysis

6.2.9.1 Sequencing, Read Mapping, Annotation and RNAseq data processing

The libraries RMN013 and RMN014 were sequenced on the GAII platform (Illumina), generating approximately 35 million 75 bp single-end reads per sample. All other libraries were sequenced on the HiSeq2000 platform (Illumina), generating approximately 200 million 75 bp single-end reads per sample.

Quality of reads was determined using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and RSeQC (Wang et al 2012).

RNAseq data was analysed using R programming language (version 3.1.1) and packages from Bioconductor project (version 2.14) (<http://www.bioconductor.org/>). Genes were defined with GenomicFeatures package (Lawrence et al 2013) using “mm10” genome by annotations available in UCSC Genome Browser (Karolchik et al 2014) and grouping exons by gene, resulting in total of 23725 genes. Annotation for Entrez gene identifiers (gene symbol, refseq identifiers, GO terms) was created using org.Mm.eg.db package (R package version 2.14.0.).

Counts per gene for each sample were generated using Rsamtools package (<http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>). Test for differential expression was performed with DESeq2 package (Love et al 2014) with replacement of outliers based on Cook’s distance. We performed pairwise comparisons of the three time points measured. The p-values were adjusted for multiple hypotheses testing using the method of Benjamini-Hochberg (Benjamini & Hochberg 1995). A gene was defined as differentially regulated if adjusted p-value was <0.05.

For the reads distribution analysis a custom perl script was used. This script analysed the mapping coordinates and CIGAR string for each mapping record and compared it with a reference annotation to classify the type of mapping - UTR5/UTR3/exonic/intronic/intergenic and the mapping alignment (sense/antisense).

Gene body coverage and directionality of reads were calculated using RSeQC (Wang et al 2012).

The expression levels of different genomic classes were calculated using the RPKM method (Mortazavi et al 2008).

6.2.9.2 *De novo* transcript assembly for the discovery of novel transcripts

Cufflinks was used for the discovery of novel transcripts and antisense transcripts (Trapnell et al 2010). All mapped reads (mapped to mm10 using TopHat) were passed to Cufflinks together with an annotation file, which contained the positions of the RefSeq genes (coding and non-coding) together with the newly identified lncRNAs (TCONS) by Ramos et al (2013). Cufflinks reported several gene classes, of which 40602 transcripts fell into the class of known transcripts (“complete match of intron chain”) and 6809 transcripts represented potential novel isoforms. Two gene classes were not specified in the provided annotation file, and thus likely represent not yet annotated transcripts: i) 5632 transcripts fell into the category “exonic overlap with reference transcript opposite strand”, which overlapped 4374 coding genes in antisense direction (see 4.6); ii) 21229 transcripts fell into the category “unknown intergenic transcript” (see 4.7).

Count data were generated using HT-Seq (Anders et al 2015). Normalization and differential expression analysis of novel lncRNAs and TCONS by Ramos et al 2013 was carried out using DESeq Upper-Quartile Normalization (UQN) (Anders & Huber 2010) and edgeR (TMM) (Robinson & Oshlack 2010). Genes were called differentially expressed if they exhibited a Benjamini-Hochberg (FDR) adjusted p-value < 5% and a mean fold change of >2. Normalized mean counts from replicates were used for fold change calculations.

For the classification of lncRNAs, a custom script was used that employed "bedtools intersect (v2.17.0)" (Quinlan & Hall 2010) function to perform genomic intersection on various RNAs of interest.

6.2.9.3 Cluster Analysis and Gene Ontology Analysis

Cluster analysis was performed using Cluster 3.0 (Eisen et al 1998) and JavaTreeview (Saldanha 2004) software. Regularized log-transformation (RLT) was applied to normalize the count data and the average expression at each time point was computed per gene.

Differentially expressed genes were clustered by the Average Linkage clustering method using a centered correlation similarity matrix.

Gene ontology analysis was performed using DAVID (Huang et al 2008) and ToppGene (<https://toppgene.cchmc.org/>).

6.2.9.4 Neighbouring gene correlation analysis

The program "closest-features" part of the BEDOP suite (v2.2.0) (Neph et al 2012) was used to identify the closest protein-coding neighbour. The direction of the two genes was ignored and genomic overlaps allowed.

Pearson correlations were calculated based on count expression values of the two neighbouring ncRNA-mRNA genes.

7 ABBREVIATIONS

Amp	Ampicillin
as	antisense
ATP	Adenosine tri phosphate
bp	base pair
BrdU	Bromodeoxyuridine
CA1-3	Cornu Ammonis area 1-3 (parts of the hippocampus)
CAG	CMV early enhancer/chicken β actin
CAGE	cap analysis of gene expression
cDNA	complementary DNA
CDS	coding sequence
CEL-Seq	cell expression by linear amplification and sequencing
ceRNA	competing endogenous RNA
CFP	cyan fluorescent protein
cfu	colony forming unit
ChIP	chromatin immunoprecipitation
CNS	central nervous system
C _T	cycle threshold (qPCR)
DAPI	4',6-Diamidino-2-phenylindole
DAVID	Database for Annotation, Visualization and Integrated Discovery
DCX	doublecortin
DG	dentate gyrus (part of the hippocampus)
dATP	Desoxyadenosinmonophosphat
dCTP	Desoxycytidintriphosphat
dGTP	Desoxyguanosintriphosphat
dNTP	Deoxynucleotides triphosphate
dTTP	Desoxythymidintriphosphat
dUTP	Desoxyuridintriphosphat
DNA	Desoxyribonucleic acid
dpi	days post injection

DSN	duplex-specific nuclease
DTT	dithiothreitol
E	embryonic day
EB buffer	Elution buffer
ECM	extracellular matrix
E.coli	Escherichia coli
EGF	epidermal growth factor
ENCODE	ENCyclopedia of DNA Elements
ERCC	external RNA controls consortium
ESC	embryonic stem cell
EtBr	Ethidium Bromide
EtOH	Ethanol
FACS	flow- or fluorescence-activated cell sorting
FAIRE-Seq	formaldehyde-assisted isolation of regulatory elements
FANTOM	Functional Annotation of the Mammalian Genome
FBS	fetal bovine serum
FGF	fibroblast growth factor
FISH	fluorescence in situ hybridization
FPKM	fragments per kilobase per million mapped reads
fwd	forward
GABA	Gamma-aminobutyric acid
GAPDH	Glyceraldehyd-3-phosphate dehydrogenase
GCL	granule cell layer of DG
GENCODE	genome research of the ENCODE project
GFAP	glial fibrillary acidic protein
GFP	green fluorescent protein
GO	gene ontology
Grin	ionotropic glutamate receptor
Grm	metabotropic glutamate receptor
GWAS	genome wide association studies
GZ	granular zone

h	hours
HBSS	Hank's balanced salt solution
HC	hippocampus
HEK	human embryonic kidney cell
ip	intraperitoneal
iPSc	induced pluripotent stem cells
IPC	intermediate progenitor cell
IRES	internal ribosomal entry site
ISH	<i>in situ</i> hybridization
IVT	<i>in vitro</i> transcription
kb	kilo base
ko	knockout
L	liter
LCM	laser capture microdissection
lincRNA	long intergenic non-coding RNA
lncRNA	long non-coding RNA
LTP	long-term potentiation
LV	lateral ventricle
M	molar mass
m	milli (10^{-3})
μ	micro (10^{-6})
MARS-Seq	massively parallel single-cell RNA-Seq
min	minutes
miRNA	micro RNA
MLL1	myeloid/lymphoid or mixed-lineage leukemia protein 1
MMLV	Moloney murine leukemia virus
mRNA	messenger RNA
n	nano (10^{-9})
NAT	natural antisense transcript
ncRNA	non-coding RNA
NGS	Next-generation sequencing

NSC	neural stem cell
NPC	neural progenitor cell
NPC	nuclear pore complex
nt	nucleotide
OB	olfactory bulb
P	postnatal day
p	p-value for statistical analysis
p	pico (10^{-12})
PBS	Phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
rev	reverse
PFA	Paraformaldehyde
piRNA	Piwi-interacting RNA
PMA	Phi29-mRNA amplification
PNK	Polynucleotide kinase
PRC	Polycomb repressive complex
qRT-PCR	quantitative real time PCR
RBP	RNA binding protein
rcf	centrifugal force (g force)
RGL	radial glia like stem cell
RIP	RNA-immunoprecipitation
RLT	regularized log-transformation
RMS	rostral migratory stream
RNA	Ribonucleic acid
RPKM	reads per kilobase per million mapped reads
rpm	round per minute
RRBS	reduced representation bisulfite sequencing
rRNA	ribosomal RNA
RT	reverse transcription
RT	room temperature

SC3-seq	single-cell mRNA 3-prime end sequencing
sec	second
SGZ	subgranular zone (of the dentate gyrus)
SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension
siRNA	short interfering RNA
SMART	switching mechanism at the 5' end of the RNA transcript
smFISH	single-molecule fluorescence in situ hybridization
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
SNV	single nucleotide variation
SOLID	Sequencing by Oligonucleotide Ligation and Detection
SP	sequencing primer
STRT	SMART-based single-cell tagged reverse transcription
SVZ	subventricular zone (of the lateral ventricle)
TC	total count
TdT	terminal deoxynucleotidyl transferase
TF	transcription factor
TGS	Third generation sequencing
TPM	transcript per million
tRNA	transfer RNA
TSO	template switching oligonucleotide
TSS	transcription start site
TTO	terminal tagging oligo
UP	universal primer sequence
UQN	upper-quartile normalization
UTR	untranslated region
UV	ultraviolet
WGA	whole-genome amplification
WT	wild type

