Aus dem Adolf-Butenandt-Institut
Lehrstuhl Physiologische Chemie
der Ludwig-Maximilians-Universität München
Vorstand: Prof. Dr. Andreas Ladurner

Cell-type-specific approaches for dissecting gene activity and chromatin structure within the *Drosophila* head

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität

vorgelegt von
Tamás Schauer
aus Mohács (Ungarn)
2013

Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuer: Prof. Dr. Andreas Ladurner
Zweitgutachter: Prof. Dr. Ralph A.W. Rupp
Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR
Tag der mündlichen Prüfung: 01.08.2014

Eidesstattliche Versicherung

Schauer, Tamás

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

"Cell-type-specific approaches for dissecting gene activity and chromatin structure within the *Drosophila* head"

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum Unterschrift Doktorand

Acknowledgements

Foremost I would like to thank Prof. Dr. Andreas Ladurner for giving me the opportunity to work on my PhD project in his group at EMBL and in his department at LMU. It would have been impossible to start, manage and finish my PhD work without his guidance and support.

I am especially grateful to Dr. Carla Margulies for the successful co-work, the motivating discussions and the useful feedback that helped the progress of my PhD project. Thank you for the scientific and non-scientific input, the active co-writing of our papers and the careful reading of my thesis manuscript.

My special thanks go to Dr. Petra Schwalie for the fruitful collaboration and discussions. Thank you for your enthusiastic work in analyzing my data what was crucial to write this manuscript.

I wish to acknowledge Sandra Esser at LMU and Bianca Nijmejer at EMBL for the technical support.

I am also grateful to Dr. Anton Eberharter and Dr. Corey Laverty for all the scientific and management-related help including reading this thesis.

I also would to thank all members of the Ladurner group/department and Margulies group, especially Ava Handley for the discussions and helping each others´ work as well as Dr. Gyula Timinszky and Dr. Markus Hassler for the scientific advices.

At last but not at least I thank my family, my parents, my friends in Hungary and in Germany as well as my life partner for supporting and motivating me in good and bad times.

Table of Contents

1 Summary

1.1 Summary (in English)

Multicellular organisms develop from a single cell (the zygote) and each cell type inherits the same genetic material from the zygote. Adult organisms are composed of terminally differentiated cell populations that carry the same genome but differential epigenomes. The epigenome consists of modifications or marks of the genome that determine which genes are activated or repressed. The differential activity of genes in distinct cells maintains their phenotype, identity and function. However, there have been few tools available until recently that would allow us to profile gene activity at the level of specific cell types. The lack of easily applicable, efficient cell-type-specific tools prompted me to develop novel biochemical methods and to refine existing protocols for profiling transcription, chromatin and mRNA levels genome-wide. I used cephalic (head) cell types of the adult fruit fly (*Drosophila melanogaster*) as a model system to study differential gene activity in differentiated cell types including neurons, glia and the fat body (adipocytes).

First, I developed a biochemical method, CAST-ChIP (Chromatin Affinity Purification from Specific cell Types), a combination of the UAS/Gal4 expression system and the affinity purification of tagged chromatin-bound reporters. To study transcription in distinct cell types, I expressed a tagged subunit of the RNA polymerase II complex in the cell type of interest and used the tag to generate cell-type-specific, genome-wide ChIP profiles. RNA polymerase II marks about 1500 genes unique to neurons or glia. Genes identified as neuronal share characteristic cellular function such as axon guidance of neurons and are expressed in other neuronal tissues, such as the larval central nervous system. Furthermore, I demonstrated that genomic regions marked by cell-type-specific RNA polymerase II show GFP-reporter activity localized within the labeled cell populations. This incidates that RNA polymerase II profiling is a suitable tool to distinguish gene activity in different cell types.

Second, I applied CAST-ChIP to study chromatin structure of cell types by profiling the incorporation of the active histone variant (H2A.Z), as a proof-of-principle

to study differences in chromatin structure between unrelated cell types. I found H2A.Z present at expressed genes and absent from inactive genes, as shown previously. However, H2A.Z-enriched regions do not completely overlap with RNA polymerase II regions. Interestingly, RNA polymerase II-bound genes lacking H2A.Z differ the most in their expression among dissected tissues. Therefore, I hypothesized that H2A.Z labels genes that are expressed in a cell-type-independent manner. To test this, I used CAST-ChIP to compare the cell-type-specific incorporation of H2A.Z. Surprisingly, H2A.Z profiles are remarkably similar in neurons and glia, with only about hundred significant differences. In addition, H2A.Z is present at those regions which share RNA polymerase II in both cell types and is absent from cell-type-specific regions. ChIP analysis of the fat body, which is another head cell type with a different developmental origin, led to the same results. To validate these findings by comparing distinct developmental stages, I found only a few H2A.Z and thousands of RNA polymerase II regions that differ between the embryo and adult head tissues. Thus, CAST-ChIP revealed a novel function of H2A.Z in marking genes with ubiquitous, cell-type-invariant expression. Using this approach, I could distinguish between ubiquitous (house-keeping) and specifically regulated genes. Together with analyses conducted by other groups, I found that ubiquitous genes share common regulatory features including promoter structure and gene length, and they form clusters marked by insulator binding proteins.

Third, I refined a fly RNA profiling approach (TRAP: Translating Ribosome Affinity Purification), first developed for the mouse, to obtain information about cell-type-specific post-transcriptional processes that regulate cellular function downstream to transcription. TRAP measures the ribosome-bound fraction of RNA and therefore identifies genes that are not only transcribed but also translated (translatome). The dynamic range of TRAP was greater compared to the previous ChIP-based methods and I identified twice as many transcripts as RNA polymerase II-bound genes using CAST-ChIP, indicating the greater resolution of the ribosome-tagging method. Using TRAP I uncovered transcripts carrying relevant neuronal functions that were hidden in the CAST-ChIP data lacking RNA polymerase II peaks. Several studies revealed that mild stress conditions induce changes only on the translational level; therefore, TRAP is a suitable tool to study such responses in various cell types.

In summary, in my PhD thesis I present and compare cell-type-specific methods to profile gene activity in *Drosophila* differentiated cells. I developed a novel method

(CAST-ChIP) and applied an existing method (TRAP) to map 1) transcription using RNA polymerase II CAST-ChIP; 2) chromatin structure using H2A.Z CAST-ChIP and 3) the translatome of ribosome-bound mRNA using TRAP. My results give useful, novel information for the scientific community: 1) the cell-type-specific profiles serve as a compendium of genes involved in the maintenance of cell identity and function; 2) using these approaches, I discovered a novel function of H2A.Z marking ubiquitous/ housekeeping genes, highlighting the differential regulation of cell-type-specific genes; 3) ChIP profiling does not identify all differences among cell types and therefore post-transcriptional profiling has to be involved in the analysis.

Cell-type-specific approaches presented in this thesis are promising tools that will allow us to describe cellular responses upon environmental perturbation, identifying differential responses to environmental change in distinct cell populations.

1.2 Zusammenfassung

Mehrzellige Organismen entwickeln sich aus einer einzigen Zelle (der Zygote) und jede Zelle erbt das gleiche genetische Material aus der Zygote. Die adulten Organismen sind aus terminal differenzierten Zellpopulationen zusammengesetzt, die das gleiche Genom, aber unterschiedliche Epigenome tragen. Das Epigenom besteht aus Modifikationen oder Markierungen des Genoms. Diese bestimmen, welche Gene aktiviert oder reprimiert werden. Die unterschiedliche Aktivität von Genen in unterschiedlichen Zellen erhält deren Phänotyp, Identität und Funktion aufrecht. Der Mangel an leicht anwendbaren und effizienten Zelltyp-spezifischen Werkzeugen hat mich dazu veranlasst, neuartige biochemische Methoden zu entwickeln und bestehende Protokolle zu verfeinern. Dadurch kann die Transkription, die Chromatinstruktur und die Menge an mRNA Zelltyp-spezifisch, genomweit profiliert werden. Ich benutzte Zellen vom Kopf der adulten Fruchtfliege (*Drosophila melanogaster*) als Modellsystem um die Genaktivität der differenzierten Zelltypen wie Neuronen, Glia und Fettzellen zu untersuchen.

Zuerst entwickelte ich ein biochemisches Verfahren, CAST-ChIP (Chromatin Affinitätsreinigung von spezifischen Zelltypen) genannt, welches eine Kombination aus dem UAS/Gal4 Expressionssystem und aus einer Affinitätsreinigung von einem

markierten Chromatin-gebundenen Reporter darstellt. Um die Transkription in verschiedenen Zelltypen zu untersuchen, exprimierte ich eine markierte Untereinheit des RNA Polymerase II Komplex im Zelltyp von Interesse und erzeugte Zelltyp-spezifische, genomweite ChIP Profile. RNA-Polymerase II markiert etwa 1500 Gene spezifisch für Neuronen oder Gliazellen. Gene, die als Neuron-spezifisch identifiziert wurden, haben charakteristische zelluläre Funktionen, wie zum Beispiel die Axon Führung von Neuronen. Sie werden auch in anderen neuronalen Geweben sowie im larvalen Zentralnervensystem exprimiert. Außerdem zeigte ich, dass genomische Regionen, die von Zelltyp-spezifischer RNA Polymerase II gebunden sind, GFP-Reporter-Aktivität innerhalb der markierten Zellpopulationen zeigen, was darauf hindeutet, dass das RNA-Polymerase II "Profiling" eine geeignete Methode ist, um Zelltyp-spezifische Gen-Aktivitäten unterscheiden zu können.

Zweitens, ich verwendete CAST-ChIP zur Untersuchung der Chromatin-Struktur verschiedener Zelltypen und erstellten Profile für den Einbau der aktiven Histon Variante H2A.Z ins Chromatin. Ich fand eine Inkorporation von H2A.Z bei exprimierten Genen und keinen H2A.Z Einbau bei inaktiven Genen, wie bereits gezeigt wurde. Allerdings überlappten die H2A.Z angereicherten Regionen nicht vollständig mit den RNA-Polymerase II Regionen. Interessanterweise unterschieden sich die Gene, die von RNA-Polymerase II jedoch nicht von H2A.Z gebunden wurden, in ihrer Expression zwischen sezierten Geweben. Daher stellte ich die Hypothese auf, dass die H2A.Z-markierten Gene in einer Zelltyp-unabhängigen Weise exprimiert werden. Um dies zu testen, vergliche ich den Zelltyp-spezifischen Einbau von H2A.Z mit der CAST-ChIP Methode. Überraschenderweise sind die H2A.Z Profile bemerkenswert ähnlich in Neuronen und Gliazellen, mit nur etwa hundert signifikanten Unterschieden. Darüber hinaus ist H2A.Z anwesend in den Regionen, die auch RNA-Polymerase II in beiden Zelltypen (Neuronen und Gliazellen) aufweisen, fehlt jedoch in Zelltyp-spezifischen Regionen. ChIP "Profiling" in einem anderen Zelltyp mit einer unterschiedlichen Entwicklungsherkunft (z.B. Fettzellen im Kopf) ergab die gleichen Ergebnisse. Um die Resultate durch einen Vergleich verschiedener Entwicklungsstadien zu validieren, fand ich nur ein paar H2A.Z und Tausende von RNA Polymerase II unterschiedliche Regionen zwischen Embryos und Kopfgewebe im adulten Stadium. So ergab die Anwendung von CAST-ChIP eine neue Funktion von H2A.Z als eine bestimmte Markierung von Genen, die ubiquitär und Zelltyp-unabhängig exprimiert werden. Mit Hilfe dieser Methode

konnte ich zwischen ubiquitär ("Housekeeping") und speziell regulierten Genen unterscheiden. Wir und andere stellten fest, dass die ubiquitär exprimierten Gene gemeinsame Eigenschaften wie Promoterstruktur und Genlänge haben und auch Cluster bilden, die von Insulator Proteinen markiert werden.

Drittens, wendete ich eine RNA-"Profiling"-Methode (TRAP: "Translating Ribosome Affinity Purification") an, um Informationen über Zelltyp-spezifische post-transkriptionelle Prozesse zu erhalten, die zelluläre Funktion auf vorgelagerten Stufen regulieren. TRAP misst die Ribosom-gebundene Fraktion von RNA und identifiziert daher Gene, die nicht nur transkribiert, sondern auch translatiert werden (Translatom). Der dynamische Bereich von TRAP war größer im Vergleich zu den vorherigen ChIP-basierten Methoden und so identifizierte ich etwa doppelt so viele Transkripte als RNA Polymerase II-gebundene Gene als mit CAST-ChIP Methode. Ich entdeckte mit der TRAP Methode die Transkripte, die eine neuronal- relevante Funktion besaßen, jedoch in den CAST-ChIP Daten nicht entdeckt wurden, da sie keine RNA Polymerase II gebunden hatten. Mehrere Studien zeigten, dass milde Stressbedingungen nur auf der Ebene des Translatoms induziert werden. Daher ist TRAP eine geeignete Technik, um solche Reaktionen in verschiedenen Zelltypen zu untersuchen.

Zusammenfassend präsentiere und vergleiche ich in meiner Doktorarbeit Zelltyp-spezifische "Profiling" Methoden, um die Aktivität von Genen in differenzierten Zellen von *Drosophila* zu messen. Ich entwickelte die neuartige CAST-ChIP Methode und wendete bestehende Methoden wie TRAP an, um 1) die Transkription durch RNA-Polymerase II mittels CAST-ChIP zu kartieren; 2) die Kartierung der Chromatin-Struktur mit H2A.Z durch CAST-ChIP und 3) das Translatom von Ribosom-gebundener RNA mittels TRAP darzustellen. Meine Ergebnisse geben wertvolle, neue Informationen für die wissenschaftliche Gemeinschaft: 1) die Zelltyp-spezifische Profile dienen als ein Kompendium von Genen, die bei der Aufrechterhaltung der Zell-Identität und der Zell-Funktion beteiligt sind; 2) durch die Verwendung dieser Methoden entdeckte ich eine neue Funktion von H2A.Z als eine Kennzeichnung für ubiquitäre "housekeeping" Gene, deren Regulation unterschiedlich zu den Zelltyp-spezifischen Genen ist, und 3) ChIP-"Profiling" kann nicht alle Unterschiede zwischen den Zelltypen identifizieren und somit muss auch post-transkriptionelles "Profiling" in der Analyse angewendet werden.

Die Zelltyp-spezifischen Methoden, die in dieser Doktorarbeit vorgestellt wurden, sind vielversprechende Werkzeuge für die Zukunft, um die zellulären

Reaktionen auf Störungen der Umweltbedingungen zu beschreiben und um diese bestimmten Änderungen in verschiedenen Zellpopulationen identifizieren zu können.

2 Introduction

Metazoans by definition consist of various cell types forming tissues and organs. All mature forms of multicellular organisms derive from a single cell, the fertilized egg, and develop through a complex differentiation process. The genetic information is inherited from this single cell by all other cells in the organism. Thus, there is the same set of genes, the same genome in each cell of an individual (excepting programmed and random genetic rearrangements and cells without nucleus). The complexity of cell types in such organisms cannot be solely explained by the DNA sequence of their genome. **Epigenetics** describes such regulatory mechanisms that are "epi-" (above) the genomic sequence. Epigenetics, by its early definition from Waddington (1942), is "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being". These epigenetic interactions differ from cell type to cell type. In order to understand biological processes during development, or even within a terminally differentiated system, we therefore need to dissect how gene activity is regulated in various cell types. Several approaches have been developed for cell-type-specific genomics, yet they are time- and material consuming, require special equipment or they use treatments that are a potential stress for cells (see **chapter 2.3**)

In this thesis, I present novel cell-type-specific approaches that I have developed to explore gene regulation within different cell types of the fruit fly ***Drosophila melanogaster***. The fruit fly is a suitable model for such studies, being a complex organism with distinct organs, tissues and cell types (see **chapter 2.1**), showing conserved gene regulatory features (see **chapter 2.2**), a highly annotated genome and powerful forward and reverse genetic tools readily available. I took advantage of the genetic repertoire of *Drosophila* by expressing genetically encoded tagged reporters in the cell type of interest. I used the tag for cell-type-specific affinity purification either for Chromatin ImmunoPrecipitation (ChIP) assays or for mRNA isolation, followed by high-throughput DNA sequencing. I developed a biochemical protocol, **CAST-ChIP** (Chromatin Affinity Purification from Specific cell Types - ChIP; **chapter 3 and 4**) and adapted a method for *Drosophila*, **TRAP** (Translating Ribosome Affinity Purification; **chapter 5**). These approaches are quick, efficient and sensitive, contributing to describe and understand biological processes that distinguish cell types.

2.1 Epigenetic landscape of development

In recent years, high-throughput methods have allowed us to determine whole genome sequences. Since 2000, when the *Drosophila* and also the human **genomes** were first published (Adams et al., 2000; Lander et al., 2001), next-generation sequencing technologies appeared that made genomics a daily routine; genomes of individuals are now publicly available. Although each individual carries only one genome, there are hundreds of epigenomes in distinct cell types. The **epigenome** is the programming of gene expression in specific cells. Using next-generation techniques, we can now in principle also investigate cellular function at the level of epigenomes.