8 REFERENCES

2008. Method of the Year. *Nat Meth* 5: 1-1
- Aasebø IEJ, Blankvoort S, Tashiro A. 2011. Critical maturational period of new neurons in adult dentate gyrus for their involvement in memory formation. *European Journal of Neuroscience* 33: 1094-100
- Abrous DN, Wojtowicz JM. 2015. Interaction between Neurogenesis and Hippocampal Memory System: New Vistas. *Cold Spring Harbor Perspectives in Biology* 7
- Abusaad I, Mackay D, Zhao J, Stanford P, Collier DA, Everall IP. 1999. Stereological estimation of the total number of neurons in the murine hippocampus using the optical disector. *The Journal of Comparative Neurology* 408: 560-66
- Akers KG, Martinez-Canabal A, Restivo L, Yiu AP, De Cristofaro A, et al. 2014. Hippocampal Neurogenesis Regulates Forgetting During Adulthood and Infancy. *Science* 344: 598-602
- Altman J, Das GD. 1965. Autoradiographic and Histological Evidence of Postnatal Hippocampal Neurogenesis in Rats. *The Journal of Comparative Neurology* 124: 319-35
- Alvarez-Buylla A, Garcia-Verdugo JM, Tramontin AD. 2001. A unified hypothesis on the lineage of neural stem cells. *Nat Rev Neurosci* 2: 287-93
- Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. 2011. lncRNADB: a reference database for long noncoding RNAs. *Nucleic Acids Research* 39: D146-D51
- Amir E-aD, Davis KL, Tadmor MD, Simonds EF, Levine JH, et al. 2013. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology* 31: 545-52
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11: R106-R06
- Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, et al. 2013. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protocols* 8: 1765-86
- Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166-69
- Antoniou D, Stergiopoulos A, Politis PK. 2014. Recent advances in the involvement of long non-coding RNAs in neural stem cell biology and brain pathophysiology. *Frontiers in Physiology* 5: 155
- Apple DM, Fonseca RS, Kokovay E. 2016. The role of adult neurogenesis in psychiatric and cognitive disorders. *Brain Research*
- Aprèa J, Calegari F. 2015. Long non-coding RNAs in corticogenesis: deciphering the non-coding code of the brain. *The EMBO Journal* 34: 2865-84
- Arruda-Carvalho M, Sakaguchi M, Akers KG, Josselyn SA, Frankland PW. 2011. Posttraining Ablation of Adult-Generated Neurons Degrades Previously Acquired Memories. *The Journal of Neuroscience* 31: 15113-27
- Balu DT, Lucki I. 2009. Adult hippocampal neurogenesis: Regulation, functional implications, and contribution to disease pathology. *Neuroscience & Biobehavioral Reviews* 33: 232-52
- Batish M, van den Bogaard P, Kramer FR, Tyagi S. 2012. Neuronal mRNAs travel singly into dendrites. *Proceedings of the National Academy of Sciences of the United States of America* 109: 4645-50
- Batista Pedro J, Chang Howard Y. 2013. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* 152: 1298-307
- Beckervordersandforth R, Tripathi P, Ninkovic J, Bayam E, Lepier A, et al. 2010. In Vivo Fate Mapping and Expression Analysis Reveals Molecular Hallmarks of Prospectively Isolated Adult Neural Stem Cells. *Cell Stem Cell* 7: 744-58
- Beckervordersandforth R, Zhang C-L, Lie DC. 2015. Transcription-Factor-Dependent Control of Adult Hippocampal Neurogenesis. *Cold Spring Harbor Perspectives in Biology* 7

- Beiter T, Reich E, Weigert C, Niess AM, Simon P. 2007. Sense or antisense? False priming reverse transcription controls are required for determining sequence orientation by reverse transcription-PCR. *Analytical Biochemistry* 369: 258-61
- Bendall Sean C, Davis Kara L, Amir E-ad D, Tadmor Michelle D, Simonds Erin F, et al. 2014. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. *Cell* 157: 714-25
- Bendall SC, Simonds EF, Qiu P, Amir E-ad, Krutzik PO, et al. 2011. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science (New York, N.y.)* 332: 687-96
- Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57: 289-300
- Bergmann O, Spalding KL, Frisén J. 2015. Adult Neurogenesis in Humans. *Cold Spring Harbor Perspectives in Biology* 7
- Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, et al. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* 29: 3082-93
- Bertani S, Sauer S, Bolotin E, Sauer F. 2011. The Noncoding RNA Mistral Activates Hoxa6 and Hoxa7 Expression and Stem Cell Differentiation by Recruiting MLL1 to Chromatin. *Molecular cell* 43: 1040-46
- Bhan A, Hussain I, Ansari KI, Kasiri S, Bashyal A, Mandal SS. 2013. Antisense Transcript Long Noncoding RNA (lncRNA) HOTAIR is Transcriptionally Induced by Estradiol. *Journal of Molecular Biology*
- Bhartiya D, Pal K, Ghosh S, Kapoor S, Jalali S, et al. 2013. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database: The Journal of Biological Databases and Curation* 2013: bat034
- Bonaguidi Michael A, Wheeler Michael A, Shapiro Jason S, Stadel Ryan P, Sun Gerald J, et al. 2011. In Vivo Clonal Analysis Reveals Self-Renewing and Multipotent Adult Neural Stem Cell Characteristics. *Cell* 145: 1142-55
- Bonasio R, Shiekhattar R. 2014. Regulation of Transcription by Long Noncoding RNAs. *Annual review of genetics* 48: 433-55
- Bond Allison M, Ming G-I, Song H. 2015. Adult Mammalian Neural Stem Cells and Neurogenesis: Five Decades Later. *Cell Stem Cell* 17: 385-95
- Bracko O, Singer T, Aigner S, Knobloch M, Winner B, et al. 2012. Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis. *The Journal of Neuroscience* 32: 3376-87
- Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, et al. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Meth* 10: 1093-95
- Brenner S, Jacob F, Meselson M. 1961. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* 190: 576-81
- Breuss M, Heng Julian I-T, Poirier K, Tian G, Jaglin Xavier H, et al. 2012. Mutations in the β -Tubulin Gene TUBB5 Cause Microcephaly with Structural Brain Abnormalities. *Cell Reports* 2: 1554-62
- Brill MS, Ninkovic J, Winpenny E, Hodge RD, Ozen I, et al. 2009. Adult generation of glutamatergic olfactory bulb interneurons. *Nat Neurosci* 12: 1524-33
- Brown JP, Couillard-Després S, Cooper-Kuhn CM, Winkler J, Aigner L, Kuhn HG. 2003. Transient expression of doublecortin during adult neurogenesis. *The Journal of Comparative Neurology* 467: 1-10
- Brunner J, Neubrandt M, Van-Weert S, András T, Kleine Borgmann FB, et al. 2014. Adult-born granule cells mature through two functionally distinct states. *eLife* 3: e03104
- Brus M, Keller M, Levy F. 2013. Temporal features of adult neurogenesis: differences and similarities across mammalian species. *Frontiers in Neuroscience* 7

- Burbach GJ, Dehn D, Nagel B, Del Turco D, Deller T. 2004. Laser microdissection of immunolabeled astrocytes allows quantification of astrocytic gene expression. *Journal of Neuroscience Methods* 138: 141-48
- Burns JC, Kelly MC, Hoa M, Morell RJ, Kelley MW. 2015. Single-cell RNA-Seq resolves cellular complexity in sensory organs from the neonatal inner ear. *Nature Communications* 6: 8557
- Burns LT, Wenthe SR. 2014. From Hypothesis to Mechanism: Uncovering Nuclear Pore Complex Links to Gene Expression. *Molecular and Cellular Biology* 34: 2114-20
- Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, et al. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology* 16: 20
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development* 25: 1915-27
- Cai L, Dalal CK, Elowitz MB. 2008. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455: 485-90
- Cameron HA, McKay RD. 2001. Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. *The Journal of Comparative Neurology* 435: 406-17
- Cameron HA, Woolley CS, McEwen BS, Gould E. 1993. Differentiation of newly born neurons and glia in the dentate gyrus of the adult rat. *Neuroscience* 56: 337-44
- Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer MW. 2010. Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes. *Cell* 140: 372-83
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. 2005. The Transcriptional Landscape of the Mammalian Genome. *Science* 309: 1559-63
- Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, et al. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491: 454-57
- Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136: 642-55
- Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, et al. 2011. A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. *Cell* 147: 358-69
- Chan W-L, Huang H-D, Chang J-G. 2014. lncRNAMap: A map of putative regulatory functions in the long non-coding transcriptome. *Computational Biology and Chemistry* 50: 41-49
- Chancey JH, Poulsen DJ, Wadiche JI, Overstreet-Wadiche L. 2014. Hilar Mossy Cells Provide the First Glutamatergic Synapses to Adult-Born Dentate Granule Cells. *The Journal of Neuroscience* 34: 2349-54
- Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453: 544-47
- Chattopadhyay P, Roederer M. 2012. Cytometry: Today's Technology and Tomorrow's Horizons. *Methods (San Diego, Calif.)* 57: 251-58
- Che S, Ginsberg SD. 2003. Amplification of RNA transcripts using terminal continuation. *Lab Invest* 84: 131-37
- Cheetham SW, Gruhl F, Mattick JS, Dinger ME. 2013. Long noncoding RNAs and the genetics of cancer. *British Journal of Cancer* 108: 2419-25
- Chen G, Wang Z, Wang D, Qiu C, Liu M, et al. 2013. lncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Research* 41: D983-D86
- Chen J, Sun M, Kent WJ, Huang X, Xie H, et al. 2004. Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Research* 32: 4812-20
- Chi KR. 2008. The year of sequencing. *Nat Meth* 5: 11-14
- Ching T, Huang S, Garmire LX. 2014. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* 20: 1684-96
- Christian KM, Song H, Ming G-I. 2014. Functions and Dysfunctions of Adult Hippocampal Neurogenesis. *Annual Review of Neuroscience* 37: 243-62

- Clark BS, Blackshaw S. 2014. Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Frontiers in Genetics* 5: 164
- Clark MB, Mercer TR, Bussotti G, Leonardi T, Haynes KR, et al. 2015. Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat Meth* 12: 339-42
- Clelland CD, Choi M, Romberg C, Clemenson GD, Fragniere A, et al. 2009. A Functional Role for Adult Hippocampal Neurogenesis in Spatial Pattern Separation. *Science* 325: 210-13
- Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, et al. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Meth* 5: 613-19
- Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology* 17: 13
- Couillard-Despres S, Winner B, Karl C, Lindemann G, Schmid P, et al. 2006. Targeted transgene expression in neuronal precursors: watching young neurons in the old brain. *European Journal of Neuroscience* 24: 1535-45
- Creer DJ, Romberg C, Saksida LM, van Praag H, Bussey TJ. 2010. Running enhances spatial pattern separation in mice. *Proceedings of the National Academy of Sciences* 107: 2367-72
- Cui P, Lin Q, Ding F, Xin C, Gong W, et al. 2010. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* 96: 259-65
- D'Angelo Maximiliano A, Gomez-Cavazos JS, Mei A, Lackner Daniel H, Hetzer Martin W. 2012. A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation. *Developmental Cell* 22: 446-58
- Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, et al. 2015. A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences of the United States of America* 112: 7285-90
- David DJ, Samuels BA, Rainer Q, Wang J-W, Marsteller D, et al. 2009. Neurogenesis-Dependent and -Independent Effects of Fluoxetine in an Animal Model of Anxiety/Depression. *Neuron* 62: 479-93
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, et al. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America* 99: 5261-66
- DeCarolis NA, Eisch AJ. 2010. Hippocampal neurogenesis as a target for the treatment of mental illness: A critical evaluation. *Neuropharmacology* 58: 884-93
- Derrien T, Guigó R, Johnson R. 2012a. The long non coding RNAs (lncRNAs) : a new (p)layer in the "dark matter". *Frontiers in Genetics* 2
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, et al. 2012b. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research* 22: 1775-89
- Deshpande A, Bergami M, Ghanem A, Conzelmann K-K, Lepier A, et al. 2013. Retrograde monosynaptic tracing reveals the temporal evolution of inputs onto new neurons in the adult dentate gyrus and olfactory bulb. *Proceedings of the National Academy of Sciences* 110: E1152-E61
- Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. 2015. Integrated genome and transcriptome sequencing of the same cell. *Nat Biotech* 33: 285-89
- Dieni CV, Nietz AK, Panichi R, Wadiche JI, Overstreet-Wadiche L. 2013. Distinct Determinants of Sparse Activation during Granule Cell Maturation. *The Journal of Neuroscience* 33: 19131-42
- Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics* 14: 671-83
- Ding B, Zheng L, Zhu Y, Li N, Jia H, et al. 2015. Normalization and noise reduction for single cell RNA-seq experiments. *Bioinformatics* 31: 2225-27
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, et al. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research* 18: 1433-45

- Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS. 2009. NRED: a database of long noncoding RNA expression. *Nucleic Acids Research* 37: D122-D26
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, et al. 2012. Landscape of transcription in human cells. *Nature* 489: 101-08
- Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, et al. 1992. Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences of the United States of America* 89: 3010-14
- Ebisuya M, Yamamoto T, Nakajima M, Nishida E. 2008. Ripples from neighbouring transcription. *Nat Cell Biol* 10: 1106-13
- Eisch AJ, Petrik D. 2012. Depression and Hippocampal Neurogenesis: A Road to Remission? *Science* 338: 72-75
- Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* 95: 14863-68
- ENCODE. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57-74
- ENCODE PC. 2011. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol* 9: e1001046
- Erickson HS, Albert PS, Gillespie JW, Rodriguez-Canales J, Linehan WM, et al. 2009. Quantitative RT-PCR gene expression analysis of laser microdissected tissue samples. *Nature protocols* 4: 902-22
- Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn A-M, Nordborg C, et al. 1998. Neurogenesis in the adult human hippocampus. *Nat Med* 4: 1313-17
- Ernst A, Alkass K, Bernard S, Salehpour M, Perl S, et al. 2014. Neurogenesis in the Striatum of the Adult Human Brain. *Cell* 156: 1072-83
- Espina V, Wulfschuhle JD, Calvert VS, VanMeter A, Zhou W, et al. 2006. Laser-capture microdissection. *Nat. Protocols* 1: 586-603
- Espósito MS, Piatti VC, Laplagne DA, Morgenstern NA, Ferrari CC, et al. 2005. Neuronal Differentiation in the Adult Hippocampus Recapitulates Embryonic Development. *The Journal of Neuroscience* 25: 10074-86
- Fabel K, Wolf S, Ehninger D, Babu H, Galicia P, Kempermann G. 2009. Additive effects of physical exercise and environmental enrichment on adult hippocampal neurogenesis in mice. *Frontiers in Neuroscience* 3
- Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, et al. 2008. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of [beta]-secretase. *Nat Med* 14: 723-30
- Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug M, et al. 2010. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biology* 11: R56
- Faissner A, Reinhard J. 2015. The extracellular matrix compartment of neural stem and glial progenitor cells. *Glia* 63: 1330-49
- Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15: 7-21
- Fatima R, Akhade VS, Pal D, Rao SMR. 2015. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Molecular and Cellular Therapies* 3: 5
- Ferrández-Huertas C, Gil-Mínguez M, Luján R. 2012. Regional expression and subcellular localization of the voltage-gated calcium channel β subunits in the developing mouse brain. *Journal of Neurochemistry* 122: 1095-107
- Freedman ML, Monteiro ANA, Gayther SA, Coetzee GA, Risch A, et al. 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics* 43: 513-18
- Fritah S, Niclou SP, Azuaje F. 2014. Databases for lncRNAs: a comparative evaluation of emerging tools. *RNA* 20: 1655-65

- Fu GK, Xu W, Wilhelmy J, Mindrinis MN, Davis RW, et al. 2014. Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. *Proceedings of the National Academy of Sciences of the United States of America* 111: 1891-96
- Gangaraju VK, Lin H. 2009. MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol* 10: 116-25
- Gao Z, Ure K, Ables JL, Lagace DC, Nave K-A, et al. 2009. Neurod1 is essential for the survival and maturation of adult-born neurons. *Nat Neurosci* 12: 1090-92
- Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Meth* 8: 469-77
- Ge S, Goh ELK, Sailor KA, Kitabatake Y, Ming G-I, Song H. 2006. GABA regulates synaptic integration of newly generated neurons in the adult brain. *Nature* 439: 589-93
- Ge S, Yang C-h, Hsu K-s, Ming G-I, Song H. 2007. A Critical Period for Enhanced Synaptic Plasticity in Newly Generated Neurons of the Adult Brain. *Neuron* 54: 559-66
- Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10: 94-108
- Ginsberg SD. 2005. RNA amplification strategies for small sample populations. *Methods* 37: 229-37
- Ginsberg SD, Che S. 2004. Combined histochemical staining, RNA amplification, regional, and single cell cDNA analysis within the hippocampus. *Lab Invest* 84: 952-62
- Girós A, Morante J, Gil-Sanz C, Fairén A, Costell M. 2007. Perlecan controls neurogenesis in the developing telencephalon. *BMC Developmental Biology* 7: 29-29
- Gloss BS, Dinger ME. 2016. The specificity of long noncoding RNA expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859: 16-22
- Goff LA, Groff AF, Sauvageau M, Traves-Gibson Z, Sanchez-Gomez DB, et al. 2015. Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* 112: 6855-62
- Goldman SA, Nottebohm F. 1983. Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain. *Proceedings of the National Academy of Sciences* 80: 2390-94
- Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, et al. 2013. RNA-sequencing from single nuclei. *Proceedings of the National Academy of Sciences of the United States of America* 110: 19802-07
- Grun D, Kester L, van Oudenaarden A. 2014. Validation of noise models for single-cell transcriptomics. *Nat Meth* 11: 637-40
- Gu Y, Arruda-Carvalho M, Wang J, Janoschka SR, Josselyn SA, et al. 2012. Optical controlling reveals time-dependent roles for adult-born dentate granule cells. *Nat Neurosci* 15: 1700-06
- Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. 2013. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Research* 23: 2126-35
- Gutschner T, Diederichs S. 2012. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biology* 9: 703-19
- Guttman M, Amit I, Garber M, French C, Lin MF, et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458: 223-27
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477: 295-300
- Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* 482: 339-46
- Haddad F, Qin AX, Giger JM, Guo H, Baldwin KM. 2007. Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. *BMC Biotechnology* 7: 21-21