All cells of an organism derive from a single, fertilized egg via cellular proliferation and differentiation. The genome of the single cell is multiplied during proliferation through DNA replication and mitotic cell division. Epigenomes are established via differentiation processes when cells start to change their morphology, shape, size and so their function. Waddington originally visualized this with a slope of a hill with different downhill paths (**Figure 2.1**; Hochedlinger et al., 2009). At the top lies a ball symbolizing a **totipotent** cell, which can differentiate to all cell types, as the ball can also reach different local minima. The end points are individual cells, some located close to each other, meaning that they share similar function, and therefore they carry similar epigenomic profiles. In between, the gene expression profile of cells changes during specialization. Some cells are **pluri/multipotent** (e.g. stem cells) keeping a more plastic status and being able to differentiate to other cell types; from their local minimum they can get to other minima. Other cells (e.g. muscles, neurons) have a very specialized function and they change their gene activity only in a predetermined way. Cells also have cellular memory (they "remember" their origin) and maintain their status by sending signals to their neighbors and themselves. The basis of how cells make a decision, choose and preserve their fate is defined by which genes are turned on and off and how this is regulated epigenetically. On top of the development into a specific cell type, cells have to be able to adapt to sudden or chronic environmental changes, usually without loosing their epigenetic status or identity.

In this section, I will introduce how the main cell-lineages diverge in the early development of *Drosophila* and I describe the major cell types of a fully **differentiated fruit fly head**, which I used in my gene activity studies.

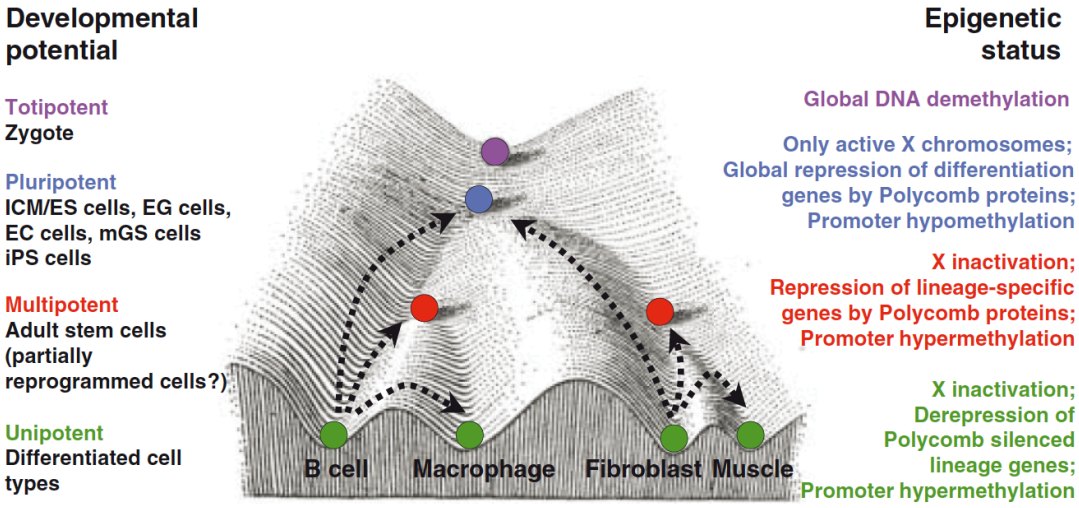


Figure 2.1 Modified figure of Waddington´s epigenetic landscape.
The totipotent zygote can develop to any type of cells, whereas differentiated cell types are unipotent. The epigenetic status indicates examples of programming ("downhill") and re-programming (arrows "uphill") events. Source: Hochedlinger et al., 2009.

2.1.1 Development of major cell-lineages in *Drosophila*

The early development of many insects, including *Drosophila*, is special compared to other model organisms. The zygotic nucleus divides mitotically without separating the daughter cells with membranes, and the nuclei migrate to the periphery surrounding the central yolk. This process is called superficial cleavage, and the nuclei without cell membranes form the syncytial **blastoderm**. Although the nuclei share a common cytoplasm, there are RNA and protein gradients deposited maternally, defining the axes of the embryo. The next event during nuclear division is the polar bud formation, where the first cells start to cellularize forming the pole cells. After that, the somatic nuclei are also separated by cell membranes, forming what is called the cellular blastoderm. During the cellularization process, transcription turns on taking over the regulation from maternal transcripts. In that early phases there are only three cell types present: somatic cells at the periphery, pole cells at the posterior end and vitellophages in the yolk. See textbooks: Campos-Ortega et al., 1985; Gilbert, 2000.

The formation of **germ layers** (ecto-, meso- and endoderm) occurs during **gastrulation**. Gastrulation in insects is especially complex, including several invaginations and cell migrations. The first invagination event is the formation of the

ventral furrow, which leads to development of the mesoderm and the anterior endoderm. At the same time as the ventral furrow appears, the cephalic furrow becomes visible, separating the head from the rest of the embryo. Shortly after, the anterior and posterior midgut invagination (endoderm formation) begins and is controlled by terminal group genes such as *tailless* and *huckebein* (CamposOrtega et al., 1985; Gilbert et al., 2000).

Invagination at the ventral furrow is initiated by the dorsal group genes, such as *twist* and *snail*. The interplay of *twist* and *snail* determine the mesoderm, an inner cell layer forming the later internal organs e.g. muscles, heart and the **fat body**. There are three different types of **muscles: somatic, visceral,** and **heart muscle**. They are under the control of *twist*, which acts as a master regulator on other transcription factors, such as *tinman* and *Mef2* (Furlong et al., 2001; Sandmann et al., 2007). Between the visceral musculature lies the fat body, whose development is promoted by *serpent* (Riechmann et al., 1998). The fat body is an organ specific to insects that retains endocrine and storage functions of the vertebrate liver. At the border of the meso- and ectoderm, *snail* suppresses mesectodermal genes as a transcriptional repressor, such as *single-minded* in the mesoderm (Kasai et al., 1992vb).

At the closure of the ventral furrow, the ventral midline appears, in which the neurogenic transcription factor *single-minded* is expressed. *Notch* signaling is also involved in the regulation of *single-minded* (MartinBermudo et al., 1995). This ventral region is also called mesectoderm and gives rise to the midline structures of the central nervous system. Laterally to the midline, at the neuroectoderm, **neuroblasts** are formed by the delamination from the surface epithelium. They move into the interior, building an orthogonal grid. Proneural genes, *achaete/scute* and *lethal of scute* are responsible for the development of neuroblasts (Skeath et al., 1994). In contrary, neurogenic genes, such as *Notch*, inhibit neuroblast formation and promote epidermal development (CamposOrtega et al., 1995). Neuroblasts divide asymmetrically in a stem cell-based manner and their later fate is determined in a cell cycle-dependent way (Fichelson et al., 2005jp). **Neuroblasts** express sequentially the transcription factors *Hunchback*, *Kruppel*, *Pdm1*, *Castor* and *Grainyhead* (Brody et al., 2002; Pearson et al., 2003). During lineage specification, layered domains of neuroblasts are formed expressing temporarily these factors. Neuroblasts also differentiate to glioblast and neurogliobasts, which produce glia and mixed neuronal/glial lineages respectively. The

binary switch of glia development is the transiently expressed transcription factor *glial cells missing* (*gcm*; Jones et al., 2005). *Gcm* promotes glial differentiation by activating *repo* and *pointed* and blocks the neuronal pathway via *tramtrack*. During gastrulation the embryo starts to be subdivided into parasegments and later into segments along the anterior-posterior axis that is defined by the hierarchical system of gap genes, pair-rule genes and segment polarity genes (NussleinVolhard et al., 1987).

2.1.2 Cell types of the fly head

The *Drosophila* head consists of seven segments, whose development differs from the trunk and is driven by the transcription factors *engrailed* and *wingless* (Schmidt-Ott et al., 1992). External parts of the head derive from imaginal discs, mainly from the eye-antennal disc (antenna, eyes and maxillary palps). The main part of the head is occupied by different compounds of the eye (**Figure 2.2**). The **Drosophila retina** is basically an array of 800 ommatidia, which contain the photoreceptor cells (Pichaud et al., 2001). Photoreceptors are neuronal cells sending axons to the optical lobe of the brain. Other cell types of the eye are cone and pigment cells as well as glia cells, which migrate along the axons of the photoreceptors.

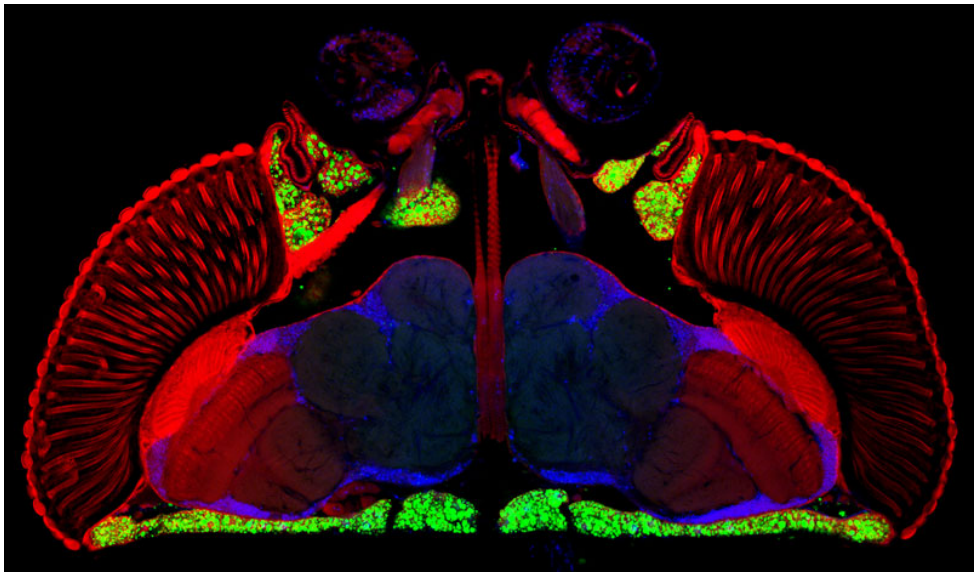


Figure 2.2 Cell types of the fly head.
The main cell types in the fly brain are neurons in the brain (blue), fat body cells (green) and parts of the eye (red). Image taken from http://www.nimr.mrc.ac.uk/research/alex-gould/

The ***Drosophila* brain** occupies the other major part of the head cavity. It consists of several sub-anatomical structures, ganglia and projections (see Flybrain Neuron Database; Shinomiya et al., 2011). Peripheral sensory inputs reach the first order neuropil regions such as the antennal lobe (olfactory pathway) and the optic lobe (visual system). Projection neurons from the glomeruli of the antennal lobe forward signals to higher centers including the **mushroom body**. The mushroom body is a complex structure consisting of the calyx (Kenyon cells), peduncle and the bifurcated lobes (Tanaka et al., 2008). The mushroom body is known to be involved in olfactory learning and memory (McGuire et al., 2001; Akalal et al., 2006; Davis et al., 2012). The higher-order center for visual information, orientation and locomotor control is the **central complex**, including the ellipsoid body, the fan-shaped body and other structures (Martin et al., 1999).

The head **fat body** surrounds the brain. The fat body plays a role in lipid metabolism, connecting metabolism to behavior and secreting hormones and pheromones to the hemolymph (Arrese et al., 2010). There is sex-biased expression of genes in the fat body regulating mating behavior and reproduction. Fat body cells (adipocytes) are separated from the brain with an insect-specific "blood-brain barrier" composed of glia cells (Stork et al., 2008).

2.1.2.1 Neurons

Neurons forming the *Drosophila* brain derive from about 100 embryonic neuroblasts (YounossiHartenstein et al., 1996). These are primary neurons, which wire the larval brain (Truman et al., 1990) and are remodeled during metamorphosis. Secondary lineages are adult-specific and have larval origin (Baek et al., 2009). The adult brain consists of about hundred lineages; the most characterized ones are the mushroom body and antennal lobe lineages (Ito et al., 1997). This diversity of neuronal cells is established and maintained by the combinatorial expression of transcription factors (Brody et al., 2002; Pearson et al., 2003) and signaling molecules including neurotransmitters and receptors. However, there are also ubiquitous, pan-neuronal marker genes, which are shared among all neurons. Usually these markers are used as Gal4 insertion lines (enhancer-trap; Brand et al., 1993), ensuring the neuron-specific ectopic expression of the gene of interest.

The best-characterized and most commonly used pan-neuronal marker is the ***elav*** gene (*embryonic lethal, abnormal vision*; Robinow et al., 1988), which encodes a nuclear-localized, RNA-binding protein involved in splicing. ***Elav*** is required for the differentiation and maintenance of neuronal cell fate. It regulates the alternative splicing of the neuron-specific isoforms of *neuroglian*, *erected wing*, and *armadillo* (Koushika et al., 2000; Lisbin et al., 2000; Soller et al., 2003). The *elav* promoter is commonly used as a Gal4 driver or antibodies against the ELAV protein in immunohistochemistry studies to specifically label neuronal cells in tissues (Robinow et al., 1991). One other often used, pan-neuronal Gal4 line carries the promoter of the *n-Syb* gene (Pospisilik et al., 2010). *n-Syb* (*neuronal Synaptobrevin*) is a vesicular SNARE protein playing a role in exocytosis (Sweeney et al., 1995). Loss-of-function mutations of *n-Syb* lead to slow neuro-degeneration (Haberman et al., 2012). Other well-characterized pan-neuronal genes such as the transcription factors *deadpan* and *asense* are only found in early development and are absent in adult tissues (Southall et al., 2009).

Most of the characterized neuronal genes play a role in neuron differentiation and are transiently expressed during development or are expressed only in a subsets of neurons. For example, during neurogenesis the timing of transcription factor binding and activation is very important (Maurange et al., 2008). Due to these temporal events, in many cases there is no or very low expression of these transcription factors in the adult head.

We can classify neurons according to their function and so by the **neurotransmitters** and **receptors** they express. The main excitatory neurotransmitter in *Drosophila* is **acetylcholine**, which is synthetized by the choline acetyltransferase (ChAT) and is bound by muscarinic and nicotinic acetylcholine receptors (AchR; Kolodziejczyk et al., 2008). These receptors are composed of several subtypes and subunits of which some are neuron-specific. The expression pattern of all different types of AchR has not been described yet. In contrary to acetylcholine, the main inhibitory system is modulated by **GABA**. This involves transporters, such as vGAT (vesicular GABA transporter), and GABA receptors, which are composed of the three subunits GRD, RDL, LCCH3 (Harrison et al., 1996; Kolodziejczyk et al., 2008). NMDA receptors respond to **glutamate** and have a role in complex behavior such as learning and memory (Wu et al., 2007). NMDA receptors fulfill this function in specific sub-

anatomical structures of the fly brain (including the mushroom body and the ellipsoid body; Wu et al., 2007). Other neurotransmitters are involved in behaviors, such as serotonin (**5-HT**) and neuropeptide F (***npf***) in aggression (Dierick et al., 2007) or **dopamine** in courtship and sleep (Liu et al., 2008; Andretic et al., 2005). Genes encoding enzymes producing these transmitters (i.e. TRH: tryptophan-hydroxylase, TH: tyrosine hydroxylase) are expressed in specific subsets of neurons (Coleman et al., 2005; FriggiGrelin et al., 2003). The TH-Gal4 is a generally used driver line used for marking dopaminergic cells. There are a several serotonin receptors (e.g *5-HT1A*) and dopamine receptors (e.g. *DopR*) encoded in the *Drosophila* genome. **Octopamine** (OA) is analogous to the mammalian noradrenaline. OA is synthesized by the enzyme tyrosine decarboxyalse (TDC) and is expressed in about 100 neurons in the *Drosophila* brain (Busch et al., 2009). One of the octopamine receptors, OAMB (octopamine receptor in the mushroom body) is found in the mushroom body and in the central complex (Han et al., 1998). Beside the classical neurotransmitters, *Drosophila* also has signaling peptides such as the short neuropeptide F (**sNPF**). sNPF is distributed in several diverse populations of neurons (Nassel et al., 2008) and has a neuroendocrine function linking behavior to feeding and growth (Lee et al., 2009). Another link between brain function and metabolism are insulin-like peptides (**dILP**s) produced by a specific subset of neurosecretory cells in the *pars intercerebralis* (Geminard et al., 2009).

Taken together, there are several genes shared among all neurons, there are specific subsets of genes expressed in defined sets of neurons and also genes that are enriched in seemingly diverse cell populations.