- Hangauer MJ, Vaughn IW, McManus MT. 2013. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genet* 9: e1003569
- Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Reports* 2: 666-73
- Hattori T, Baba K, Matsuzaki S, Honda A, Miyoshi K, et al. 2007. A novel DISC1-interacting partner DISC1-Binding Zinc-finger protein: implication in the modulation of DISC1-dependent neurite outgrowth. *Mol Psychiatry* 12: 398-407
- Hebenstreit D. 2012. Methods, Challenges and Potentials of Single Cell RNA-seq. *Biology* 1: 658-67
- Hebenstreit D, Fang M, Gu M, Charoensawan V, van Oudenaarden A, Teichmann SA. 2011. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Molecular Systems Biology* 7: 497-97
- Heng JI-T, Chariot A, Nguyen L. 2010. Molecular layers underlying cytoskeletal remodelling during cortical development. *Trends in Neurosciences* 33: 38-47
- Henke K. 2010. A model for memory systems based on processing modes rather than consciousness. *Nat Rev Neurosci* 11: 523-32
- Hirose T, Mishima Y, Tomari Y. 2014. Elements and machinery of non-coding RNAs: toward their taxonomy. *EMBO reports* 15: 489-507
- Hodge RD, Kowalczyk TD, Wolf SA, Encinas JM, Rippey C, et al. 2008. Intermediate Progenitors in Adult Hippocampal Neurogenesis: Tbr2 Expression and Coordinate Regulation of Neuronal Output. *The Journal of Neuroscience* 28: 3707-17
- Hodge RD, Nelson BR, Kahoud RJ, Yang R, Mussar KE, et al. 2012. Tbr2 is essential for hippocampal lineage progression from neural stem cells to intermediate progenitors and neurons. *The Journal of Neuroscience* 32: 6275-87
- Hou Y, Guo H, Cao C, Li X, Hu B, et al. 2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res*
- Hu HY, He L, Khaitovich P. 2014. Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes. *BMC Genomics* 15: 457
- Hu W, Alvarez-Dominguez JR, Lodish HF. 2012. Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Rep* 13: 971-83
- Huang DW, Sherman BT, Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* 4: 44-57
- Huang L, Ma F, Chapman A, Lu S, Xie XS. 2015. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annual Review of Genomics and Human Genetics* 16: 16.1-16.24
- Ibarra A, Hetzer MW. 2015. Nuclear pore proteins and the control of genome functions. *Genes & Development* 29: 337-49
- Iseki K, Hagino S, Zhang Y, Mori T, Sato N, et al. 2011. Altered expression pattern of testican-1 mRNA after brain injury. *Biomedical Research* 32: 373-78
- Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, et al. 2012. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat. Protocols* 7: 813-28
- Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, et al. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research* 21: 1160-67
- Islam S, Zeisel A, Joost S, La Manno G, Zajac P, et al. 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Meth* 11: 163-66
- Iwakiri J, Hamada M, Asai K. 2016. Bioinformatics tools for lncRNA research. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859: 23-30
- Iyengar BR, Choudhary A, Sarangdhar MA, Venkatesh KV, Gadgil CJ, Pillai B. 2014. Non-coding RNA interact to regulate neuronal development and function. *Frontiers in Cellular Neuroscience* 8: 47

- Jackson RJ, Hellen CUT, Pestova TV. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11: 113-27
- Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318-56
- Jagasia R, Steib K, Englberger E, Herold S, Faus-Kessler T, et al. 2009. GABA-cAMP Response Element-Binding Protein Signaling Regulates Maturation and Survival of Newly Generated Neurons in the Adult Hippocampus. *The Journal of Neuroscience* 29: 7966-77
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, et al. 2014. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science* 343: 776-79
- Jalali S, Bhartiya D, Lalwani MK, Sivasubbu S, Scaria V. 2013. Systematic Transcriptome Wide Analysis of lncRNA-miRNA Interactions. *PLoS ONE* 8: e53823
- Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, et al. 2011. Synthetic spike-in standards for RNA-seq experiments. *Genome Research* 21: 1543-51
- Jin J, Liu J, Wang H, Wong L, Chua N-H. 2013. PLncDB: plant long non-coding RNA database. *Bioinformatics* 29: 1068-71
- Jin W, Tang Q, Wan M, Cui K, Zhang Y, et al. 2015. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* 528: 142-46
- Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, et al. 2015. Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci* 18: 637-46
- Johri A, Beal MF. 2012. Mitochondrial Dysfunction in Neurodegenerative Diseases. *The Journal of Pharmacology and Experimental Therapeutics* 342: 619-30
- Juliá M, Telenti A, Rausell A. 2015. Sincell: an R/Bioconductor package for statistical assessment of cell-state hierarchies from single-cell RNA-seq. *Bioinformatics*
- Jung M-Y, Lorenz L, Richter JD. 2006. Translational Control by Neuroguidin, a Eukaryotic Initiation Factor 4E and CPEB Binding Protein. *Molecular and Cellular Biology* 26: 4277-87
- Kadakkuzha BM, Liu X-A, McCrate J, Shankar G, Rizzo V, et al. 2015. Transcriptome analyses of adult mouse brain reveal enrichment of lncRNAs in specific brain regions and neuronal populations. *Frontiers in Cellular Neuroscience* 9: 63
- Kalverda B, Pickersgill H, Shloma VV, Fornerod M. 2010. Nucleoporins Directly Stimulate Expression of Developmental and Cell-Cycle Genes Inside the Nucleoplasm. *Cell* 140: 360-71
- Kandel ER. 2001. The Molecular Biology of Memory Storage: A Dialogue Between Genes and Synapses. *Science* 294: 1030-38
- Kang E, Wen Z, Song H, Christian KM, Ming G-l. 2016. Adult Neurogenesis and Psychiatric Disorders. *Cold Spring Harbor Perspectives in Biology*
- Kang Y, Norris MH, Zarzycki-Siek J, Nierman WC, Donachie SP, Hoang TT. 2011. Transcript amplification from single bacterium for transcriptome analysis. *Genome Research* 21: 925-35
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, et al. 2007. RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science* 316: 1484-88
- Kapusta A, Feschotte C. 2014. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends in genetics : TIG* 30: 439-52
- Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, et al. 2013. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genetics* 9: e1003470
- Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, et al. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Research* 42: D764-D70
- Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, et al. 2005. Antisense Transcription in the Mammalian Transcriptome. *Science* 309: 1564-66
- Kee N, Teixeira CM, Wang AH, Frankland PW. 2007. Preferential incorporation of adult-generated granule cells into spatial memory networks in the dentate gyrus. *Nat Neurosci* 10: 355-62

- Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology* 13: R107-R07
- Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, et al. 2014. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 111: 6131-38
- Kempermann G, Gast D, Kronenberg G, Yamaguchi M, Gage FH. 2003. Early determination and long-term persistence of adult-generated new neurons in the hippocampus of mice. *Development* 130: 391-99
- Kempermann G, Jessberger S, Steiner B, Kronenberg G. 2004. Milestones of neuronal development in the adult hippocampus. *Trends in Neurosciences* 27: 447-52
- Kempermann G, Kuhn HG, Gage FH. 1997. Genetic influence on neurogenesis in the dentate gyrus of adult mice. *Proceedings of the National Academy of Sciences of the United States of America* 94: 10409-14
- Kempermann G, Song H, Gage FH. 2015. Neurogenesis in the Adult Hippocampus. *Cold Spring Harbor Perspectives in Biology* 7
- Kerever A, Mercier F, Nonaka R, de Vega S, Oda Y, et al. 2014. Perlecan is required for FGF-2 signaling in the neural stem cell niche. *Stem cell research* 12: 492-505
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences* 106: 11667-72
- Kharchenko PV, Silberstein L, Scadden DT. 2014. Bayesian approach to single-cell differential expression analysis. *Nat Meth* 11: 740-42
- Khodosevich K, Inta D, Seeburg PH, Monyer H. 2007. Gene Expression Analysis of In Vivo Fluorescent Cells. *PLoS ONE* 2: e1151
- Kim Daniel H, Marinov Georgi K, Pepke S, Singer Zakary S, He P, et al. 2015a. Single-Cell Transcriptome Analysis Reveals Dynamic Changes in lncRNA Expression during Reprogramming. *Cell Stem Cell* 16: 88-101
- Kim JK, Kolodziejczyk AA, Illicic T, Teichmann SA, Marioni JC. 2015b. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat Commun* 6
- Kim K-T, Lee HW, Lee H-O, Kim SC, Seo YJ, et al. 2015c. Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. *Genome Biology* 16
- Kindler S, Kreienkamp H-J. 2012. Dendritic mRNA Targeting and Translation In *Synaptic Plasticity*, ed. MR Kreutz, C Sala, pp. 285-305: Springer Vienna
- Kitamura T, Saitoh Y, Takashima N, Murayama A, Niibori Y, et al. 2009. Adult Neurogenesis Modulates the Hippocampus-Dependent Period of Associative Fear Memory. *Cell* 139: 814-27
- Kivioja T, Vaharautio A, Karlsson K, Bonke M, Enge M, et al. 2012. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Meth* 9: 72-74
- Klein Allon M, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. 2015. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* 161: 1187-201
- Klinge S, Voigts-Hoffmann F, Leibundgut M, Ban N. 2012. Atomic structures of the eukaryotic ribosome. *Trends in Biochemical Sciences* 37: 189-98
- Knoth R, Singec I, Ditter M, Pantazis G, Capetian P, et al. 2010. Murine Features of Neurogenesis in the Human Hippocampus across the Lifespan from 0 to 100 Years. *PLoS ONE* 5: e8809
- Kolodziejczyk Aleksandra A, Kim JK, Svensson V, Marioni John C, Teichmann Sarah A. 2015. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell* 58: 610-20
- Kondrashov N, Pusic A, Stumpf CR, Shimizu K, Hsieh Andrew C, et al. 2011. Ribosome-Mediated Specificity in Hox mRNA Translation and Vertebrate Tissue Patterning. *Cell* 145: 383-97
- Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of long non-coding RNA transcription. *BMC Biology* 11: 59-59