2.1.2.2 Glia

Glial lineages separate from the neuronal differentiation pathway by the activity of transcription factors such as *gcm*, ***repo***, *pointed* and ***tramtrack***. *Gcm* is transiently expressed during differentiation in the embryo and is absent from adult tissues (Jones et al., 2005). The most commonly used pan-glial marker is *repo* (reversed polarity). There are *repo*-driven Gal4 lines and antibodies against **REPO** available specific to glia cells. *Repo* encodes a nuclear-localized transcription factor that regulates glial differentiation, induces glial markers and in general maintains glial identity (Yuasa et al., 2003). *Repo* cooperates with other transcription factors such as *pointed* (Klaes et al., 1994) and activates the expression of *loco*, a regulator of G protein signaling

(Granderath et al., 2000). *Loco* co-localizes with *moody*, a G protein coupled receptor subunit, in the plasma membrane of surface glia cells; *moody*-driven Gal4 expression is also used to mark surface glia (Schwabe et al., 2005). *Repo* and *pointed* are activators of glial development, whereas *tramtrack* is a repressor of the neural cell fate in glia (Jones et al., 2005).

Glia cells can be classified according to their function and histological location (Edwards et al., 2010). At the surface of the neuropil, the perineurial and sub-perineurial glia are found, forming the **blood-brain-barrier** between neurons and the hemolymph (Awasaki et al., 2008). Contrary to surface glia, glia cells infiltrating the neuropil (neuropil glia) can be sub-grouped to ensheathing and astrocyte-like glia. The latter ones associate with synapses and express excitatory amino-acid transporters (EAAT-1 and EAAT-2; Rival et al., 2004). EAAT1 plays a crucial role in glutamate uptake and its inactivation leads to glutamate-mediated neuro-degeneration (Rival et al., 2004).

Glia cells are actively involved in mediating **behavior** (Jackson et al., 2008). The gene *ebony* is known to have phenotypes in pigmentation, vision and circadian behavior. *Ebony* is localized to *repo*-positive nuclei especially in the epithelial glia of the optic system (Suh et al., 2007). More complex behaviors such as long-term memory (LTM) formation also require glia cell function. The over-expression of a cathepsin encoding gene *crammer* in glia cells but not in mushroom body neurons decreases LTM (Comas et al., 2004). Glia cells are also important in maintaining metabolic homeostasis. There are glia cell-specific insulin-like peptides (dILP6) (SousaNunes et al., 2011) or the Apolipoprotein D homolog GLaz (Glial Lazarillo) was also found in glia regulating lipid metabolism.

Glia cells have a diverse function in neuron differentiation, axonogenesis, metabolism, insulation and neurotransmitter clearance (Edwards et al., 2010). Therefore, glia cells or subsets of glia cells express a variety of genes involved in these functions.

2.1.2.3 Fat body

The Drosophila fat body shares functions of the vertebrate liver and adipose tissue (Arrese et al., 2010). It is the major tissue involved in **energy storage** and utilization and also in the **hormonal regulation** of metabolism and feeding behavior.

Triglycerides (TAG) are stored in lipid droplets in fat body cells. In order to use these lipids, fat body cells express genes playing a role in **lipid mobilization**. The gene *brummer* encodes a TAG lipase, which is up-regulated upon fasting (Gronke et al., 2005). There are other putative TAG lipases such as *CG5966* and *CG6113* showing similar up-regulation (Fujikawa et al., 2009). The break-down of glycogen (sugar storage) depends on the enzyme glycogen phosphorylase and leads to the synthesis of trehalose (the main sugar in the hemolymph in insects; Arrese et al., 2010).

Metabolism is regulated via signaling pathways such as the **cAMP pathway** and the **insulin-like signaling** pathway. The cAMP signaling leads to the phophorylation and activation of the transcription factor CREB via the protein kinase A (PKA). Expressing a dominant negative isoform of CREB in the adult fat body reduces glycogen levels and increases lipid levels in the head (Iijima et al., 2009). Signaling molecules of the insulin-like pathway are insulin-like peptides (ILP) mainly produced by neurosecretory cells in the *pars intercerebralis* (see **section 2.1.2.1**). The insulin pathway branches to different kinase cascades such as the TOR and Akt kinases (Teleman et al., 2010). The main transcription factor phophorylated by Akt is FOXO (Puig et al., 2003). Unlike CREB, FOXO is active in an unphosphorylated form. FOXO decreases protein synthesis and growth by regulating the translation factor 4E-BP, and it also sensitizes the insulin pathway by binding to the insulin receptor gene (*InR*) (Gershman et al., 2007; Alic et al., 2011). FOXO enrichment at its target genes is increased upon starvation suggesting a role in regulating low blood sugar levels (Alic et al., 2011).

The fat body not only mediates **metabolism** but also links metabolism to **behavior**. The fat body-specific gene ***takeout***, a putative juvenile hormone (JH) binding protein plays a role in this process (SarovBlat et al., 2000; Meunier et al., 2007). *Takeout* has a rhythmic expression in circadian time, preferentially in the fat body shown by northern blot and Gal4 activity (SarovBlat et al., 2000; Dauwalder et al., 2002). *Takeout* mutants have altered feeding behavior and TAG levels (Meunier et al., 2007). Secreted, fat body-derived proteins encoded by *takeout* or *dissatisfaction* are also involved in sexual behavior (Dauwalder et al., 2002; Finley et al., 1998). Masculinization and feminization of the fat body alters mating, suggesting a role of fat body in courtship behavior (Lazareva et al., 2007). In summary, the fat body acts as fuel storage and plays a role in connecting metabolism to neuronal and glial function.

2.1.3 Gene regulatory networks

The key regulators of gene expression are **transcription factors** (TFs), whose combinatorial binding and activity turns on and off genes required for cellular differentiation and cell fate maintenance (Spitz et al., 2012). Transcription factors recognize and bind specific DNA motifs in so-called **cis-regulatory modules** (CRM; Lee et al., 2002; Wilczynski et al., 2010). These are 100-1000 bp DNA elements that serve as a platform for TF binding. The pattern of TF occupancy determines whether a close-by gene is active or repressed in a particular cell type and time. On the other hand, one CRM can regulate more genes in proximal and distal genomic location. The genome-wide binding events of hundreds of transcription factors form gene regulatory networks that are important in fine-tuning gene activity in a spatiotemporal manner.

2.1.3.1 Transcription factors

Transcription factors belong to the most studied class of proteins (Yusuf et al., 2012). They are subdivided in basal and specific transcription factors. Basal or general TFs are necessary for transcription initiation. Their binding site is at the promoter, close to the transcription start site (TSS), where they form the pre-initiation complex (PIC) with RNA polymerase II (Thomas et al., 2006; Lenhard et al., 2012). In contrast, specific transcription factors bind also to distal *cis*-regulatory elements and mediate gene transcription from distance (Spitz et al., 2012). As discussed above, they are involved in development, cell growth and cell fate maintenance, but also in responding to internal and external environmental changes.

Transcription factors usually contain a **DNA-binding domain** (DBD) and a **trans-activating domain**; some of them also a ligand binding domain (Latchman et al., 1997). The DBD recognizes specific DNA sequences and bind them with high affinity. The most common structural sub-types of DBDs are composed of helix-loop-helix, leucine-zipper, or Zn-finger motifs (Stegmaier et al., 2004). Trans-activation domains are responsible for protein-protein interaction with other transcriptional co-factors or with the RNA polymerase complex itself. These proteins forming complexes can recruit histone modifying and remodeling enzymes as well as histone chaperones that make the chromatin structure accessible (or inaccessible in case of repression) for the

transcription machinery (Spitz et al., 2012; see **section 2.2**). Thus, transcription factors directly or indirectly regulate gene transcription.

2.1.3.2 Cis-regulatory modules and networks

The stretches of DNA where transcription factors bind are called ***cis*-regulatory modules** (CRMs). These are modular units containing multiple binding sites of several transcription factors. CRM activity is predominantly determined by the expression level and binding affinity of TFs (Bonn et al., 2008; Wilczynski et al., 2010). Transcription factors cooperatively bind to the same CRM either via direct protein-protein interaction or via DNA-mediated interaction. Negative regulators may repress gene expression by inhibiting other TFs. The presence or absence of these repressors in the given cell type gives the possibility of spatial regulation.

Depending on their activity, CRMs can be **enhancer, silencer** or **insulator elements** (Spitz et al., 2012). Enhancers are bound by activating TFs and they therefore positively influence gene expression. Enhancers may be located far from the promoter of the regulated genes and they form loops by positioning enhancer elements close to the promoter physically. In *Drosophila*, enhancer trap lines (insertions of Gal4; Brand et al., 1993) are used to map the spatiotemporal expression pattern of the genomic locus. Silencer elements have the opposite function by binding negative or repressive TFs. Another way of regulation is to separate promoters from enhancers by insulator elements and binding proteins (e.g. CTCF, BEAF-32, CP190 and Su(Hw) in *Drosophila*; Negre et al., 2010; Srinivasan et al., 2012). Insulators block enhancer or even silencer activity, thus making gene expression independent of the particular element.

The interplay between CRMs forms complex ***cis*-regulatory networks** (CRNs; **Figure 2.3**). *Cis*-regulatory networks are under the control of **master regulators**, which are transcription factors that regulate other transcription factors (Bonn et al., 2008; Spitz et al., 2012). Master regulators are common in development, such as *twist* in mesoderm formation (Sandmann et al., 2007), or *gcm* in glia specification (Jones et al., 2005). The dynamic crosstalk of TFs determines the expression of the target genes ensuring the correct developmental process in space and time. *Cis*-regulatory networks are coordinated by feed-forward and feedback loops, which fine-tune the expression level, the binding affinity and transcriptional activity of TFs.

In this section, I gave an overview of gene activity regulation in *Drosophila* development, how the main cell lineages divide, what are the main cell types in an adult, differentiated organ (e.g. brain) and how transcription factors initiate and maintain cellular differentiation and identity.

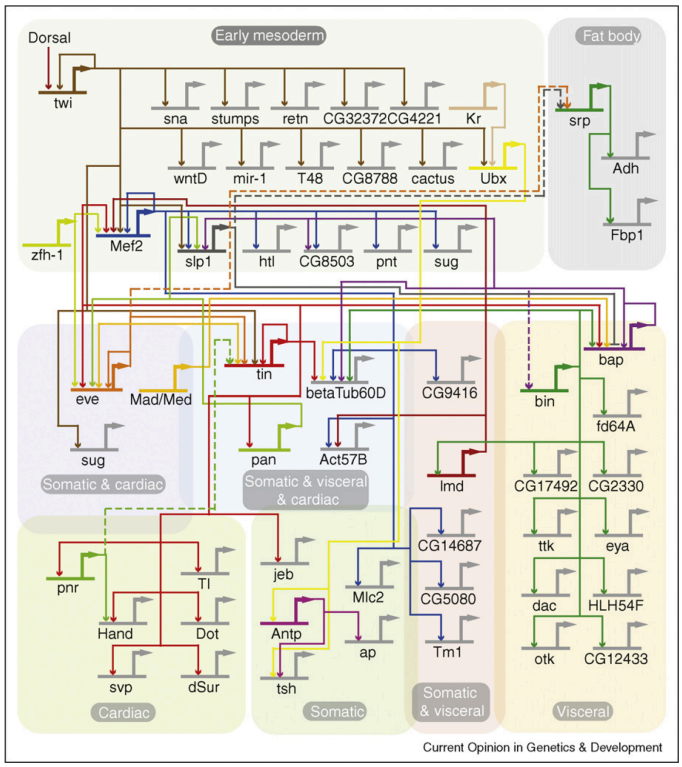


Figure 2.3 Summary of a *cis*-regulatory network in *Drosophila* mesoderm development.

In this example, there are 57 genes involved in the differentiation process on the level of CRMs. Image taken from Bonn et al., 2008.

2.2 Regulation of gene expression

The central dogma of molecular biology, postulated by Francis Crick (Crick et al., 1970), describes the still valid core of gene expression meaning the main information transfer is from DNA to RNA and from RNA to proteins. I am interested in how these processes differ between differentiated cell types of the *Drosophila* head. The cell-type-specific tools I present later in this thesis are novel approaches to dissect various regulatory steps of the central dogma.

As we have seen during development, transcription factors initiate the first steps of gene expression (see **section 2.1.3**). They recruit the general transcription factors, the RNA polymerase complex and other co-factors that allow the efficient RNA synthesis. In this chapter, I will cover the key steps of RNA polymerase II-mediated transcription that occur in a chromatin environment, where the DNA is packed (either

tightly or loosely) with histone and non-histone proteins. Additionally, the regulation of gene expression also occurs at the level of post-transcriptional events including RNA splicing and translation.

2.2.1 RNA polymerase II-mediated transcription

Transcription of genes to messenger RNA is mediated by the RNA polymerase II complex. RNA polymerase II is a **multi-subunit complex** consisting of 12 core subunits (see structure: Cramer et al., 2000; Cramer et al., 2001). The largest subunit, RPB1 contains the catalytic site for RNA synthesis and also a C-terminal domain (CTD) that plays a regulatory role. The CTD consists of 52 heptade repeats in humans and 44 in *Drosophila* from which serine and tyrosine residues get phosphorylated during the transcription cycle (Hirose et al., 2007). The second two largest subunits, RPB2 and RPB3 are involved in the complex assembly (Kolodziej et al., 1991). RPB4 and RPB7, a dissociable heterodimer mediates a step during initiation subsequent to promoter DNA binding (Orlicky et al., 2001). The 12-subunit RNA polymerase complex does not function on its own; general transcription factors (GTF), co-activators are needed to form even larger complexes such as the pre-initiation complex, the initiation complex and the elongation complex.

2.2.1.1 Transcription initiation and Pol II pausing

Transcription initiation occurs at the promoter regions of genes. According to cap analysis of gene expression (CAGE), promoters can be classified into broad and peak transcription start sites (TSS; Ni et al., 2010; Hoskins et al., 2011). Transcription at peak promoters begins from a narrow, few base pairs long position, whereas at broad promoters from a wide region. Peak promoters usually associate with defined core **promoter motifs** such as TATA box, Initiator (Inr) and DPE (downstream promoter elements). Temporal expression profiles throughout *Drosophila* development are highly variable among genes with peak promoters and genes with the lowest specificity mainly carry broad promoters (Hoskins et al., 2011). In addition, broad promoters usually have well-positioned nucleosomes architecture associated with histone modifications and variants (Nozaki et al., 2011; see **section 2.2.2**).

Since broad promoters in general lack core promoter elements, our knowledge of basal or general transcription factor binding to these is focused on peak promoters (Juven-Gershon et al., 2010). These elements are not universal and also not all basal transcription factors are general. *In vitro* purified mix of RNA polymerase II and the **basal transcription factors** TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH is able to mediate transcription from TATA box containing promoters (Lewis et al., 2005). This complex is called the pre-initiation complex (PIC; Thomas et al., 2006) and transcription initiation begins with its assembly at the promoter (**Figure 2.4**; Shandilya et al., 2012). In this process the first factor engaged to the core promoter is TFIID, which contains the TATA binding protein (TBP). Next, TFIIB gets recruited and stabilizes the ternary complex (Deng et al., 2007; Shandilya et al., 2012). The other large complex that helps Pol II binding in this early steps is the mediator complex. Efficient initiation also requires TFIIH in the process called promoter melting, where the two strands of DNA get separated at the TSS (Kim et al., 2000).

TFIIH, using its Cdk7 subunit phosphorylates the Ser-5 residues of RPB1 C-terminal domain (CTD; **Figure 2.4**; Buratowski et al., 2009; Shandilya et al., 2012). The **Ser5-P-CTD** serves as a signal for other factors, such as mRNA capping enzymes, which create the 5′ trimethylguanosine cap of short RNA fragments produced in the early elongation phase (Fabrega et al., 2003). The CTD Ser5 phosphorylation mark also recruits chromatin-modifying enzymes, such as the Set1 methyl-transferase, which establishes the active mark H3K4me3 (see **section 2.2.2**; Ng et al., 2003). This early elongation is inefficient and Pol II pauses after 25–50 nucleotide RNA synthesis (Rasmussen et al., 1993), mediated by the complex of DSIF (DRB Sensitivity-Inducing Factor) and NELF (Negative Elongation Factor; Yamaguchi et al., 1999). Analysis of short RNAs, produced at this stage, revealed that Pol II backtracks after the initial pausing to a thermodynamically stable state (Nechaev et al., 2010). The transcription factor TFIIS is required to keep Pol II in a transcriptional competent state and also functions as cleavage factor of the 3′ end of short transcripts (Adelman et al., 2006; Nechaev et al., 2010). Stalled Pol II stays in a paused state waiting for induction signals as in the case of heat shock (Rasmussen et al., 1993), or developmental signals (Muse et al., 2007; Zeitlinger et al., 2007). Arrested Pol II may undergo early termination (Nrd1/Nab3/Sen1 complex) instead of elongation (Buratowski et al., 2009; Terzi et al.,

2011). Active transcription occurs when the initiation complex turns into a productive elongation complex (**Figure 2.4**).

2.2.1.2 Transcription elongation and the CTD code

In order to release Pol II from the promoter paused state other **phophorylation** events occur at the Serine 2 residues of the CTD. This is catalyzed by the positive transcription elongation factor b (P-TEFb) Cdk9 kinase subunit (Peterlin et al., 2006). In addition, P-TEFb phophporylates NELF and DSIF that leads to their dissociation from the paused complex (Yamaguchi et al., 1999). Next, chromatin-modifying enzymes such as Set2, a H3K36 methyltransferase, are recruited to the **Ser2-P-CTD** ensuring the chromatin environment for active elongation (Hampsey et al., 2003; Selth et al., 2010; **Figure 2.4** and **section 2.2.2**). Other histone modifications including histone H3 acetylation by the GCN5 containing SAGA complex is also required for efficient transcription (Green et al., 2005). During Pol II progression nucleosomes may be displaced or rearranged by **ATP-dependent chromatin remodeling enzymes** and histone chaperones. The best-characterized remodelers involved in transcription are SWI/SNF, ISWI, CHD, and INO80/SWR (Becker et al., 2002; Clapier et al., 2009). Activity of these enzymes depends on histone modification e.g. the SWI/SNF family member RSC on acetylation (Carey et al., 2006). The most important histone chaperones mediating changes in chromatin structure during transcription are the histone H2A-H2B binding FACT complex (facilitates chromatin transcription; Belotserkovskaya et al., 2003; Hondele et al., 2013) and the elongation factor Spt6 (Kaplan et al., 2003). Other histone chaperones such as Nap1, Asf1, HIRA are also required for transcription-associated histone exchange (Avvakumov et al., 2011).

The phosphorylation status of the CTD, also called **the CTD code**, defines other co-transcriptional events such as splicing (Hirose et al., 2007; Egloff et al., 2008bf). The hyper-phosphorylated CTD recruits splicing factors and allows the assembly of the spliceosome. There are several other CTD modifications whose function is not completely clear so far. Ser7 phosphorylation was found on the gene bodies (Chapman et al., 2007) and is involved in snRNA transcription (Egloff et al., 2007). CTD Thr4 phosphorylation is required for histone RNA 3´ processing (Hsin et al., 2011) and Tyr1 phosphorylation activates binding of Spt6 elongation factor and interferes with the recruitment of termination factors (Mayer et al., 2012).

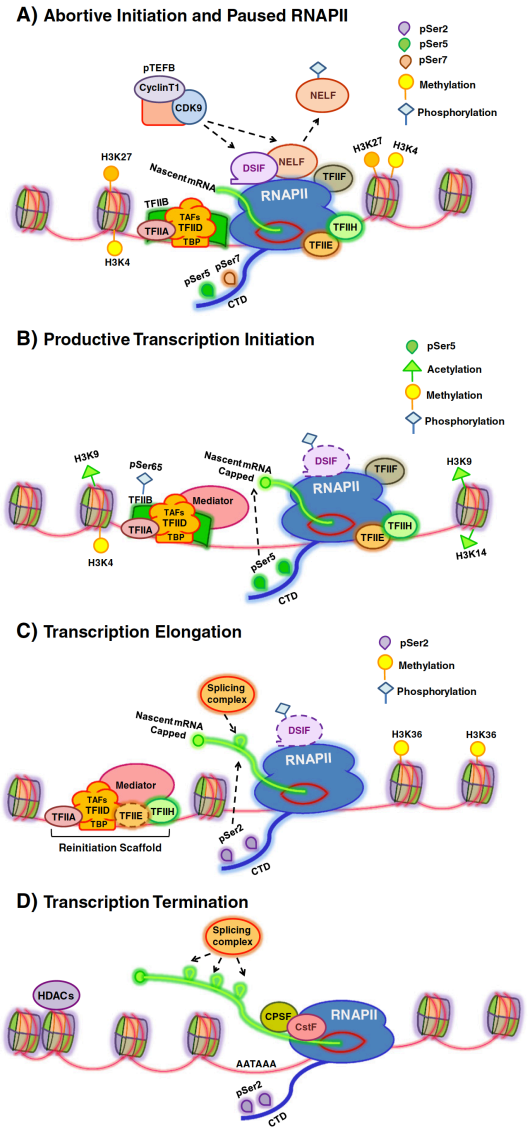


Figure 2.4 The key steps of RNA polymerase II mediated transcription.

A) Initiation and Pausing: pre-initiation complex assembly, CTD Ser5 phosphorylation, H3K4 trimethylation.
B-C) Elongation: CTD Ser2 phosphorylation, H3K36 trimethylation
D) Termination
Source: Shandilya et al., 2012.

2.2.2 Chromatin structure and function

In vivo transcription does not occur on "naked" DNA because DNA is packed into the highly organized chromatin structure in eukaryotic cells (Li et al., 2007). Chromatin is an array of nucleosomes consisting of an octamer of histone proteins (H2A, H2B, H3, H4) and 145-147 bp of DNA wrapped around (Luger et al., 1997). Packaging of DNA not only functions to compact such a long molecule into a relatively small nucleus, but gives the opportunity of regulating any DNA-related processes, such as DNA replication, repair (Groth et al., 2007; Dinant et al., 2008) and transcription (Li et al., 2007; Jiang et al., 2009). In this section, I will discuss three possibilities of chromatin regulation related to transcription. 1) The N-terminal tails and the globular domains of histones

are targets for **post-translational modifications** (PTMs; Kouzarides et al., 2007 and **section 2.2.2.1**). 2) Beside the canonical histone, **histone variants** are incorporated to nucleosomes (Talbert et al., 2010 and **section 2.2.2.2**). 3) Nucleosomes are not randomly distributed in the genome; they are actively positioned by **ATP-dependent remodeling enzymes** (Becker et al., 2002; Clapier et al., 2009) determining **higher order chromatin structure** (VargaWeisz et al., 2006; Sajan et al., 2012 and section 2.2.2.3).

2.2.2.1 Histone post-translational modifications

Since the 1960´s, when the first histone acetylation and methylation was reported (Allfrey et al., 1964), more than 60 different PTMs have been discovered (Kouzarides et al., 2007; Bannister et al., 2011). Histone arginine (R) residues can be methylated, lysines (K) methylated, acetylated, ubiquitinated, ADP-ribosylated, and sumoylated; as well as serines and threonines phosphorylated. Furthermore, there are examples of mono-, di- and trimethylation that give even more variations and complexity. Combinations of histone marks, such as synergistic effects of H3Ser10 phosphorylation and H3K9 acetylation and many other examples, led to the so-called **histone code hypothesis** (Strahl et al., 2000; Jenuwein et al., 2001) and models involving cooperative interactions, such as "binary switches" ((Fischle et al., 2003; Hake et al., 2006). The histone code enormously extends the regulation potential of gene expression by the combinatorial nature of histone modifications. The mechanism how the code is written lays on the cooperation of enzymes that create (**writers**) and remove (**erasers**) as well as proteins recognizing (**readers**) the modified residues by specific protein domains (Wang et al., 2007).

Histone acetylation, unlike other marks, occurs at multiple lysine sites both on H3 and H4, usually close to promoter regions (**Figure 2.5**). Acetylation alters the net charge of histone tails loosening DNA-nucleosome interactions (Zhao et al., 2005, Chandy et al., 2006). H4 acetylation on lysines 5, 8, and 12 are relatively non-specific marks and rather their cumulative effect determines transcription (Dion et al., 2005). Writer enzymes of acetylation are histone acetyl transferases (HATs) and erasers histone deacetylases (HDACs). One of the most characterized HAT complex involved in gene expression is the **SAGA complex**, member of the GNAT (Gcn5 N-acetyl-transferases) family (Lee et al., 2007). SAGA preferentially acetylates H3K9 residues,

which is mutually exclusive with the heterochromatin mark H3K9 trimethylation (Wang et al., 2008). H4K16 acetylation has specialized roles being involved in *Drosophila* dosage compensation (MOF histone acetyl transferase in the MSL complex; Akhtar et al., 2000). Histone acetylation is a dynamic process, several classes of HDAC (histone deacetlyase) enzymes (erasers) control the balance of acetylation state during development. HDACs are molecular targets of HDAC inhibitors, potential drugs in cancer therapy (Haberland et al., 2009). Readers of histone acetylation are globular protein modules and include – most notably – the bromodomain family of proteins (Dhalluin et al., 1999; Jacobson et al., 2000), which can recruit other histone modifying enzymes to acetylation sites (Ladurner et al., 2003; Kasten et al., 2004).

The other most common mark involved in transcription is **histone methylation**. As we have seen in section 2.2.1, the Set1 methyl-transferase is recruited to the Pol II CTD and modifies **H3K4** residues (Ng et al., 2003). Mono-, di- and tri-methylation shows distinct patterns over the promoter and gene body (**Figure 2.5**). Mono-methylation of H3K4 shows a wide distribution over active genes, whereas tri-methylation accumulates close to the transcription start sites (Pokholok et al., 2005). Set1 binds Pol II at the 5´ end of genes and also interacts with the PAF1 complex that is required for di- and tri-methylation (Krogan et al., 2003; Adelman et al., 2006). H3K36 di- and tri-methylation is enriched towards to the 3´ end of actively transcribed genes and is mediated by Set2 methyl-transferase (Rao et al., 2005). Specific readers of the **H3K36me** mark include EAF3, a chromodomain-containing subunit of the Rpd3S histone deacetylase complex, which stabilizes the amount of histone acetylation at ORFs (Carrozza et al., 2005; Joshi et al., 2005). Another chromodomain protein MRG15 is also specific to the H3K36me mark (Zhang et al., 2006) and recruits chromatin-remodeling complexes (i.e. Tip60), involved in histone variant (H2A.Z) incorporation (Kusch et al., 2004).

Histone methylation has the completely opposite role at **H3K9** residues (Bannister et al., 2001; Schotta et al., 2002), forming constitutively repressed **heterochromatin** (**Figure 2.5**). The writer enzyme for H3K9me3 is Su(var)3-9 and one of the readers is the chromodomain protein **HP1** (Jacobs et al., 2002; Ebert et al., 2006). HP1 maintains heterochromatin at spread out, large genomic regions (Cheutin et al., 2003) including pericentric heterochromatin and the chromocenter of polytene chromosomes in *Drosophila* (James et al., 1989).

The third group of histone methylation is related to **Polycomb group (PcG)** proteins, which have a key role in developmental silencing (Simon et al., 2009). PcGs bind to polycomb response elements (PREs) and recruit the PRC2 complex, which creates the H3K27me3 mark. PcG mediated silencing forms a regulated type of heterochromatin to control cell fate, development and cancer (Sparmann et al., 2006).

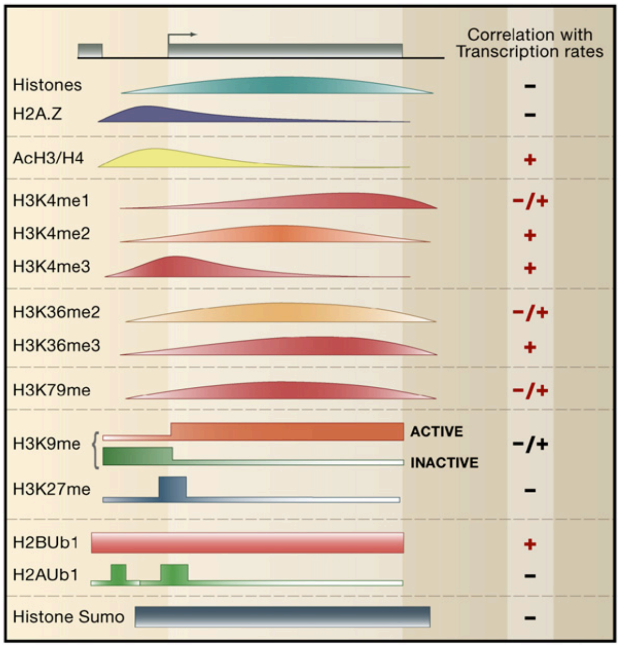


Figure 2.5 Histone modifications involved in gene regulation.

Active genes carry unspecific acetylation marks and specific methylation marks such as H3K4me and H3K36me. There are also histone variants (e.g. H2A.Z) incorporated to active genes. Image taken from Li et al., 2007.

2.2.2.2 Histone variants

Another level of gene regulation is the incorporation of histone variants, which differ in their primary amino acid sequence from the four canonical histones (Talbert et al., 2010). Histone variants are not encoded by genes forming arrays, but by a single copy (or few) gene(s) in the genome. The other common difference from canonical histones is that they are also incorporated into nucleosomes outside of S-phase, so are thought to be replication independent, and may have introns, leading to alternative splice products with distinct functions (note the NAD metabolite-binding and metabolite-insensitive vertebrate macroH2A.1 isoforms; Kustatscher et al., 2005). Histone variants are generally highly conserved, so probably function as "universal" histone variants in common regulatory mechanisms. There are also lineage-specific histone variants with (likely) very specific functions (Hake et al., 2006; Talbert et al., 2012).

Histone variants involved in transcription regulation, such as **H3.3 and H2A.Z**, are conserved among eukaryotes (Talbert et al., 2012). A common regulatory mechanism is that the dynamically exchange of histones and histone variants helps to

overcome the nucleosome barrier at actively transcribed genes (Petesch et al., 2012). These histone variants are re-placed by canconical as well as histone variant-specific chaperones (H2A.Z: Nap1, Chz1; H3.3: HIRA, Daxx) or by chromatin remodelers (H2A.Z: Swr1) during transcription (Avvakumov et al., 2011).

H3.3 differs only by four amino acids from the canonical H3; one residue is on the N-terminal tail, three residues on the core (Talbert et al., 2012). H3.3 containing nucleosomes are assembled either in a replication-dependent or independent way (Ahmad et al., 2002). The S-phase-independent assembly is mediated by the histone regulator A (HIRA) histone chaperone complex (Tagami et al., 2004). In this process, HIRA forms a complex with Asf1 (a general H3-H4 chaperone). Alternatively, H3.3 is associated with Atrx and Daxx proteins at telomeric regions in a HIRA independent way (Goldberg et al., 2010). **H3.3** nucleosomes are less stable compared to H3 nucleosomes shown by salt extractions (Jin et al., 2007). Genome-wide studies revealed that H3.3 is present at active genes having RNA polymerase II and H3K4me marks (Mito et al., 2005) and also at boundaries of *cis*-regulatory domains (Mito et al., 2007). Furthermore, Jin and colleagues found that H3.3/H2A.Z double variant nucleosomes are the least stable and are incorporated to "nucleosome-free" regions, such as active promoters, enhancers and insulator-bound regions (Jin et al., 2007; Jin et al., 2009). Although H3.3 co-localizes with H2A.Z at active sites, unlike in the case of H2A.Z, the distance of H3.3 nucleosomes from transcription start sites does not correlate with gene expression levels (Bargaje et al., 2012).

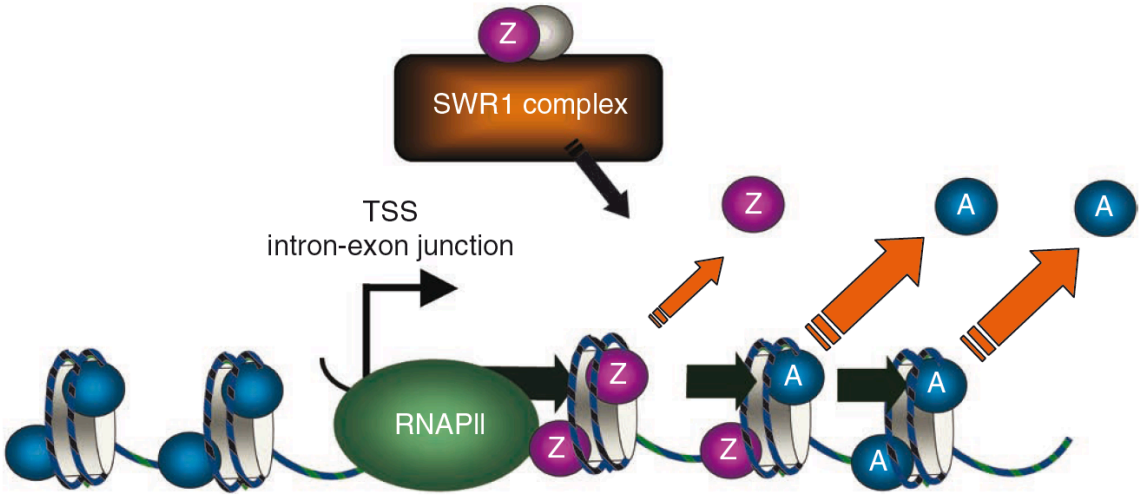


Figure 2.6 Histone variant H2A.Z is present at active genes carrying RNA polymerase II.
The histone variant H2A.Z is incorporated to nucleosomes close to active transcription start sites by the Swr1 remodeling complex. Figure taken from Weber et al., 2010.