- Koyama Y, Hattori T, Nishida T, Hori O, Tohyama M. 2015. Alterations in dendrite and spine morphology of cortical pyramidal neurons in DISC1-binding zinc finger protein (DBZ) Knockout mice. *Frontiers in Neuroanatomy* 9
- Koyama Y, Hattori T, Shimizu S, Taniguchi M, Yamada K, et al. 2013. DBZ (DISC1-binding zinc finger protein)-deficient mice display abnormalities in basket cells in the somatosensory cortices. *Journal of Chemical Neuroanatomy* 53: 1-10
- Kuhn HG. 2015. Control of Cell Survival in Adult Mammalian Neurogenesis. *Cold Spring Harbor Perspectives in Biology* 7
- Kuhn HG, Dickinson-Anson H, Gage FH. 1996. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *The Journal of Neuroscience* 16: 2027-33
- Kukurba KR, Montgomery SB. 2015. RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*
- Kurimoto K, Yabuta Y, Ohinata Y, Ono Y, Uno KD, et al. 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Research* 34: e42-e42
- Kurimoto K, Yabuta Y, Ohinata Y, Saitou M. 2007. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat. Protocols* 2: 739-52
- Kuwabara T, Hsieh J, Muotri A, Yeo G, Warashina M, et al. 2009. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat Neurosci* 12: 1097-105
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of Novel Genes Coding for Small Expressed RNAs. *Science* 294: 853-58
- Laity JH, Lee BM, Wright PE. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology* 11: 39-46
- Laplagne DA, Espósito MS, Piatti VC, Morgenstern NA, Zhao C, et al. 2006. Functional Convergence of Neurons Generated in the Developing and Adult Hippocampus. *PLoS Biol* 4: e409
- Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, et al. 2013. Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology* 9: e1003118
- Lecault V, White AK, Singhal A, Hansen CL. 2012. Microfluidic single cell analysis: from promise to practice. *Current Opinion in Chemical Biology* 16: 381-90
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445: 168-76
- Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, et al. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Meth* 7: 709-15
- Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. 2014. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research* 42: D92-D97
- Li Y, Stam FJ, Aimone JB, Goulding M, Callaway EM, Gage FH. 2013. Molecular layer perforant path-associated cells contribute to feed-forward inhibition in the adult dentate gyrus. *Proceedings of the National Academy of Sciences* 110: 9106-11
- Liang Y, Franks TM, Marchetto MC, Gage FH, Hetzer MW. 2013. Dynamic Association of NUP98 with the Human Genome. *PLoS Genetics* 9: e1003308
- Lie D-C, Colamarino SA, Song H-J, Desire L, Mira H, et al. 2005. Wnt signalling regulates adult hippocampal neurogenesis. *Nature* 437: 1370-75
- Lie DC, Song H, Colamarino SA, Ming G-I, Gage FH. 2004. Neurogenesis In The Adult Brain: New Strategies for Central Nervous System Diseases. *Annual Review of Pharmacology and Toxicology* 44: 399-421
- Lim SM, Koraka P, Osterhaus ADME, Martina BEE. 2013. Development of a strand-specific real-time qRT-PCR for the accurate detection and quantitation of West Nile virus RNA. *Journal of Virological Methods* 194: 146-53

- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* 133: 523-36
- Liu K, Yan Z, Li Y, Sun Z. 2013. Linc2GO: a human LincRNA function annotation resource based on ceRNA hypothesis. *Bioinformatics* 29: 2221-22
- Liu L, Shi M, Wang L, Hou S, Wu Z, et al. 2011. Ndr2 expression in neurogenic germinal zones of embryonic and postnatal mouse brain. *Journal of Molecular Histology* 43: 27-35
- Liu N, Liu L, Pan X. 2014a. Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos. *Cellular and Molecular Life Sciences* 71: 2707-15
- Liu X, Ramirez S, Pang PT, Puryear CB, Govindarajan A, et al. 2012. Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* 484: 381-85
- Liu Y, Namba T, Liu J, Suzuki R, Shioda S, Seki T. 2010. Glial fibrillary acidic protein-expressing neural progenitors give rise to immature neurons via early intermediate progenitors expressing both glial fibrillary acidic protein and neuronal markers in the adult hippocampus. *Neuroscience* 166: 241-51
- Liu Y, Zhou J, White KP. 2014b. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* 30: 301-04
- Lledo P-M, Alonso M, Grubb MS. 2006. Adult neurogenesis and functional plasticity in neuronal circuits. *Nat Rev Neurosci* 7: 179-93
- Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba A. 2015. Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell* 17: 329-40
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550
- Lovén J, Orlando David A, Sigova Alla A, Lin Charles Y, Rahl Peter B, et al. 2012. Revisiting Global Gene Expression Analysis. *Cell* 151: 476-82
- Luo Y, Coskun V, Liang A, Yu J, Cheng L, et al. 2015. Single-Cell Transcriptome Analyses Reveal Signals to Activate Dormant Neural Stem Cells. *Cell* 161: 1175-86
- Lupu F, Alves A, Anderson K, Doye V, Lacy E. 2008. Nuclear Pore Composition Regulates Neural Stem/Progenitor Cell Differentiation in the Mouse Embryo. *Developmental Cell* 14: 831-42
- Lv J, Cui W, Liu H, He H, Xiu Y, et al. 2013. Identification and Characterization of Long Non-Coding RNAs Related to Mouse Embryonic Brain Development from Available Transcriptomic Data. *PLoS ONE* 8: e71152
- Ma L, Bajic VB, Zhang Z. 2013. On the classification of long non-coding RNAs. *RNA Biology* 10: 925-34
- Macaulay IC, Haerty W, Kumar P, Li Yi, Hu TX, et al. 2015. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Meth* 12: 519-22
- Macaulay IC, Voet T. 2014. Single Cell Genomics: Advances and Future Perspectives. *PLoS Genetics* 10: e1004126
- Macosko Evan Z, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161: 1202-14
- Mahar I, Bambico FR, Mechawar N, Nobrega JN. 2014. Stress, serotonin, and hippocampal neurogenesis in relation to depression and antidepressant effects. *Neuroscience & Biobehavioral Reviews* 38: 173-92
- Mahata B, Zhang X, Kolodziejczyk Aleksandra A, Proserpio V, Haim-Vilmovsky L, et al. 2014. Single-Cell RNA Sequencing Reveals T Helper Cells Synthesizing Steroids De Novo to Contribute to Immune Homeostasis. *Cell Reports* 7: 1130-42
- Malone CD, Hannon GJ. 2009. Small RNAs as Guardians of the Genome. *Cell* 136: 656-68
- Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, et al. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Meth* 7: 111-18
- Mardis ER. 2008a. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24: 133-41

- Mardis ER. 2008b. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 9: 387-402
- Marguerat S, Bähler J. 2012. Coordinating genome expression with cell size. *Trends in Genetics* 28: 560-65
- Marín-Burgin A, Mongiat LA, Pardi MB, Schinder AF. 2012. Unique Processing During a Period of High Excitation/Inhibition Balance in Adult-Born Neurons. *Science* 335: 1238-42
- Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, et al. 2014. From single-cell to cell-pool transcriptomes: Stochasticity in gene expression and RNA splicing. *Genome Research* 24: 496-510
- Markwardt SJ, Dieni CV, Wadiche JI, Overstreet-Wadiche L. 2011. Ivy/neurogliaform interneurons coordinate activity in the neurogenic niche. *Nat Neurosci* 14: 1407-09
- Marr HS, Basalamah MA, Bouldin TW, Duncan AW, Edgell CJ. 2000. Distribution of testican expression in human brain. *Cell Tissue Res* 302: 139-44
- Mattick JS. 2011. The central role of RNA in human development and cognition. *FEBS Letters* 585: 1600-16
- Meng Y, Zhang Y, Tregoubov V, Janus C, Cruz L, et al. 2002. Abnormal Spine Morphology and Enhanced LTP in LIMK-1 Knockout Mice. *Neuron* 35: 121-33
- Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10: 155-59
- Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* 105: 716-21
- Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, et al. 2011. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature biotechnology* 30: 99-104
- Mercer TR, Mattick JS. 2013. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* 20: 300-07
- Metzker ML. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* 11: 31-46
- Miller BR, Hen R. 2015. The current state of the neurogenic theory of depression and anxiety. *Current opinion in neurobiology* 0: 51-58
- Miller JA, Nathanson J, Franjic D, Shim S, Dalley RA, et al. 2013. Conserved molecular signatures of neurogenesis in the hippocampal subgranular zone of rodents and primates. *Development (Cambridge, England)* 140: 4633-44
- Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, et al. 2015. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* 349: 1351-56
- Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, et al. 2015. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat Biotech* 33: 269-76
- Moison C, Arimondo PB, Guieysse-Peugeot A-L. 2011. Commercial reverse transcriptase as source of false-positive strand-specific RNA detection in human cells. *Biochimie* 93: 1731-37
- Mongiat LA, Schinder AF. 2011. Adult neurogenesis and the plasticity of the dentate gyrus network. *European Journal of Neuroscience* 33
- Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, Cocho JA. 2013. A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism* 110: 3-24
- Morlan JD, Qu K, Sinicropi DV. 2012. Selective Depletion of rRNA Enables Whole Transcriptome Profiling of Archival Fixed Tissue. *PLoS ONE* 7: e42882
- Morris J, Singh JM, Eberwine JH. 2011. Transcriptome Analysis of Single Cells. *Journal of Visualized Experiments : JoVE*: 2634
- Morris KV, Mattick JS. 2014. The rise of regulatory RNA. *Nat Rev Genet* 15: 423-37
- Morrison SJ, Spradling AC. 2008. Stem Cells and Niches: Mechanisms That Promote Stem Cell Maintenance throughout Life. *Cell* 132: 598-611

- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* 5: 621-28
- Mu L, Berti L, Masserdotti G, Covic M, Michaelidis TM, et al. 2012. SoxC Transcription Factors Are Required for Neuronal Differentiation in Adult Hippocampal Neurogenesis. *The Journal of Neuroscience* 32: 3067-80
- Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, et al. 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502: 59-64
- Nakamura T, Yabuta Y, Okamoto I, Aramaki S, Yokobayashi S, et al. 2015. SC3-seq: a method for highly parallel and quantitative measurement of single-cell gene expression. *Nucleic Acids Research* 43: e60-e60
- Nakashiba T, Cushman Jesse D, Pelkey Kenneth A, Renaudineau S, Buhl Derek L, et al. 2012. Young Dentate Granule Cells Mediate Pattern Separation, whereas Old Granule Cells Facilitate Pattern Completion. *Cell* 149: 188-201
- Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28: 1919-20
- Neves G, Cooke SF, Bliss TVP. 2008. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nat Rev Neurosci* 9: 65-75
- Ng S-Y, Lin L, Soh BS, Stanton LW. 2013. Long noncoding RNAs in development and disease of the central nervous system. *Trends in Genetics* 29: 461-68
- Niazi F, Valadkhan S. 2012. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* 18: 825-43
- Niland CN, Merry CR, Khalil AM. 2012. Emerging roles for large non-coding RNAs in Cancer and Neurological Disorders. *Frontiers in Genetics* 3
- Ohlstein B, Kai T, Decotto E, Spradling A. 2004. The stem cell niche: theme and variations. *Current Opinion in Cell Biology* 16: 693-99
- Ozsolak F. 2012. Third Generation Sequencing Techniques and Applications to Drug Discovery. *Expert Opinion on Drug Discovery* 7: 231-43
- Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, et al. 2010. Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. *Cell* 143: 1018-29
- Ozsolak F, Milos PM. 2011a. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12: 87-98
- Ozsolak F, Milos PM. 2011b. Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdisciplinary Reviews. RNA* 2: 565-70
- Ozsolak F, Platt AR, Jones DR, Reifengerger JG, Sass LE, et al. 2009. Direct RNA sequencing. *Nature* 461: 814-18
- Pan X, Durrett RE, Zhu H, Tanaka Y, Li Y, et al. 2013. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proceedings of the National Academy of Sciences of the United States of America* 110: 594-99
- Pang KC, Frith MC, Mattick JS. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in genetics : TIG* 22: 1-5
- Paraskevopoulou MD, Georgakilas G, Kostoulas N, Reczko M, Maragkakis M, et al. 2013. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Research* 41: D239-D45
- Paraskevopoulou MD, Hatzigeorgiou AG. 2016. Analyzing MiRNA–LncRNA Interactions In *Long Non-Coding RNAs: Methods and Protocols*, ed. Y Feng, LZhang, pp. 271-86. New York, NY: Springer New York
- Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, et al. 2016. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Research* 44: D231-D38

- Parent JM, Yu TW, Leibowitz RT, Geschwind DH, Sloviter RS, Lowenstein DH. 1997. Dentate Granule Cell Neurogenesis Is Increased by Seizures and Contributes to Aberrant Network Reorganization in the Adult Rat Hippocampus. *The Journal of Neuroscience* 17: 3727-38
- Park C, Yu N, Choi I, Kim W, Lee S. 2014. IncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* 30: 2480-85
- Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, et al. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research* 37: e123-e23
- Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, et al. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (New York, N.Y.)* 344: 1396-401
- Patzke N, Spocter MA, Karlsson KÆ, Bertelsen MF, Haagenen M, et al. 2013. In contrast to many other mammals, cetaceans have relatively small hippocampi that appear to lack adult neurogenesis. *Brain Structure and Function* 220: 361-83
- Pease J, Sooknanan R. 2012. A rapid, directional RNA-seq library preparation workflow for Illumina sequencing - application note. *Nat Meth* 9
- Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. *Nat Rev Genet* 14: 880-93
- Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Research* 35: e128-e28
- Phillips J, Eberwine JH. 1996. Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. *Methods* 10: 283-88
- Piatti VC, Davies-Sala MG, Espósito MS, Mongiat LA, Trincherro MF, Schinder AF. 2011. The Timing for Neuronal Maturation in the Adult Hippocampus Is Modulated by Local Network Activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31: 7715-28
- Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Meth* 10: 1096-98
- Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* 9: 171-81
- Pickrell WO, Rees MI, Chung S-K. 2012. Chapter One - Next Generation Sequencing Methodologies - An Overview In *Advances in Protein Chemistry and Structural Biology*, ed. IR Mark, pp. 1-26: Academic Press
- Plessy C, Bertin N, Takahashi H, Simone R, Salimullah M, et al. 2010. Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nature methods* 7: 528-34
- Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, et al. 2014. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature biotechnology* 32: 1053-58
- Ponjavic J, Ponting CP, Lunter G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Research* 17: 556-65
- Ponting CP, Hardison RC. 2011. What fraction of the human genome is functional? *Genome Research* 21: 1769-76
- Ponting CP, Oliver PL, Reik W. 2009. Evolution and Functions of Long Noncoding RNAs. *Cell* 136: 629-41
- Ptak C, Aitchison JD, Wozniak RW. 2014. The multifunctional nuclear pore complex: a platform for controlling gene expression. *Current opinion in cell biology* 0: 46-53
- Qiu S, Luo S, Evgrafov O, Li R, Schroth GP, et al. 2012. Single-neuron RNA-Seq: technical feasibility and reproducibility. *Frontiers in Genetics* 3: 124
- Quek XC, Thomson DW, Maag Jesper L, Bartonicek N, Signal B, et al. 2015. IncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Research* 43: D168-D73
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-42

- Quinodoz S, Guttman M. 2014. Long non-coding RNAs: An emerging link between gene regulation and nuclear organization. *Trends in cell biology* 24: 651-63
- Qureshi I, Mehler M. 2013. Long Non-coding RNAs: Novel Targets for Nervous System Disease Diagnosis and Therapy. *Neurotherapeutics*: 1-15
- Qureshi IA, Mehler MF. 2012. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci* 13: 528-41
- Raices M, D'Angelo MA. 2012. Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. *Nat Rev Mol Cell Biol* 13: 687-99
- Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biology* 4: e309
- Ramos Alexander D, Diaz A, Nellore A, Delgado Ryan N, Park K-Y, et al. 2013. Integration of Genome-wide Approaches Identifies lncRNAs of Adult Neural Stem Cells and Their Progeny In Vivo. *Cell Stem Cell* 12: 616-28
- Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, et al. 2012. Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. *Nature biotechnology* 30: 777-82
- Rinn J, Guttman M. 2014. RNA and dynamic nuclear organization: Long noncoding RNAs may function as organizing factors that shape the cell nucleus. *Science (New York, N.Y.)* 345: 1240-41
- Rinn JL, Chang HY. 2012. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry* 81: 145-66
- Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11: R25-R25
- Rossner MJ, Hirrlinger J, Wichert SP, Boehm C, Newrzella D, et al. 2006. Global Transcriptome Analysis of Genetically Identified Neurons in the Adult Cortex. *The Journal of Neuroscience* 26: 9956-66
- Rotem A, Ram O, Shores N, Sperling RA, Schnall-Levin M, et al. 2015. High-Throughput Single-Cell Labeling (Hi-SCL) for RNA-Seq Using Drop-Based Microfluidics. *PLoS ONE* 10: e0116328
- Sabin Leah R, Delás MJ, Hannon Gregory J. 2013. Dogma Derailed: The Many Influences of RNA on the Genome. *Molecular cell* 49: 783-94
- Sackmann EK, Fulton AL, Beebe DJ. 2014. The present and future role of microfluidics in biomedical research. *Nature* 507: 181-89
- Sahay A, Hen R. 2007. Adult hippocampal neurogenesis in depression. *Nat Neurosci* 10: 1110-15
- Sahay A, Scobie KN, Hill AS, O'Carroll CM, Kheirbek MA, et al. 2011a. Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature* 472: 466-70
- Sahay A, Wilson Donald A, Hen R. 2011b. Pattern Separation: A Common Function for New Neurons in Hippocampus and Olfactory Bulb. *Neuron* 70: 582-88
- Saldanha AJ. 2004. Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20: 3246-48
- Saliba A-E, Westermann AJ, Gorski SA, Vogel J. 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research* 42: 8845-60
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi Pier P. 2011. A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* 146: 353-58
- Sanchez-Flores A, Abreu-Goodger C. 2014. A Practical Guide to Sequencing Genomes and Transcriptomes. *Current Topics in Medicinal Chemistry* 14: 398-406
- Santarelli L, Saxe M, Gross C, Surget A, Battaglia F, et al. 2003. Requirement of Hippocampal Neurogenesis for the Behavioral Effects of Antidepressants. *Science* 301: 805-09
- Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, et al. 2013. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biology* 14: R31-R31
- Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* 2: e01749

- Schmidt-Hieber C, Jonas P, Bischofberger J. 2004. Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature* 429: 184-87
- Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative CT method. *Nat. Protocols* 3: 1101-08
- Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome Regulation by Polycomb and Trithorax Proteins. *Cell* 128: 735-45
- Schwartz YB, Pirrotta V. 2007. Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* 8: 9-22
- Scobie KN, Hall BJ, Wilke SA, Klemenhausen KC, Fujii-Kuriyama Y, et al. 2009. Krüppel-Like Factor 9 Is Necessary for Late-Phase Neuronal Maturation in the Developing Dentate Gyrus and during Adult Hippocampal Neurogenesis. *The Journal of Neuroscience* 29: 9875-87
- Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, et al. 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498: 236-40
- Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, et al. 2014. Single cell RNA Seq reveals dynamic paracrine control of cellular variation. *Nature* 510: 363-69
- Shapiro E, Biezuner T, Linnarsson S. 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 14: 618-30
- Shi Z, Barna M. 2015. Translating the Genome in Time and Space: Specialized Ribosomes, RNA Regulons, and RNA-Binding Proteins. *Annual Review of Cell and Developmental Biology* 31: 31-54
- Shin J, Berg Daniel A, Zhu Y, Shin Joseph Y, Song J, et al. 2015. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* 17: 360-72
- Shin J, Ming G-I, Song H. 2014. Decoding neural transcriptomes and epigenomes via high-throughput sequencing. *Nat Neurosci* 17: 1463-75
- Shishkin AA, Giannoukos G, Kucukural A, Ciulla D, Busby M, et al. 2015. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat Meth* 12: 323-25
- Shors TJ, Anderson ML, Curlik li DM, Nokia MS. 2012. Use it or lose it: How neurogenesis keeps the brain fit for learning. *Behavioural Brain Research* 227: 450-58
- Sierra A, Encinas JM, Deudero JJP, Chancey JH, Enikolopov G, et al. 2010. Microglia Shape Adult Hippocampal Neurogenesis through Apoptosis-Coupled Phagocytosis. *Cell Stem Cell* 7: 483-95
- Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, et al. 2013. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences* 110: 2876-81
- Simon JA, Kingston RE. 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 10: 697-708
- Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 15: 121-32
- Siomi H, Siomi MC. 2009. On the road to reading the RNA-interference code. *Nature* 457: 396-404
- Siomi H, Siomi MC. 2010. Posttranscriptional Regulation of MicroRNA Biogenesis in Animals. *Molecular Cell* 38: 323-32
- Sonenberg N, Hinnebusch AG. 2009. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* 136: 731-45
- Song J, M. Christian K, Ming G-I, Song H. 2012a. Modification of hippocampal circuitry by adult neurogenesis. *Developmental Neurobiology* 72: 1032-43
- Song J, Sun J, Moss J, Wen Z, Sun GJ, et al. 2013. Parvalbumin interneurons mediate neuronal circuitry-neurogenesis coupling in the adult hippocampus. *Nat Neurosci* 16: 1728-30
- Song J, Zhong C, Bonaguidi MA, Sun GJ, Hsu D, et al. 2012b. Neuronal circuitry mechanism regulating adult quiescent neural stem-cell fate decision. *Nature* 489: 150-54
- Spaethling JM, Eberwine JH. 2013. Single-cell transcriptomics for drug target discovery. *Current Opinion in Pharmacology* 13: 786-90

- Spalding Kirsty L, Bergmann O, Alkass K, Bernard S, Salehpour M, et al. 2013. Dynamics of Hippocampal Neurogenesis in Adult Humans. *Cell* 153: 1219-27
- Spalding KL, Bhardwaj RD, Buchholz BA, Druid H, Frisén J. 2005. Retrospective Birth Dating of Cells in Humans. *Cell* 122: 133-43
- Spigoni G, Gedressi C, Mallamaci A. 2010. Regulation of Emx2 Expression by Antisense Transcripts in Murine Cortico-Cerebral Precursors. *PLoS ONE* 5: e8658
- Spitale RC, Tsai M-C, Chang HY. 2011. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6: 539-43
- St. Laurent G, Wahlestedt C, Kapranov P. 2015. The Landscape of long noncoding RNA classification. *Trends in Genetics* 31: 239-51
- Stegle O, Teichmann SA, Marioni JC. 2015. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 16: 133-45
- Steib K, Schäffner I, Jagasia R, Ebert B, Lie DC. 2014. Mitochondria Modify Exercise-Induced Development of Stem Cell-Derived Neurons in the Adult Brain. *The Journal of Neuroscience* 34: 6624-33
- Steiner B, Klempin F, Wang L, Kott M, Kettenmann H, Kempermann G. 2006. Type-2 cells as link between glial and neuronal lineage in adult hippocampal neurogenesis. *Glia* 54: 805-14
- Streets AM, Zhang X, Cao C, Pang Y, Wu X, et al. 2014. Microfluidic single-cell whole-transcriptome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 111: 7048-53
- Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 14: 103-05
- Subramanian L, Sarkar A, Shetty AS, Muralidharan B, Padmanabhan H, et al. 2011. Transcription factor Lhx2 is necessary and sufficient to suppress astroglialogenesis and promote neurogenesis in the developing hippocampus. *Proceedings of the National Academy of Sciences of the United States of America* 108: E265-E74
- Suh H, Deng W, Gage FH. 2009. Signaling in Adult Neurogenesis. *Annual Review of Cell and Developmental Biology* 25: 253-75
- Sun GJ, Sailor KA, Mahmood QA, Chavali N, Christian KM, et al. 2013. Seamless Reconstruction of Intact Adult-Born Neurons by Serial End-Block Imaging Reveals Complex Axonal Guidance and Development in the Adult Hippocampus. *The Journal of Neuroscience* 33: 11400-11
- Sun GJ, Zhou Y, Stadel RP, Moss J, Yong JHA, et al. 2015a. Tangential migration of neuronal precursors of glutamatergic neurons in the adult mammalian brain. *Proceedings of the National Academy of Sciences* 112: 9484-89
- Sun J, Bonaguidi MA, Jun H, Guo JU, Sun GJ, et al. 2015b. A septo-temporal molecular gradient of sfrp3 in the dentate gyrus differentially regulates quiescent adult hippocampal neural stem cell activation. *Molecular Brain* 8: 52
- Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. 2011a. Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* 332: 472-74
- Suter DM, Molina N, Naef F, Schibler U. 2011b. Origins and consequences of transcriptional discontinuity. *Current Opinion in Cell Biology* 23: 657-62
- Suzuki A, Matsushima K, Makinoshima H, Sugano S, Kohno T, et al. 2015. Single-cell analysis of lung adenocarcinoma cell lines reveals diverse expression patterns of individual cells invoked by a molecular target drug treatment. *Genome Biology* 16: 66
- Tadmouri A, Kiyonaka S, Barbado M, Rousset M, Fablet K, et al. 2012. Cacnb4 directly couples electrical activity to gene expression, a process defective in juvenile epilepsy. *The EMBO Journal* 31: 3730-44
- Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* 29: 288-99
- Takizawa T, Gudla PR, Guo L, Lockett S, Misteli T. 2008. Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes & Development* 22: 489-98

- Tang F, Barbacioru C, Nordman E, Xu N, Bashkirov VI, et al. 2010. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature protocols* 5: 10.1038/nprot.2009.236
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Meth* 6: 377-82
- Tang F, Lao K, Surani MA. 2011. Development and applications of single cell transcriptome analysis. *Nature methods* 8: S6-11
- Tanti A, Belzung C. 2013. Neurogenesis along the septo-temporal axis of the hippocampus: Are depression and the action of antidepressants region-specific? *Neuroscience* 252: 234-52
- Tashiro A, Makino H, Gage FH. 2007. Experience-Specific Functional Modification of the Dentate Gyrus through Adult Neurogenesis: A Critical Period during an Immature Stage. *The Journal of Neuroscience* 27: 3252-59
- Tashiro A, Sandler VM, Toni N, Zhao C, Gage FH. 2006. NMDA-receptor-mediated, cell-specific integration of new neurons in adult dentate gyrus. *Nature* 442: 929-33
- Temprana Silvio G, Mongiat Lucas A, Yang Sung M, Trincherro Mariela F, Alvarez Diego D, et al. 2015. Delayed Coupling to Feedback Inhibition during a Critical Period for the Integration of Adult-Born Granule Cells. *Neuron* 85: 116-30
- Ting DT, Wittner BS, Ligorio M, Jordan NV, Shah AM, et al. 2014. Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells. *Cell reports* 8: 1905-18
- Toni N, Laplagne DA, Zhao C, Lombardi G, Ribak CE, et al. 2008. Neurons born in the adult dentate gyrus form functional synapses with target cells. *Nat Neurosci* 11: 901-07
- Toni N, Schinder AF. 2016. Maturation and Functional Integration of New Granule Cells into the Adult Hippocampus. *Cold Spring Harbor Perspectives in Biology* 8
- Toni N, Teng EM, Bushong EA, Aimone JB, Zhao C, et al. 2007. Synapse formation on neurons born in the adult hippocampus. *Nat Neurosci* 10: 727-34
- Torarinsson E, Sawera M, Havgaard JH, Fredholm M, Gorodkin J. 2006. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Research* 16: 885-89
- Torarinsson E, Yao Z, Wiklund ED, Bramsen JB, Hansen C, et al. 2008. Comparative genomics beyond sequence-based alignments: RNA structures in the ENCODE regions. *Genome Research* 18: 242-51
- Tozuka Y, Fukuda S, Namba T, Seki T, Hisatsune T. 2005. GABAergic Excitation Promotes Neuronal Differentiation in Adult Hippocampal Progenitor Cells. *Neuron* 47: 803-15
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, et al. 2014. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nature biotechnology* 32: 381-86
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. 2010. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature biotechnology* 28: 511-15
- Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, et al. 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single cell RNA-seq. *Nature* 509: 371-75
- Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, et al. 2010. The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Molecular cell* 39: 925-38
- Tronel S, Belnoue L, Grosjean N, Revest J-M, Piazza P-V, et al. 2012. Adult-born neurons are necessary for extended contextual discrimination. *Hippocampus* 22: 292-98
- Tsang JCH, Yu Y, Burke S, Buettner F, Wang C, et al. 2015. Single-cell transcriptomic reconstruction reveals cell cycle and multi-lineage differentiation defects in Bcl11a-deficient hematopoietic stem cells. *Genome Biology* 16: 178
- Tzadok S, Caspin Y, Hachmo Y, Canaani D, Dotan I. 2013. Directionality of noncoding human RNAs: How to avoid artifacts. *Analytical Biochemistry* 439: 23-29