H2A.Z is another well-studied histone variant that is involved in transcription (**Figure 2.6**; Li et al., 2007; Talbert et al., 2010; Petesch et al., 2012). H2A.Z forms a conserved class of H2A variants, as it is evolutionary distinguished from H2A.X and macroH2A variants (Talbert et al., 2012). The overall structure of **H2A.Z**-containing nucleosome is similar to H2A nucleosomes, however its incorporation might affect interaction between H2A.Z-H2B and H3-H4 dimers (Suto et al., 2000). Studies of the physical properties and stability of H2A.Z nucleosomes show contradictory results depending on the approach (Zlatanova et al., 2008). Sedimentation coefficient of reconstituted H2A.Z nucleosome particles depends on ionic strength, indicating reduced stability (Abbott et al., 2001). Zhang and colleagues confirmed this finding (Zhang et al., 2005) that H2A.Z is released from purified yeast chromatin more readily than H2A. In contrary, a FRET based, *in vitro* assay indicates that H2A.Z stabilizes nucleosomes (Park et al., 2004). In addition, Thakar *et al.* found by electrophoretic analysis subtle differences in compaction and stability of H2A.Z nucleosomes compared to H2A (Thakar et al., 2009). Immunoprecipitation studies revealed that H2A.Z-H3 containing nucleosomes are as stable as H2A-H3 nucleosomes particles, although interestingly H2A.Z-H3.3 nucleosomes are the least stable among all variations (Jin et al., 2007).

H2A.Z is essential *Drosophila* during development; mutant embryos fail proliferation and differentiation (van Daal et al., 1992; Faast et al., 2001). H2A.Z has contradictory roles in chromatin regulatory processes, such as gene activation, chromosome segregation and repression in heterochromatin (Zlatanova et al., 2008; Altaf et al., 2009). Reports on *Drosophila* polytene chromosomes revealed a non-random distribution of H2A.Z (Leach et al., 2000). Recent genome-wide studies revealed H2A.Z is preferentially bound to promoter regions and promotes RNA polymerase II recruitment (Barski et al., 2007; Mavrich et al., 2008; Hardy et al., 2009). In yeast, H2A.Z is incorporated to nucleosomes flanking the transcription start site at the 5´end of genes (Zhang et al., 2005; Raisner et al., 2005; Guillemette et al., 2005). In *Drosophila*, well-positioned nucleosomes carry H2A.Z mainly at the TSS +1 position and gradually decreasingly towards the 3´end of genes (Mavrich et al., 2008). H2A.Z nucleosomes often associate with paused RNA polymerase just in front of the +1 nucleosome (Mavrich et al., 2008). Human data, similarly to yeast, indicate H2A.Z occupancy in both upstream and downstream of the TSS (Schones et al., 2008) and the distance between H2A.Z nucleosomes and the TSS correlates with gene expression (Bargaje et al., 2012).

On the other hand, correlation between H2A.Z and gene expression levels is not that clear. On average, silent genes lack H2A.Z, whereas expressed genes carry H2A.Z, but there is no clear co-linearity between H2A.Z and RNA levels (Barski et al., 2007; Jin et al., 2009; Weber et al., 2010). H2A.Z nucleosomes at active promoters share additional features: they are homotypic for H2A.Z (Weber et al., 2010; **Figure 2.6**) and they also carry the histone variant H3.3 (Jin et al., 2009). In some cases, such as at the INO1 gene in yeast (Brickner et al., 2007) or at flowering genes in *Arabidopsis* (Deal et al., 2007), H2A.Z is present at repressed genes that can get rapidly reactivated. In addition, H2A.Z is quickly evicted from heat shock genes upon heat shock, probably ensuring a rapid activation (Zanton et al., 2006; Kumar et al., 2010; Kotova et al., 2011). These findings suggest that H2A.Z may play a role in "transcriptional memory" at genes with low expression, but that are ready to be strongly reactivated.

2.2.2.3 Higher order chromatin structure

In the previous sections, I discussed how gene expression is regulated by transcription factors, RNA polymerase II recruitment, histone modifications and incorporation of histone variants to nucleosomes. Transcription occurs in a chromatin environment where the nucleosomes are positioned by remodeling enzymes in an actively regulated manner (Becker et al., 2002; Clapier et al., 2009 and see their role in in transcription: **section 2.2.1.2**). However, chromatin is not only an array of individual nucleosomes on the DNA, but it is further organized into higher order structure (VargaWeisz et al., 2006; Li et al., 2011).

The classical types of chromatin are the open, actively transcribed **euchromatin** and the compact, silent **heterochromatin** (Heitz et al., 1928; see review: Grewal et al., 2007). Using genome-wide technologies, we can further distinguish between smaller chromatin domains. Epigenetic domains are functional domains carrying characteristic histone modifications and specific chromatin binding proteins (Filion et al., 2010; Kharchenko et al., 2010; Larson et al., 2010). Physical domains, called topological domains, are three-dimensional physical interaction patterns, defined by high-resolution chromosomal contact maps (Dixon et al., 2012; Sexton et al., 2012; Nora et al., 2012). There is a strong association between linear epigenetic domains and physical domain structure (Sexton et al., 2012).

Epigenetic domains were first defined by a comprehensive map of 53 chromatin associated proteins and four histone modifications in *Drosophila* cells using a DamID-based method (Filion et al., 2010). The authors identified - and named according to color - **5 functional domains** based on principal component analysis of these maps (**Figure 2.7**). Beside the two classical heterochromatic regions (GREEN: HP1-type [H3K9me3] and BLUE: Polycomb-type [H3K27me3]), they found another type of silent, gene-poor chromatin (BLACK) that carries no active histone marks (lack of H3K4me2 and H3K79me3) and genes on average with low expression. In addition, mRNA levels of expressed genes in BLACK chromatin show high deviation among dissected tissues in the so-called FlytAtlas (Chintapalli et al., 2007). Two different active chromatin domains could be distinguished by the histone modification H3K36me3 and by its associated chromo-domain protein, MRG15 (present in the YELLOW chromatin, absent in RED). Based on gene ontology analysis, the authors suggest that the YELLOW chromatin contains constitutively active whereas the RED dynamically active genes (**Figure 2.7**). These findings suggest that the characterization of chromatin is not restricted only to euchromatin and heterochromatin but there are other features that mark chromatin domains that are constitutive ("housekeeping") or regulated in different cell types and developmental stages.

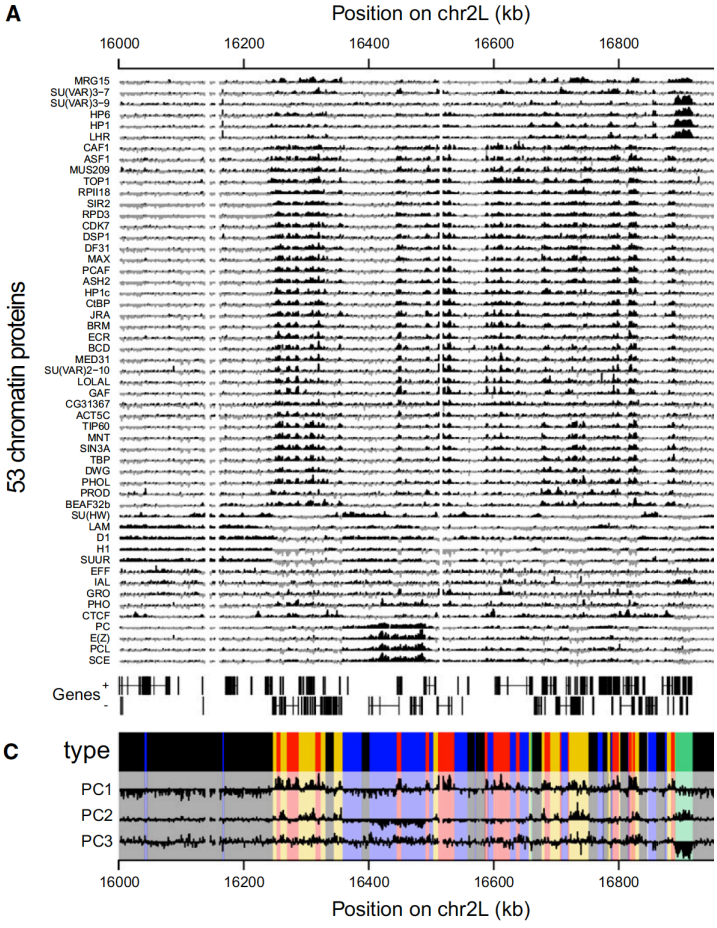


Figure 2.7 Chromatin is classified to epigenetic domains.

The 5-state chromatin domains were obtained by mapping 53 chromatin-associated proteins. YELLOW and RED chromatin carries proteins involved in active gene expression. BLUE chromatin is marked by repressive Polycomb group proteins, whereas GREEN chromatin by heterochromatin protein 1 (HP1). The inactive BLACK chromatin lacks most of the proteins investigated except lamin or histone H1. Source: Filion et al., 2010.

Another similar classification of chromatin domains was done in the **modENCODE** project (Kharchenko et al., 2010). They describe a **9-state model** based on ChIP-array of histone modifications and chromatin-bound proteins. In that model, active chromatin is subdivided into TSS-proximal (H3K4me3), active elongation (H3K36me3) and intronic regions with high H3K27ac. Additionally, there is a special type of chromatin domain involved in X chromosome hyper-activation in male *Drosophila* marked by H4K16 acetylation (Kharchenko et al., 2010; reviewed by Straub et al., 2011). They also find the H3K27me3 (PcG) and H3K9me3 (HP1) domains as distinct heterochromatin types. State 8 and 9 are similar to the BLACK chromatin in the 5-state model, except that state 8 carries moderate levels of H3K9me2/me3. Thus, the two independent approaches in *Drosophila*; and a third human study (Ernst et al., 2011); came to similar conclusions. However, some domains (e.g. different active domains) in the 9-state model can be merged into the 5-state model, making it simpler for comparisons.

Chromatin is located in the three-dimensional nuclear space carrying **physical interactions** between intra- and inter-chromosomal regions. The nucleus is organized into sub-compartments such as the nuclear periphery (e.g. nuclear envelope [Taddei et al., 2004; Pickersgill et al., 2006; Akhtar et al., 2007] and nuclear pore [Tran et al., 2006ds; Brickner et al., 2007]) or internal structures (e.g. Polycomb bodies [Buchenau et al., 1998; Saurin et al., 1998], transcription factories [Jackson et al., 1998; Osborne et al., 2004]) playing an important role in regulating transcription and gene activity (see review: Sexton et al., 2007). Early microscopy studies extended with chromosome conformation capture approaches (3C, 4C, Hi-C etc.) revealed genome-wide maps of proximal positioned genomic regions (Dekker et al., 2002; LiebermanAiden et al., 2009; see review Sajan et al., 2011). Sexton et al. published the first genome-wide **Hi-C map** from *Drosophila* embryonic nuclei (Sexton et al., 2012), where they found association of physical domains to epigenetic domains (5-state). This and a mammalian study (Dixon et al., 2012) investigated and confirmed that **insulator-binding proteins** are enriched at the boundaries of these domains. In *Drosophila* two classes of insulator-binding proteins were identified: class I (i.e. CTCF, CP190 and BEAF-32) were found at domain boundaries (Negre et al., 2010; Schwartz et al., 2012; VanBortle et al., 2012) and class II such as Su(Hw), which is associated with the gypsy retro-transposon insulator (Adryan et al., 2007).

Topological domains are probably involved in separating co-regulated genes. Dixon *et al.* also found that not only insulator binding proteins, but also **housekeeping genes** are enriched at domain boundaries (Dixon et al., 2012). These are genes expressed independently of cell type and developmental stage. Housekeeping genes are usually short, carry few exons (DeFerrari et al., 2006) and have "broad" promoters (Hoskins et al., 2011). Ubiquitous housekeeping genes are clustered also by sequence (Lercher et al., 2002; Weber et al., 2011) forming "genomic platforms" of co-regulation. In contrary, cell-type-specific genes are outside of these clusters, in gene-poor, dynamic regions. In order to understand how ubiquitous and specific genes are differentially regulated, novel cell-type-specific tools have to be developed.

2.2.3 Post-transcriptional regulation

In order to investigate gene expression and to compare which genes are active in distinct cell types, one has to study post-transcriptional events as well. Usually the dynamic range of mRNA levels are broader compared to the occupancy of chromatin-associated factors. As an example, it is very hard to predict whether a small change in paused RNA polymerase II peak height would alter mRNA levels. There are several ways to profile total RNA pools, the **transcriptome**, or specific parts of the transcriptome. One subset of total RNA is the actually transcribed, **nascent RNA** (Core et al., 2008; Churchman et al., 2011). We can distinguish between short, nuclear RNA that is most probably produced by stalled RNA polymerase and longer nascent RNA molecules associated with the elongating RNA polymerase (Nechaev et al., 2010). Another large set of RNA is bound by the ribosome involved in translation. **Ribosome-associated RNA** is also called the **translatome**, which may differ from the whole transriptome (Halbeisen et al., 2008).

There are several methods applied to map and quantify nascent transcription such as 1) **GRO-Seq**, global run-on sequencing (Core et al., 2008), 2) isolating short, nuclear transcripts (Nechaev et al., 2010), or 3) **NET-seq**, native elongating transcript sequencing (Churchman et al., 2011). At the same time, other studies looked at the translatome by profiling ribosome-bound RNA (polysome profiling [Fu et al., 2012; Paredes et al., 2012], ribosome affinity purification [RAP: Halbeisen et al., 2009],

translating ribosome affinity purification [**TRAP**: Heiman et al., 2008; Doyle et al., 2008; Dougherty et al., 2010; Thomas et al., 2012]). In this section, I will summarize the findings about transcriptomic and translatomic approaches that allow us to investigate post-transcriptional regulation and can be potentially used for cell-type-specific studies.

2.2.3.1 Nascent transcription

Chromatin immunoprecipitation of transcription factors, chromatin modifications or the RNA polymerase II complex itself hint at gene activity. RNA polymerase enrichment predicts whether a gene is active or ready to get activated (e.g. paused Pol II at the heat shock genes). However, RNA polymerase peaks at the promoter proximal regions do not indicate active elongation (Nechaev et al., 2011). To evaluate promoter-proximal pausing, Core *et al.* established a method called global run-on-sequencing (**GRO-seq**) assay to profile transcriptionally engaged polymerase (Core et al., 2008). Using a nuclear run-on assay, nascent RNA molecules can be extended and labeled with an BrU analog. About hundred nucleotides extension by the transcriptionally engaged polymerase generates BrU-incorporated RNA for sequencing. According to these studies genes can be subdivided into classes of active, not paused polymerase; active, paused; inactive, paused and inactive with no RNA polymerase (Core et al., 2008). The limitation of GRO-Seq is that it cannot be used for complex organisms, because the BrU incorporation is only possible in cell culture or thin tissues.

Another way to investigate active elongation is to isolate RNA polymerase II-associated RNA molecules. A recent approach **NET-Seq** (native elongating transcript sequencing) is based on high-throughput sequencing of the 3´ends of nascent transcripts produced by Pol II in the yeast *S. cerevisiae* (Churchman et al., 2011). In principle, RNA is co-purified with the elongating RNA polymerase complex by immunoprecipitation. A FLAG-tagged Pol II subunit, Rpb3 was expressed endogenously in yeast and pulled-down using anti-FLAG affinity gel. NET-Seq not only revealed active transcription, but also gave insights of divergent transcription and Pol II pausing in the gene body. The method should be suitable for multicellular organism as well, by expressing a tagged Pol II subunit in a cell-type-specific manner. cDNA obtained by NET-Seq needs to be amplified to yield enough material for sequencing.

Recent techniques such as GRO-Seq and NET-Seq solved the problem of detecting active elongation by focusing on the product of transcription, the nascent

RNA. However, these methods were not yet used in a cell-type-specific manner to compare gene activity in cell types.

2.2.3.2 Ribosome-associated RNA: the translatome

Most of the functional genomics studies focus on the dynamics of transcription by profiling the **transcriptome**. On the other hand, gene expression leads to changes at the protein levels as well. A way to study this step of gene expression is to map ribosome-associated RNA, what we can call the **translatome**. Correlations between the transcriptome and translatome were high upon severe stress, but very low upon mild stress induction in yeast (Halbeisen et al., 2009). Thus, the post-transcriptional response appeared prior to changes in global transcript levels in yeast. Inducing the EGF pathway in human cells also showed that the transcriptional and translational response is extensively uncoupled and controlled mainly at the translational level (Tebaldi et al., 2012). In addition, polysome profiling from mouse liver identified dynamic changes of the translatome upon food deprivation (Fu et al., 2012). Therefore profiling ribosome engaged RNA might reveal many more changes upon perturbation than by profiling global transcription alone.