- Ulitsky I, Bartel David P. 2013. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* 154: 26-46
- Usoskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, et al. 2015. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci* 18: 145-53
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proceedings of the National Academy of Sciences of the United States of America* 87: 1663-67
- van Praag H, Schinder AF, Christie BR, Toni N, Palmer TD, Gage FH. 2002. Functional neurogenesis in the adult hippocampus. *Nature* 415: 1030-34
- Vivar C, Potter MC, Choi J, Lee J-y, Stringer TP, et al. 2012. Monosynaptic inputs to new neurons in the dentate gyrus. *Nat Commun* 3: 1107
- Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, et al. 2015. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Research* 43: 4363-64
- Vučičević D, Schrewe H, Ørom UA. 2014. Molecular mechanisms of long ncRNAs in neurological disorders. *Frontiers in Genetics* 5: 48
- Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. Understanding the transcriptome through RNA structure. *Nat Rev Genet* 12: 641-55
- Wang D, Bodovitz S. 2010. Single cell analysis: the new frontier in 'omics'. *Trends in Biotechnology* 28: 281-90
- Wang Kevin C, Chang Howard Y. 2011. Molecular Mechanisms of Long Noncoding RNAs. *Molecular cell* 43: 904-14
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472: 120-24
- Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28: 2184-85
- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57-63
- Williams RRE, Azuara V, Perry P, Sauer S, Dvorkina M, et al. 2005. Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *Journal of Cell Science* 119: 132-40
- Wills QF, Livak KJ, Tipping AJ, Enver T, Goldson AJ, et al. 2013. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat Biotech* 31: 748-52
- Wilson Nicola K, Kent David G, Buettner F, Shehata M, Macaulay Iain C, et al. 2015. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* 16: 712-24
- Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes & Development* 23: 1494-504
- Winner B, Winkler J. 2015. Adult Neurogenesis in Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology* 7
- Winter J, Jung S, Keller S, Gregory RI, Diederichs S. 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 11: 228-34
- Wirtz S, Schuelke M. 2011. Region-Specific Expression of Mitochondrial Complex I Genes during Murine Brain Development. *PLoS ONE* 6: e18897
- Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, et al. 2014. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Meth* 11: 41-46
- Wu MV, Sahay A, Duman RS, Hen R. 2015. Functional Differentiation of Adult-Born Neurons along the Septotemporal Axis of the Dentate Gyrus. *Cold Spring Harbor Perspectives in Biology* 7
- Wu P, Zuo X, Deng H, Liu X, Liu L, Ji A. 2013. Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases. *Brain Research Bulletin* 97: 69-80
- Xie C, Yuan J, Li H, Li M, Zhao G, et al. 2014. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Research* 42: D98-D103

- Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, et al. 2011. Antisense expression increases gene expression variability and locus interdependency. *Molecular Systems Biology* 7
- Xue S, Barna M. 2012. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature reviews. Molecular cell biology* 13: 355-69
- Yan L, Yang M, Guo H, Yang L, Wu J, et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* 20: 1131-39
- Yang J-H, Li J-H, Jiang S, Zhou H, Qu L-H. 2013. CHIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from CHIP-Seq data. *Nucleic Acids Research* 41: D177-D87
- Yang L, Duff MO, Graveley BR, Carmichael GG, Chen L-L. 2011. Genomewide characterization of non-polyadenylated RNAs. *Genome Biology* 12: R16-R16
- Yarmishyn AA, Kurochkin IV. 2015. Long noncoding RNAs: a potential novel class of cancer biomarkers. *Frontiers in Genetics* 6: 145
- Yi H, Cho Y-J, Won S, Lee J-E, Jin Yu H, et al. 2011. Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Research* 39: e140-e40
- Yoon J-H, Abdelmohsen K, Gorospe M. 2012a. Posttranscriptional Gene Regulation by Long Noncoding RNA. *Journal of Molecular Biology*
- Yoon J-H, Abdelmohsen K, Srikantan S, Yang X, Martindale Jennifer L, et al. 2012b. LincRNA-p21 Suppresses Target mRNA Translation. *Molecular cell* 47: 648-55
- Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, et al. 2015. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347: 1138-42
- Zeng W, Mortazavi A. 2012. Technical considerations for functional sequencing assays. *Nat Immunol* 13: 802-07
- Zhang K, Huang K, Luo Y, Li S. 2014. Identification and functional analysis of long non-coding RNAs in mouse cleavage stage embryonic development based on single cell transcriptome data. *BMC Genomics* 15: 845
- Zhang Y, Li J, Kong L, Gao G, Liu Q-R, Wei L. 2007. NATsDB: Natural Antisense Transcripts DataBase. *Nucleic Acids Research* 35: D156-D61
- Zhao C, Deng W, Gage FH. 2008. Mechanisms and Functional Implications of Adult Neurogenesis. *Cell* 132: 645-60
- Zhao C, Teng EM, Summers RG, Ming G-I, Gage FH. 2006. Distinct Morphological Stages of Dentate Granule Neuron Maturation in the Adult Mouse Hippocampus. *The Journal of Neuroscience* 26: 3-11
- Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, et al. 2010. Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell* 40: 939-53
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30: 892-97

9 APPENDIX

Eidesstaatliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung bzw. Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

Identification of molecular hallmarks and regulators during neuronal maturation in adult hippocampal neurogenesis using RNA-Sequencing

am Lehrstuhl für Entwicklungsgenetik unter der Anleitung und Betreuung durch

Univ.-Prof. Dr. W. Wurst

ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Abs. 6 und 7 Satz 2 angegebenen Hilfsmittel benutzt habe.

Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

Ich habe diese Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

Ich habe den angestrebten Doktorgrad **noch nicht** erworben und bin **nicht** in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich einverstanden.

München, den

.....

Unterschrift

CURRICULUM VITAE

Name	Elisabeth Johanna UTTENTHALER (ehem. Englberger)
E-mail	elisabeth.uttenthaler@gmx.at
Staatsangehörigkeit	Österreich
Geburtsdatum und -ort	16.02.1983, Ried im Innkreis
Familienstand	verheiratet, 2 Kinder

AUSBILDUNG

03/ 2008 - 09/ 2016	Promotion am Helmholtz-Zentrum München, Institut für Entwicklungsgenetik (Prof. W. Wurst), Arbeitsgruppe Adulte Neurogenese (Prof. Chichung Lie) Dissertationsthema: Identification of molecular hallmarks and regulators during neuronal maturation in adult hippocampal neurogenesis using RNA-Sequencing
10/ 2003 – 11/ 2007	Studium der Biologie an der Universität Salzburg Masterarbeit: Regulation of neurogenesis: Identification of novel interaction partners for the developmentally regulated Co-repressor CoREST (Prof. Hans-Christian Bauer)
10/ 2001 – 07/ 2012	Studium der Psychologie an der Universität Salzburg Diplomarbeit: The role of adult neurogenesis on learning and memory (Prof. Michael Doppelmayr)
06/ 2001	Matura am Bundesoberstufenrealgymnasium Ried im Innkreis

FORTBILDUNG

11/ 2010	40th SfN annual meeting in San Diego
10/ 2010	EMBO/EMBL Symposium “The Non-Coding Genome” in Heidelberg
03/ 2010	10th Spring School in Bioinformatics and Systems Biology in München (Prof. W. Mewes; Helmholtz Zentrum München)
09/ 2008	FEBS Praxiskurs “Structural Variations in Genome, Gene Expression, Single Cell Analysis: Arrays, Beads, High-throughput sequencing” in Prag (Prof. W. Ansorge; Ecole Polytechnique Federale Lausanne)
07 – 08/ 2005	Praktikum am Paul-Flechsig-Institut für Hirnforschung der Universität Leipzig Abteilung für Neuroanatomie (Prof. Th. Arendt) und Abteilung für Neurophysiologie (Prof. A. Reichenbach)

PUBLIKATIONEN

Jagasia R, Steib K, **Englberger E**, Herold S, Faus-Kessler T, Saxe M, Gage FH, Song H, Lie DC (2009) GABA-cAMP response element-binding protein signaling regulates maturation and survival of newly generated neurons in the adult hippocampus. *J Neurosci* 29:7966-7977.

**MANUSKRIPTE IN
VORBEREITUNG**

Uttenthaler E, Cernilogar F, Karl K, Zhu Z, Chia KH, Davies-Sala G, Aneichyk T, Covic M, Jagasia R, Ebert B, Merz K, Mu L, Doppelbauer C, Klopstock T, Schinder A, Müller N, Nagarajan N, Song H, Lie DC (2016). RNA-Sequencing of the *in vivo* neuronal maturation associated transcriptome (in preparation)