Ribosome-associated RNA profiling was also applied in a cell-type-specific manner in an approach called **TRAP** (Translating Ribosome Affinity Purification; Heiman et al., 2008; Doyle et al., 2008; Sanz et al., 2009). A tagged subunit of the ribosome (GFP-L10A) was cell-type-specifically expressed in the mouse brain using BAC clones (bacTRAP) and immunoprecipitated from the cell type of interest (Heiman et al., 2008). Applying the technique on 24 cell populations in the mouse central nervous system, Doyle *et al*. reported thousands of specific mRNAs that were not detected by whole tissue microarrays (Doyle et al., 2008). In order to identify these transcripts statistical approaches such as the specificity index were developed (Dougherty et al., 2010). The method was used not only in mammalian system, but recently was also reported in *Drosophila* (Thomas et al., 2012). Combining the TRAP with the versatile UAS/GAL4 expression system gave the possibility to study the translatome in hundreds of distinct cell populations in *Drosophila*, as I have done in my own thesis project.

Thus, translatomic studies revealed that several changes of gene expression occur on the post-translational levels. Methods such as TRAP are suitable tools to study

the translatome in a cell-type-specific manner and thus identifying transcripts hidden in whole tissue map.

2.3 Cell-type-specific approaches to dissect gene activity

One of the key challenges in molecular biology is to understand how genes are regulated in different cell types of multicellular organisms. Cell fate and cell identity is determined by sets of cell-type-specifically expressed genes across development or in terminally differentiated cells (see **chapter 2.1**). Expression of transcription factors, co-factors, chromatin modifying and remodeling enzymes is broad; their function might be pleiotropic. In contrary, combinatorial binding of these factors to a chromatin environment in a spatio-temporal manner ensures the tight regulation of gene expression (see **chapter 2.2**). Most of the chromatin and gene expression profiling studies have been focusing on cell culture or whole organism based methods, which lack the resolution of different cell types. During the course of my PhD, new cell-type-specific approaches, combined with high-throughput technologies (microarrays and sequencing), have been developed to dissect gene activity in several cell types within an organism, within complex tissues. In this chapter, I will give an overview of novel methods using physical or biochemical isolation of cell types for chromatin and mRNA mapping.

2.3.1 General workflow

Cell-type-specific gene expression profiling methods consist of at least five common steps: 1) labeling of the cell type of interest, 2) isolation of labeled cells, 3) profiling of gene expression, 4) data analysis, 5) validation of cell type specificity (**Figure 2.8**).

In general, labeling uses a transgenic construct expressing a tagged fusion protein (in many cases fluorescently), which allows also the visualization of the required cell population (**Figure 2.9**). The tagged protein sometimes is a nuclear protein (e.g. histone) or a protein localized to the nuclear envelope marking whole

nuclei. In other cases, the tagged protein is a functional part of the given protein machinery, such as the RNA polymerase II complex or the ribosome (**Figure 2.9**)

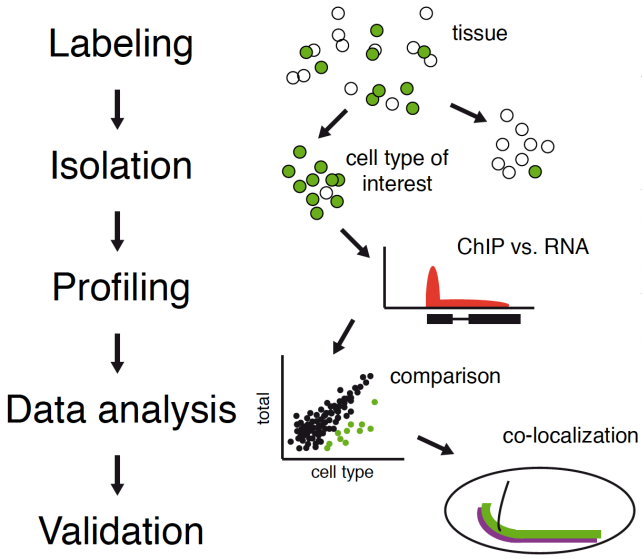


Figure 2.8 General workflow of cell-type-specific genomics.

Cell-type-specific genomic methods share the following steps: labeling of the cell type of interest, isolation of the labeled cell population, profiling gene activity (by ChIP- or RNA-Seq), finding the differences among cell types (data analysis) and validating cell type specificity.

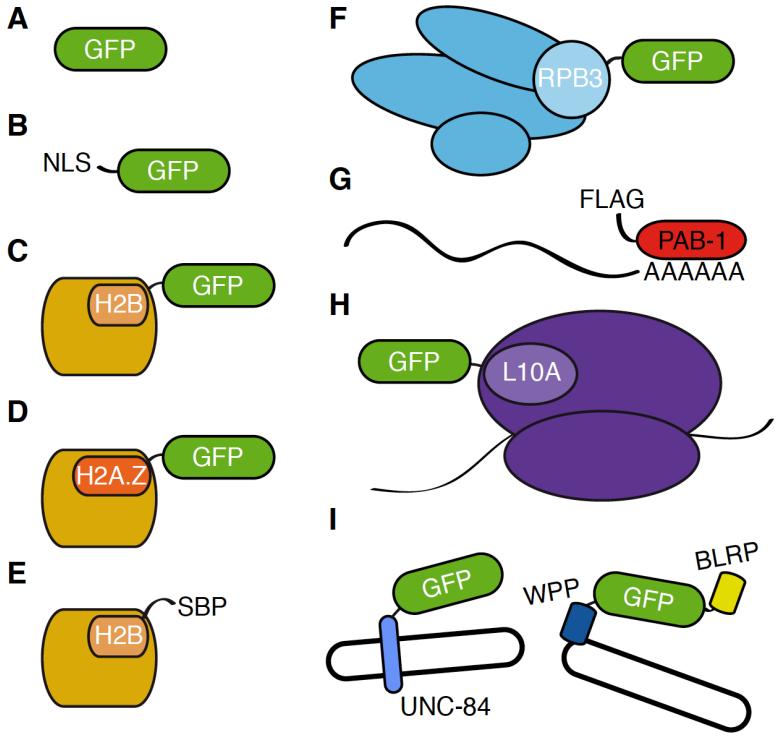


Figure 2.9. Summary of transgenic constructs for labeling cell types

Cell types can be labeled by **A)** green fluorescent protein (GFP), **B)** nuclear localized (NLS) GFP or by histone-GFP fusion protein **C)** H2B-GFP, **D)** H2A.Z-GFP, histone with a streptavidin binding peptide tag **E)** SBP-H2B) or by a fusion protein part of a protein complex such as **F)** the RNA polymerase II complex (GFP-RPB3), **H)** the ribosome (GFP-L10A) or by RNA binding proteins **G)** polyA binding protein (PAB-1) or by **I)** nuclear envelope targeting proteins (UNC84-GFP and RanGAP(WPP)-GFPP-BLRP).

The isolation is then based on the fluorescence of the reporter, immunostaining of specific proteins (no tagging) or affinity purification by the tag. The labeled cell population can be separated physically including manual dissection or FACS sorting, and also biochemically such as affinity purification of whole nuclei, chromatin-bound or RNA-associated proteins. The methods discussed here can be distinguished whether they focus on chromatin mapping or RNA profiling (**Table 2.1**). Different approaches require various treatments (crosslinking, protease dissociation etc.) or different amount of biological material, time and labor (**Table 2.2**).

Method	Full name	Reference
Manual	Manual	Hempel et al., 2007; Nagoshi et al., 2010
LCM/LDM	Laser- capture/directed microdissection	Karsten et al., 2002; van Deerlin et al., 2002 Chung et al., 2005; Rossner et al., 2006
FACS-array/ MAPCeL	Fluorescence activated cell sorting microarray	Lobo et al., 2006; Fox et al., 2005, 2007; Marsh et al., 2008; Spencer et al., 2011
FANS	fluorescence-activated nuclei sorting	Haenni et al., 2012
FACS-ChIP	Fluorescence activated cell sorting chromatin immunoprecipitation	Jiang et al., 2008; Weake et al., 2011 Cheung et al., 2010
BiTS-ChIP	Batch isolation of tissue-specific chromatin immunoprecipitation	Bonn et al., 2012
PAB RNA IP	polyA binding protein RNA tagging immunoprecipitation	Roy et al, 2002; Yang et al., 2005; van Stetina et al., 2007; Spencer et al., 2011
TRAP	Translating ribosome affinity purification	Heiman et al., 2008; Doyle et al., 2008; Mustroph et al., 2009; Dougherty et al., 2010; Thomas et al.,2012
INTACT	Isolation of nuclei tagged in specific cell types	Deal and Henikoff 2010; Steiner et al., 2012; Henry et al., 2012
CAST-ChIP	Chromatin affinity purification from specific cell types	Schauer et al., unpublished

Table 2.1 Summary of cell-type-specific methods

Methods discussed in this chapter are indicated by short name, full name and relevant references.

Method	Application	Labeling	Type	Crosslinking	Protease	Species
Manual	RNA	FP	physical	no	yes	fly
LCM/LDM	RNA	FP	physical	both	no	mouse
FACS-array	RNA	FP	physical	no	yes	mouse, worm
FACS-ChIP	ChIP	FP / IF	physical	both	no	mouse, human
BiTS-ChIP	ChIP	TP+IF	physical	yes	no	fly
PAB RNA IP	RNA	TP	biochemical	no	no	worm, fly
TRAP	RNA	TP	biochemical	no	no	mouse, plant, fly
INTACT	both	TP	biochemical	both	no	plant, worm, fly
CAST-ChIP	ChIP	TP	biochemical	yes	no	fly

FP = fluorescent protein
IF = immunofluorescence
TP = tagged protein

Table 2.2 Classification of cell-type-specific methods

Methods can by classified according to the application they are used for (RNA, ChIP), the type of labeling (FP = fluorescent protein, IF = immunofluorescence, TP =tagged protein), the type of separation (physical or biochemical), different treatments used (crosslinking and protease) and species applied in.

2.3.2 Chromatin mapping-based methods

Several reports showed that cell-type-specific chromatin maps uncover gene and enhancer activity, therefore being suitable tools to study gene activity in distinct cell types (Weake et al., 2012; Bonn et al., 2012). Experimentally, one advantage of ChIP-based assays is the relative easy purification of labeled nuclei compared to whole cells with complex morphology (e.g. neurons). In such cases, specific cell populations cannot be isolated from tissues without damage, while their nuclei remain intact.

2.3.2.1 FACS sorting nuclei for ChIP

One of the most common physical separation approaches for obtaining cell-type-specific chromatin uses fluorescent activated cell sorter (**FACS**) to isolate fluorescently labeled nuclei. The cell-type-specifically expressed tag is either a fluorescent protein (GFP; Weake et al., 2012) or a nuclear protein (e.g. histone) fused to GFP (H2B-GFP; Jiang et al., 2008) or to a short tag that can be used for immunofluorescence staining (SBP-H2B: **BiTS-ChIP**; Bonn et al., 2012). Alternatively, nuclear localized, cell-type-specific proteins can be stained directly (e.g. NeuN; Jiang et al., 2008). The method has been used for ChIP with native as well as cross-linked (formaldehyde fixed) chromatin. Unfixed nuclei keep their integrity less well, however fixed nuclei can easily form

clumps (Jiang et al., 2008). It is crucial to avoid clumping in the FACS sorter, because labeled and non-labeled nuclei can get mixed and are thus not separated. In addition, crosslinking excludes gene expression changes during sorting, strengthens protein-DNA interactions, and may reduce potential nucleosome loss and degradation. The disadvantage of FACS is the amount of time and biological material required. ChIP-Seq experiments need several million nuclei to obtain enough IP- purified DNA for sequencing. Bonn *et al.* estimated that ~8 hours of sorting yields ~40 million mesoderm nuclei from *Drosophila* embryos with 97% purity (Bonn et al., 2012). Jiang *et al.* reported up to ~50% loss of mouse, neuronal nuclei during the experimental procedure (Jiang et al., 2008).

Thus, FACS sorting is a suitable approach for cell-type-specific ChIP profiling, but it requires a lot of material, is time consuming and needs specialized, expensive equipment that may not be sufficiently available to a particular researcher.

2.3.2.2 Biochemical tagging for ChIP

Alternatively, biochemical tagging has been applied to isolate either labeled nuclei (Deal et al., 2010) or chromatin-bound proteins (our own work; Schauer *et al.*, *accepted*).

Several groups used a method called "isolation of nuclei tagged in specific cell types" (**INTACT**) in various species (Deal et al., 2010; Steiner et al., 2012; Henry et al., 2012). The method is based on tagging nuclei on the outer surface and pulling down whole nuclei only from the cell type of interest. In the original study, a relatively complicated construct was used, where a nuclear envelope targeting domain was fused to GFP (visualization) and a biotin ligase recognition peptide (BLRP, acceptor peptide; Deal and Henikoff 2010). In parallel, an *E. coli* biotin ligase (BirA) was expressed that biotinylates BLRP. The tagged nuclei were then affinity purified using streptavidin-coated magnetic beads in a homemade column system. In *Drosophila,* this approach has been combined with the bipartite UAS/Gal4 system allowing the labeling of a large variety of cell populations. In addition, Henry *et al.* modified the method to a double GFP fusion construct (Henry et al., 2012). Interestingly, they used a *C. elegans* nuclear envelope protein (UNC84-2XGFP) in *Drosophila* system, because other fusions failed to show the required sub-cellular localization. The latter GFP-based method reached a purity of 99%, with ~50% yield, whereas the biotinylation method was estimated to give 93-96% purity. The INTACT approach was used for both chromatin profiling

(MNase mapping, Steiner et al., 2012; histone modifications H3K4me3, H3K27Ac and H3K27me3; Henry et al., 2012) and for sequencing of nuclear RNA (see **section 2.3.3**).

Isolating intact nuclei needs special care and experiments that are performed relatively quickly. One way to avoid potential artifacts, including protein and DNA degradation, is fixation of the tissues immediately after homogenization. This makes the purification of nuclei difficult, but preserves any changes during chromatin preparation. Another method, Chromatin Affinity Purification from Specific cell Types - ChIP (**CAST-ChIP**), which I have developed together with Carla Margulies in my PhD, has been used to directly pull-down cross-linked, chromatin-bound GFP-tagged proteins that are expressed in a cell-type-specific manner (see **chapter 3**; Schauer *et al., accepted*). The method is a combination of the UAS/Gal4 system with GFP-ChIP followed by high throughput sequencing. CAST-ChIP was applied on RNA polymerase II (RPB3) binding (**chapter 3**) and histone variants (e.g. H2A.Z; **chapter 4**) in *Drosophila* adult tissues. CAST-ChIP is a rapid, efficient method and requires regular amounts of biological samples. On the other hand, it cannot yet be used for histone modification to map active chromatin marks. For this to be possible, we would need to tag canonical histones, coupled to a release of the cell-type-specific chromatin, which we are currently carrying out using both GFP-TEV and GFP-precission protease tagged canonical histones (unpublished data from the Margulies team). Nonetheless, CAST-ChIP provides a highly suitable tool to profile RNA polymerase II-enriched regions that reflect the spatial expression of genes within a highly complex organ, such as the adult fly brain.

2.3.3 RNA profiling-based methods

Measuring RNA levels is a more direct way of profiling the output of transcription. RNA is more abundant in cells compared to DNA, making it easier to obtain the required amounts. On the other hand, only a few percent of the total RNA molecules are actually coding mRNAs, which requires special library preparation methods for sequencing. To work with RNA is in general harder because its sensitivity to degradation and no fixation can be used prior to RNA isolation. Since RNA is localized both to the nucleus and the cytoplasm, most of the methods aim for whole cell isolation. Despite of these issues, there are a large variety of approaches to purify RNA from distinct cell types.

2.3.3.1 Manual sorting

As other methods, **manual isolation** also relies on fluorescent labeling (Hempel et al., 2007; Nagoshi et al., 2009). The setup is relatively simple, no special expensive equipment is required. First, cells have to be dissociated by protease treatment from the tissue and plated on a Petri dish. Under a fluorescent stereomicroscope the labeled nuclei are picked using a pipette. Only healthy and properly shaped cells are collected controlling the sample quality. Although the approach sounds simple, the protease treatment is a potential stress for the cells. Usually in these studies a specific group of neurons is isolated that are separated from their connection network (Hempel et al., 2007; Nagoshi et al., 2009), exposed to completely different environment during the procedure. Since neurons are intact at this stage, the changed environment might change their signaling pathways, leading to non-desired gene expression changes. At least thirty neurons were collected from mouse brain samples (Hempel et al., 2007) and about hundred from *Drosophila* (Nagoshi et al., 2009). Dealing with such little amount of material requires RNA amplification steps before profiling with microarrays or sequencing is possible. Any amplification raises the question of linearity, and rare transcripts may remain undetected.

2.3.3.2 Laser directed microdissection

Instead of dissociating cells mechanically, precise dissection techniques have been developed using laser capture (**LCM**) or laser directed micro-dissection (**LDM**). The samples are either flash frozen or preserved in formalin, such as human post mortem tissues. Fixing with formaldehyde can cause RNA fragmentation and therefore interferes with RNA quality and reproducibility (Karsten et al., 2002; VanDeerlin et al., 2002). In earlier studies, specific cells were labeled by quick immuno-staining that required fixing (Chung et al., 2005). Rossner *et al.* published a strategy where GFP was expressed in mouse neurons and they applied a fluorescence guided LDM-based separation without fixing and staining (Rossner et al., 2006). They used freeze-dried cryo-sectioning followed by micro-dissection and obtained high quality RNA for microarray analysis. This and other systems (from Leica) reached the accuracy of a micron, but contamination from neighboring cells (e.g. glia) or missing axon branches has to be expected (for a comparison see Okaty et al., 2011). This is true especially in

the nervous system where tightly packed cells (neurons and glia) are in close proximity any kind of cross-contamination should be avoided.

2.3.3.3 FACS-array

Fluorescent activated cell sorting was initially used for RNA profiling (Lobo et al., 2006; Fox et al., 2005; Fox et al., 2007). Instead of fixed nuclei, protease-dissociated whole cells are isolated and either directly sorted or cultured before sorting. The two options were compared in *C. elegans* embryonic samples and thousands of differentially expressed genes were obtained including elevated levels of proteasome subunits upon 24h culturing (Fox et al., 2007). However, **FACS-array** studies on mammalian neurons mainly used the direct sorting method (Lobo et al., 2006; Marsh et al., 2008). Using this method 98% purity can be reached with a yield of 100.000 cells and a range of 10-100 ng total RNA. FACS-array studies, as the name indicates, used microarrays for profiling. Recently, Haenni *et al.* applied also a sorting-based approach on uncross-linked nuclei (**FANS**: fluorescence-activated nuclei sorting), followed by sequencing of amplified 3' end regions of the mRNA transcripts (Haenni et al., 2012).

The amount of RNA obtained by these approaches requires amplification. Although high throughput sequencing technologies requires only about 100 ng of total RNA, this is still the lower limit needed for mRNA-Seq. Ideally, amplification should be avoided in order to obtain full-length transcripts and to stay within the linear range. The solution could be to increase the amount of sorting material or the pooling of more technical replicates. This significantly increases the time and labor required to obtain enough RNA from FACS sorted samples.

2.3.3.4 Biochemical tagging for mRNA profiling

Tightly packed tissues make it difficult to isolate whole cells with physical separation. Alternatively, isolated nuclei could be used as in the case of **INTACT** (discussed in **section 2.3.2.2**; Deal et al., 2010dh). One disadvantage of RNA purified from specific nuclei is the potential difference of cytoplasmic vs. nuclear mRNA pool. In the nucleus, there might be an overrepresentation of unspliced native transcripts or short RNA products of early transcription (Nechaev et al., 2010). Nonetheless, enrichment of known neuronal transcripts was reported by using INTACT in the *Drosophila* brain (Henry et al., 2012).

Direct tagging and affinity purification of RNA binding proteins followed by RNA isolation seems to overcome this problem. There were two targets mainly used in these studies, either using polyA binding proteins (**PAB**) to enrich polyA mRNA or using **ribosome tagging** to pull down "translating" mRNA from specific cells. The PAB-based method was reported first on *C. elegans* muscle cells (Roy et al., 2002) and was applied for *Drosophila* photoreceptors (Yang et al., 2005) and continued on *C. elegans* neurons (VonStetina et al., 2007) and thirty different cells and developmental stages (Spencer et al., 2011). No other PAB studies were published in *Drosophila*, possibly due to the lethality of the transgene when expressed under the control of various Gal4 drivers (Yang et al., 2005).

Instead of polyA binding proteins, whole ribosomes were tagged using the Translating Ribosome Affinity Purification (**TRAP**) approach in the mouse (Heiman et al., 2008; Doyle et al., 2008; Dougherty et al., 2010) and, recently, also in *Drosophila* (Thomas et al., 2012 and **chapter 5**). TRAP was used to profile several cell types in the central nervous system where mRNA transport and localized translation are thought to be very important. Since the method relies on ribosome binding to mRNA, it actually gives information about the pool of mRNAs being translated. An important step is to block new translation initiation during the procedure by using translation inhibitors such as cycloheximide. The technique is well established in mouse; there are several "bacTRAP" lines available that are specific to cell types (Doyle et al., 2008). TRAP was also combined with the powerful UAS/Gal4 system of *Drosophila* and was narrowed down to about 200 cells in the *pars intercerebralis* (Thomas et al., 2012). Enrichment ratios over the total tissue or comparing cell types to each other are quite high (up to 100 fold), making the method robust for statistical analysis.

The advantage of these biochemical methods is that they are rapid and simple without requiring the need of special equipment. Instead of sorting nuclei for many hours (Bonn et al., 2012), the immunoprecipitation takes no longer than an hour and can be carried out in a cooled bench environment. I have therefore developed a *Drosophila* TRAP approach in order to profile the translatome of *Drosophila* adult cell types. Such a readily applicable, efficient tool will allow us to identify differential gene expression changes to environmental perturbation in distinct cell populations.

2.3.4 Data analysis

Analyzing microarray and high throughput sequencing data is a quickly developing field and is not a topic of this thesis in detail. In order to identify differences between cell types, differential expression analysis (similar to different conditions) has to be applied either on identified features of ChIP-Seq, or on exon/transcript annotations for RNA-Seq. There are several packages to compare data counts at these features including Cuffdiff (Trapnell et al., 2010; Trapnell et al., 2012), baySeq, DESeq (Anders et al., 2010) and edgeR (Robinson et al., 2010) and the most recent BitSeq (Glaus et al., 2012). These tools of the Bioconductor project or other packages are publicly available (http://www.bioconductor.org/).

The principle of analyzing cell-type-specific datasets is either to the compare cell type to the total or cell types to each other. The "total" sample might be an easily dissectible body part (whole head of *Drosophila*), organ (whole brain) or a more specific tissue (striatum in mouse brain). Finding significant enrichment over the total is difficult in cases where high percentage of the labeled cell types is part of the whole (e.g. neurons of the fly brain; granule cells of the cerebellum), because the two profiles may be too similar. In that case, genes that are depleted from the given cell type but present in the total can be identified. To find the differences among cell types, cell-type-specific enrichments have to be determined by pair-wise comparisons. Dougherty *et al.* suggested a filtering procedure on TRAP data to solve these issues (Dougherty et al., 2010). Furthermore, they suggest a specificity index, which is the average rank for each gene coming from all possible pair-wise comparisons. The output of these analyses is a list of cell-type-specific genes that helps to understand how these cells maintain they function.

2.3.5 Validation of cell type specificity

There are two ways to evaluate whether a gene is specific for the cell type found by ChIP or RNA profiling: 1) computationally, by comparing to existing datasets, 2) experimentally, by imaging gene activity in the particular cell type.

Most of the studies use a computational validation such as gene ontology (GO) analysis. GO analysis is a relatively quick and easy way to see terms enriched in cell-type-specific datasets, and therefore, evaluate whether there are genes with known characteristic function. Indeed, studies showing this type of analysis find GO terms typical for the cell type e.g. ion channel activity in neurons, *etc*. (Henry et al., 2012). On the other hand, one has to be careful with GO terminology, since the annotation is far from being complete. Another valid comparison is to compare datasets obtained by the cell-type-specific method to existing data that were generated by dissecting whole organs or part of tissues (insect neurons vs. whole brain). In *Drosophila*, a compendium of microarray data using dissected tissues from different developmental stages (FlyAtlas; Chintapalli et al., 2007) has been published. Although the dissected cell population might be quite heterogeneous, it can give a good estimate of whether a gene is enriched in the particular tissue or cell type.

Further, some of the reports validate their findings by co-staining of the cell-type-specific gene product and a known cell type marker. Bonn *et al.* used fluorescent *in situ* hybridization to test the co-localization of a *Drosophila* mesoderm marker (Mef2) and RNA of mesoderm specific genes identified by BiTS-ChIP (Bonn et al., 2008). RNA *in situ* analysis confirmed their predicted active regulatory regions determined by cell-type-specific Pol II occupancy and histone modifications. In some cases, however, it is not possible to compare cytoplasmic RNA with nuclear localized marker in cell types that have diverse morphology (e.g. comparing RNA with the ELAV marker in *Drosophila* neurons). Instead of direct staining of the gene product, nuclear GFP reporters can be used driven by the cell-type-specific promoter or enhancer. In fact, I have used this approach. The *Drosophila* UAS/Gal4 system is ideal for such experiments, since there are thousands of available Gal4 insertion lines (enhancer-trap) that reflect the gene activity of the given locus. The third option is to perform co-staining of the protein encoded by the cell-type-specific gene with a marker protein. However, antibodies are available only for a limited set of *Drosophila* proteins, so the staining approach may be limited by antibody availability.

2.3.6 Perspectives

In this chapter, I summarized several new techniques to dissect gene activity in specific cell types. The different methods can be characterized by specificity and sensitivity. Specificity is the measure of how pure the isolated cell population is, whereas sensitivity shows the minimum number of cells that are required from a given biological sample. Physical separation methods reached almost 100% purity of the labeled cells. On the other hand, despite the automated sorting, they are time and material consuming, making it difficult to obtain enough samples for high throughput profiling without PCR amplification. The longer the treatment and protocol, the more artifacts can occur. Sorting of fixed nuclei for ChIP sequencing solves the problem of altered gene expression throughout the procedure. Although chromatin mapping can predict gene expression, the precise measure of gene activity is mRNA sequencing. There are biochemical methods (TRAP) that can be up-scaled or carried out in many technical replicates to pool the required amounts of RNA. The starting material for gene expression profiling should be ideally sufficient, requiring no prior PCR amplifications to avoid potential bias. Any biochemical enrichment has its drawback, meaning it has to be compared to the background (or Input) and therefore it is not 100% specific. The choice from the available methods is a balance of specificity and sensitivity and trying to exclude bias.

In this thesis, I present biochemical approaches to study gene activity in *Drosophila* cell types. I developed and applied CAST-ChIP, a rapid and robust cell-type-specific chromatin profiling approach (see **section 2.3.2.2**, **chapter 3** and **6.1**). CAST-ChIP uses cross-linked material to preserve the gene expression state of cells and exclude potential stress. CAST-ChIP is carried out on a cooled bench environment and does not require special equipment such as a FACS sorter. CAST-ChIP is an efficient procedure; it does not need more biological material and experimental time than standard ChIP protocols. CAST-ChIP is easily applicable to several chromatin-bound proteins (e.g. RNA polymerase II in **chapter 3** and H2A.Z in **chapter 4**) and distinct cell populations of *Drosophila* (e.g. neurons, glia, fat body in **chapter 3**). Since it relies on the ectopic expression of a tagged chromatin reporter, it cannot be used for transcription factors that have low expression levels, that cannot be tagged without

disrupting biological function, or (in its present form) for profiling post-translational modifications of histone and other chromatin-bound proteins.

As next generation sequencing methods quickly evolve in a competitive field, there are easier, more cost-effective and standardized protocols ready for use by the scientific community. The required amount of starting material is decreasing, as better sequencing platforms appear on the market. As an example, few years ago 10 µg total RNA was needed for standard RNA-Seq. This has now been scaled down to a few 100 ng. RNA-Seq experiments have to deal with a large proportion of ribosomal RNA in the total RNA pool. Nowadays, there are several options to remove ribosomal RNA (e.g. polyA selection, RiboMinus, RiboZero). Alternatively, the sequencing depth/coverage (number of reads over the genome) is getting so high, that rRNA can be included to the sequencing. In the meantime, single cell transcriptomes have been reported (Tang et al., 2009; Tang et al., 2011), which are based on amplification of picogram amounts of RNA. One has to be careful whether the amplification is linear, does not have bias for special sequences, the transcripts are complete in length and whether rare transcripts are also represented. Nevertheless, the technology is constantly developing to make cell-type-specific studies easier and more precise.

2.4 Aims of the thesis

To be able to dissect gene regulatory mechanisms within complex tissues of *Drosophila*,

I aimed to establish and validate a novel cell-type-specific chromatin profiling method, CAST-ChIP to map

1) RNA polymerase II occupancy (**chapter 3**)

2) the incorporation of the active histone variant H2A.Z (**chapter 4**)

and to adapt a translatomic approach, TRAP, for distinct *Drosophila* cell types to profile

3) ribosome-associated mRNA (**chapter 5**)

3 CAST-ChIP – Chromatin Affinity purification from Specific cell Types

This project was performed in a bioinformatics collaboration with Petra Schwalie (EBI, Hinxton, UK) and with the help of Carla Margulies (LMU Munich), as will be noted.

3.1 Summary

Cell fate and identity is determined by the epigenetic regulation of gene expression throughout development. Generating cell-type-specific, genome-wide gene activity profiles helps us to understand

1) what makes distinct cell populations different from each other,
2) how cell-type-specific genes are regulated,
3) how ubiquitous, house-keeping genes are maintained in most cell-types.

Several approaches have been reported to map cell-type-specific gene expression (see **chapter 2.3**). Most of these techniques are time-, labor- and material-consuming; they use protease treatment (potential stress); some methods require amplification steps (potential bias) or special equipment for extended periods of time (FACS sorter).

In my PhD project, I developed a rapid and sensitive method called CAST-ChIP (Chromatin Affinity Purification from Specific cell Types). CAST-ChIP uses the powerful repertoire of the *Drosophila* UAS/Gal4 expression system combined with the chromatin immunoprecipitation (ChIP) of a tagged reporter expressed in the cell type of interest. In this chapter, I will introduce the method as applied on the profiling of RNA polymerase II in the three major cell types of the *Drosophila* head: neurons, glia and the fat body. I validated the CAST-ChIP-seq results using computational and experimental approaches. Cell-type-specific RNA polymerase II enrichment reflects the spatial expression pattern specific to the labeled cell type. CAST-ChIP provides a compendium of active genes marked by RNA polymerase II from cell types such as neurons or glia, relevant for neurobiology, or the fat body, relevant for metabolism research.

3.2 Introduction

The fruit fly is a suitable model system to study gene activity in a cell-type-specific manner. *Drosophila* consists of complex organs and tissues e.g. the fly brain contains more than 100.000 neurons. The complexity of the *Drosophila* central nervous system ensures complex behaviors such as feeding, mating or learning and memory (see **section 2.1.2**). In order to understand such processes, it is crucial to study which genes are active in these cell types. *Drosophila* has powerful genetics that allows the expression of reporters in defined sub-population of cells. There are thousands of GAL4 driver lines available either as genomic insertions or as cloned promoter/enhancer fragments to control the cell-type-specific expression of a tagged gene from the UAS promoter (upstream activating sequence). Thus, I chose *Drosophila* to express cell-type-specific chromatin reporters using transgenic lines and to profile gene activity in the cells of interest.

Transcription is mediated by the RNA polymerase II (Pol II) complex (see **chapter 2.2**) and thus mapping genome-wide Pol II binding gives a read-out of gene activity. Pol II accumulates at transcription start sites (TSS) and its level positively correlates with gene expression levels (Barski et al., 2007; Sultan et al., 2008). Accumulation of Pol II at the TSS seems to be a rate-limiting step in transcription since Pol II 1) stays paused and produces short transcripts (Nechaev et al., 2010); 2) turns into elongation and continues transcription (Adelman et al., 2005; Guenther et al., 2007; Muse et al., 2007; Zeitlinger et al., 2007); and 3) is back-tracked or terminated from the pausing point (Nechaev et al., 2010). New techniques such GRO-Seq revealed (see **section 2.2.3**; Core et al., 2008) that actually paused Pol II is able to produce full-length transcripts in human cells. Comparison of Pol II peaks and GRO-Seq in *Drosophila* suggests that the vast majority of Pol II is engaged and competent for transcription (Core et al., 2012). Thus, RNA polymerase occupancy at promoters is a suitable tool to monitor gene activity.

In order to profile gene activity in distinct cell types, I established a protocol to ChIP a tagged Pol II subunit expressed in the three major cell lines of the *Drosophila* head (neuron, glia, fat body, respectively).

3.3 Methods

3.3.1 Experimental procedures

Experiments in this chapter were performed by Tamás Schauer

3.3.1.1 Fly stocks

Flies were kept on standard media at 25 °C. 1-3 days old flies were collected and frozen in liquid nitrogen at the same time of the day. Frozen flies were stored at -80 °C until used for chromatin preparation. The following strains were used for ChIP experiments: 2202U2 (wild-type) (Boynton et al., 1992), *elav*-GAL4 (Bloomington stock no. 458), *repo*-GAL4 (Sepp et al., 2001), *take-out*-GAL4 (Dauwalder et al., 2002), c147-Gal4 (Bloomington stock no. 6979) and UAS-EGFP-RPB3 (Yao et al., 2006).

GAL4 lines selected for validation in immunohistochemistry: NP-lines (from DGRC): NP-0003, NP-1321, NP-3310, NP-2575, NP-4234, NP-4240, NP-5412, NP-7181 (Hayashi et al., 2002); as well as C0015 (barbos-1; Dubnau et al., 2003), D0417 (ruslan; Dubnau et al., 2003), *Eaat1*-GAL4 (Rival et al., 2004) and GH146 (Bloomington stock no. 30026).

3.3.1.2 Immunohistochemistry and western blot

Brains of 1-3 days old flies were dissected in PBS, fixed, and stained, as described previously (Wu et al., 2006). Fixed brains were incubated for two-three days at 4°C with primary and secondary antibodies, respectively. The antibodies used for immuno-histochemistry were anti-GFP (Torrey Pines TP-401, 1:200), anti-ELAV (DHSB 9F8A9 and 7E8A10, 1:200), anti-REPO (DHSB 8D12, 1:100), anti-rabbit Alexa-488 (A-11034), anti-mouse Alexa-568 (A-11031) and anti-rat Alexa-647 (A-21247). Three-four brains of each Gal4 line were imaged with a confocal laser-scanning microscope (LSM-710, Zeiss). All staining images show a representative brain of each line. Imaging data were processed using ImageJ and Adobe Creative Suite package.

Polytene chromosomes were prepared and stained as described previously (Pile et al., 2002). UAS-GFP-RPB3 was driven by the salivary gland driver c147-Gal4 and stained using anti-GFP (TP401, 1:200) and anti-RPB1 (7G5 Euromedex, 1:300) antibodies.

Western blot was carried out as described in **section 5.3.1.2.**

3.3.1.3 Chromatin immunoprecipitation and DNA sequencing

Heads of frozen flies were separated using 630 µm and 400 µm sieves. 400 – 600 fly heads were homogenized in homogenization buffer (HB) at 4 °C [HB: 350 mM sucrose, 15 mM HEPES pH 7.6, 10 mM KCl, 5 mM MgCl2, 0.5 mM EGTA, 0.1 mM EDTA, 0.1% Tween, freshly completed with 1 mM DTT and Protease Inhibitor Cocktail (PIC) (Roche)]. The homogenate was fixed using 1% formaldehyde for 10 minutes at room temperature and then quenched with glycine. The tissue debris was filtered with 60 µm nylon net (Millipore). Nuclei were collected and washed with RIPA buffer at 4 °C (RIPA: 150 mM NaCl, 25 mM HEPES pH 7.6, 1 mM EDTA, 1% Triton-X, 0.1% SDS, 0.1% DOC, freshly completed with PIC). For the fragmentation of chromatin a 2 step sonication was used: Branson250 (7 cycles, intensity 5, pulsing 16 sec) and Covaris sonicator (PIP175, DC20, CB20, time 4 min). After sonication debris was collected and chromatin was stored at -80°C. Fragment size was checked after cross-link reversal on agarose gel.

10-15 µg chromatin was used for each ChIP assay. Dynabeads protein G (Invitrogen) were equilibrated in RIPA buffer with 1 µg/µl salmon sperm DNA and 1 µg/µl BSA. Chromatin was always pre-absorbed with beads without antibody. The cleared chromatin was incubated with antibodies overnight. The following antibodies were used for ChIP in this study: anti-GFP (goat polyclonal affinity purified IgG, Ladurner lab stock), anti-RPB3 [(Adelman et al., 2006) and rabbit polyclonal IgG Ladurner lab stock], anti-RPB1 (7G5, Euromedex). After IP, beads were washed four times with RIPA and once with LiCl wash buffer (250 mM LiCl, 10 mM Tris-Hcl pH 8.0, 1 mM EDTA, 0.5% NP-40, 0.5% DOC, freshly added protease inhibitors). Beads were resuspended in TE buffer and incubated overnight at 65 °C. RNA was degraded using RNase A (Fermentas) for 30 min at 37 °C and proteins were digested with proteinase K (10 mg/ml) at 55 °C for 1,5 hours. Immunoprecipitated DNA was purified using Qiagen MiniElute columns and 3-10 ng DNA was sent for deep sequencing. All samples were at least duplicated through biological replicates. In some cases, technical replicates were pooled to obtain the required amount of DNA.

The following primers were used in ChIP-qPCR analysis: Fas2 +28 forward GTGCTCTGCTTGCTGAGAGA, reverse GCCACGACCGTTAACACATA; Fas2 +271 forward CTCCTCTGCAGCTGCTCTTT, reverse TTCGTGCGTTTGGGTTCTAT; CrebB-17A -502

forward TTTCTGTGAAACAGCCGATG, reverse TCGCTCGCCTAGTGATGTAA and Ubx PRE F4 (Papp et al., 2006).

Library preparation and sequencing was performed by the EMBL Genomics Core facility using standard Illumina protocols. 36bp single run was used on the Illumina GenomeAnalyzer IIX. All ChIP-Seq was done in two biological replicates. As control, one Input lane was sequenced. In addition, two biological replicates of ChIP using anti-GFP antibody on wild-type samples were also sequenced.

3.3.2 Data analysis

Data analysis was performed by Petra Schwalie (European Bioinformatics Institute, Hinxton, UK; Schauer et al. accepted).

3.3.2.1 Sequence alignment and peak-calling

Reads were aligned to the *D. melanogaster* genome (BDGP5.13) using Bowtie version 0.12.7 (Langmead et al., 2009). All sequence, genome annotations and comparative genomics data were taken from Ensembl release 57. Aligned reads were filtered for duplicates, uncalled bases (maximum 3 Ns were allowed) and low complexity (reads with stretches of > 20 identical bases). Only chromosomes 4, X, 3L, 3R, 2L and 2R were considered in the analyses. Aligned reads were transformed in coverage files using igvtools and visualized in the IGV 2.0 Browser (Robinson et al., 2011) and IGB Browser (Nicol et al., 2009).

Regions of high ChIP enrichment were detected with CCAT 3.0 (Xu et al., 2010) on individual replicates using Input or ChIP against GFP. Regions occurring in more then one replicate at a reported FDR rate <=0.05 were merged with regions occurring in single replicates at an FDR <=0.01. All replicate information was used to restrain the region and gain maximum resolution. Regions from different cell types were merged together to obtain a set of regions of interest (ROIs) for 1) the neuron-glia comparison (Head-Neuron-Glia) and for 2) the neuron-glia-fat body comparison (Head-All).

3.3.2.2 Differential ChIP enrichment analysis

Count information in individual ROIs (1 and 2) were compared using DESeq ver. 1.8.2 (Anders et al., 2010), determining significant differences in ChIP enrichment between pairs of cell types. The cutoff was defined at the significance level of 0.01 (after multiple

testing correction) except for fat body comparisons, where 0.001 was used, because of the higher variation among replicates. Additionally, we required a fold-difference of at least 1.8 between estimated mean counts. All Venn diagrams shown are created with the Vennerable 2.1 R package (Swinton et al., 2009) and are based on these cutoffs. The number of overlapping regions resulted from the intersection of the two or three regions to be compared.

Spearman correlations were calculated based on read counts inside corresponding ROI categories. Values shown in **Figure 3.9** are means of the two replicates using the "Head-Neuron-Glia" ROI sets.

3.3.2.3 Annotation analysis

ROI localization with respect to Ensembl genome annotation was measured by calculating the fraction of total ROIs that overlap a TSS, a protein-coding gene, a non-protein coding gene, or that are located at <=5kb respectively >5kb distance of any gene. Pie charts of these distributions are shown in **Figure 3.8**.

GO term enrichment analysis was performed using the Bioconductor topGO ver. 2.8.0 package and all *Drosophila melanogaster* genes as background. Only genes occurring >=15 times were included in **Table 3.1**. To test the enrichment of brain relevant genes in our neuronal, glial or head datasets, we used all central nervous system (CNS)-active genes previously collected (Pfeiffer et al., 2008). We tested whether these CNS-active genes were overrepresented among genes bound by (a) neuron-specific Pol II ROIs; (b) glia-specific Pol II ROIs; (c) invariant Pol II ROIs or (d) no Pol II ROIs. In **Figure 3.11**, we report -log10(p-values) for this enrichment based on the one-sided Fisher's exact test.

3.3.2.4 Expression analysis

Probe-level expression values available from FlyAtlas (Chintapalli et al., 2007) were used to determine the cell type specificity of genes overlapping neuron-, glia-specific or common Pol II ROIs. Expression levels (ln(probe-set expression values+1)) were displayed as heatmaps (**Figure 3.13**) and standard deviation of probe-level values in **Figure 3.12**.

3.4 Results

3.4.1 Cell-type-specific gene activity maps using CAST-ChIP

To investigate gene activity in cell types, I chose the adult *Drosophila* head as a model system. The fruit fly head contains the major part of the central nervous system, the brain (see **section 2.1.2**). The brain consists of two basic, distinct cell types; neurons and glia cells (**Figure 3.1).** There are well-characterized tools available to label this cell types within the organism. Neurons can be stained with the pan-neuronal marker, ELAV or by the expression of a reporter using the promoter of the *elav* gene. Glia cells can be specifically marked by the REPO protein or by the *repo* promoter. The two cell populations, however, both derive from the ectoderm, do not overlap with each other and have distinct function (see **section 2.1.2** and **Figure 3.1**).

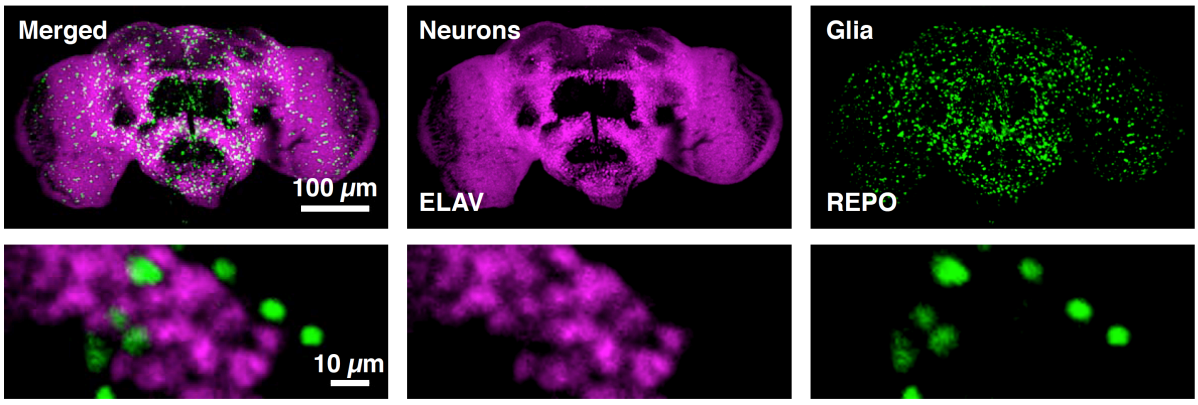


Figure 3.1 The two major cell types of the fly brain are neurons and glia.
Immunostaining of neurons (ELAV [magenta]) and glia (REPO [green]) shows two distinct cell types within the fly brain. Bottom panel indicates a zoom-in image of the top panel.

I sought to find differences in ChIP profiles of chromatin-associated proteins in neurons and glia. In order to do this, I established a method that I refer to CAST-ChIP: Chromatin Affinity Purification from Specific cell Types followed by Chromatin Immuno-Precipitation, and then subjected to high-throughput sequencing. CAST-ChIP is based on the expression of a tagged chromatin-associated reporter encoded by a transgene under the control of the UAS (Upstream Activating Sequence) promoter (**Figure 3.2**). The expression of the GFP-tagged reporter gene (GeneX-GFP) is ensured by the Gal4 transcription factor, which is controlled by a cell-type-specific enhancer (Gal4). In this

setup the chromatin reporter is only present in the cell type of interest and absent from other cell types.

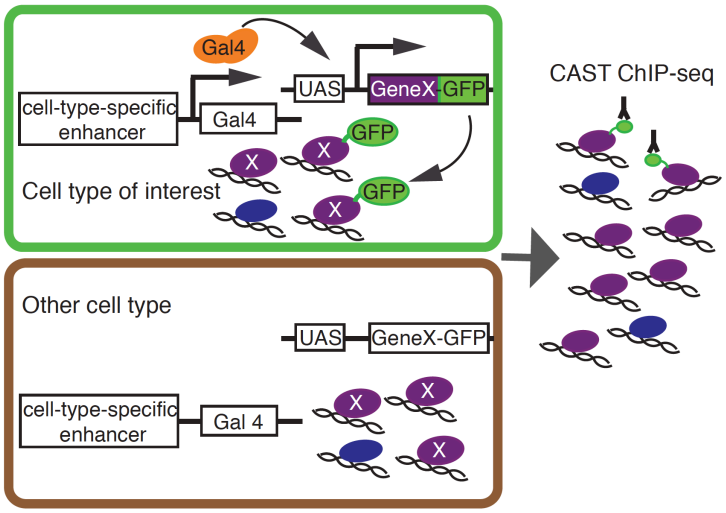


Figure 3.2 Principles of the CAST-ChIP method.

The GFP-tagged reporter (GeneX) is expressed in the cell type of interest and affinity purified using anti-GFP antibody. The purified DNA fragments are sequenced to obtain genome-wide binding sites.

3.4.1.1 Establishing CAST-ChIP

To identify gene activity variation between neurons and glia, I first profiled the genome-wide enrichment of RNA polymerase II. I used a transgenic, GFP tagged Pol II subunit, GFP-RPB3 that was described to recruit to target genes upon heat shock activation suggesting that the tagged protein is functional (Yao et al., 2006). In addition, GFP-RPB3 expressed in the salivary gland (c147-Gal4) overlaps with RPB1 on fixed polytene chromosomes (Yao et al., 2006 and reproduced on **Figure 3.3**), therefore mapping GFP-RPB3 most probably shows binding sites of the Pol II complex.

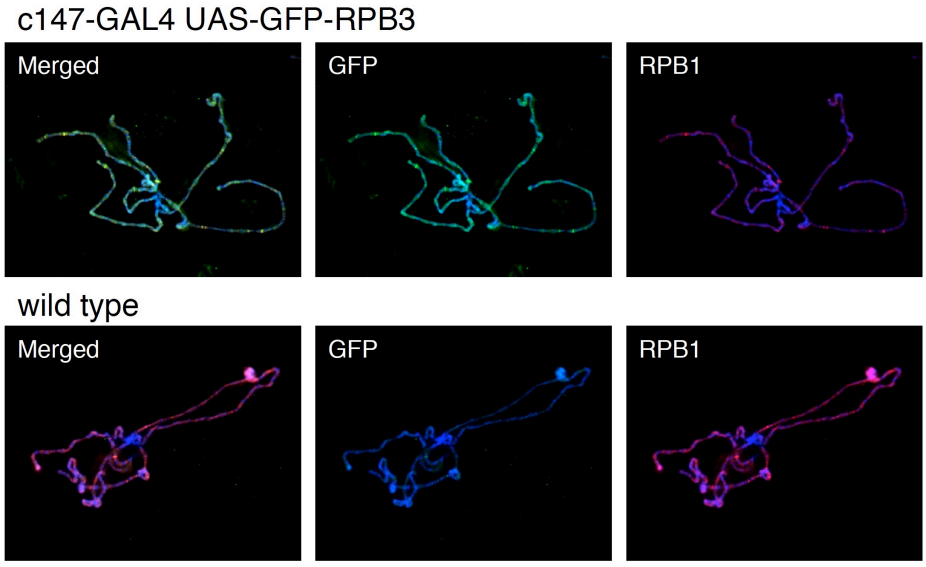
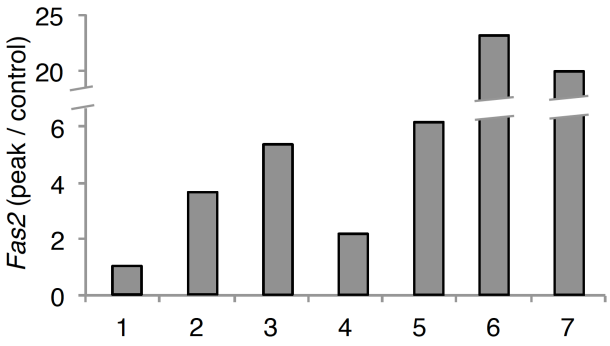


Figure 3.3 GFP-RPB3 co-localizes with RPB1 on salivary gland polytene chromosomes.

GFP-RPB3 (green) expressed by c147-Gal4 overlaps very well with the endogenous large Pol II subunit, RPB1 (red). The wild type control (bottom panel) does not show GFP staining. DNA is stained with DAPI (blue).

Next, I established and refined a crosslinking-based (X-ChIP) protocol suitable for CAST-ChIP. I used *elav*-Gal4 (neuronal) driven GFP-RPB3 in comparison to endogenous RPB3 on whole head (**Figure 3.4**). I tested in ChIP-qPCR two different anti-GFP antibodies (rabbit and goat) in combination with two bead-conjugates (protein A and G) and bead types (sepharose and magnetic). I also compared chromatin derived from animals carrying one or two copies of the *elav*-Gal4 transgenes. The best ChIP enrichment at the transcription start site (peak) of the *Fas2* gene over a neighboring region (control) is produced when using goat anti-GFP antibody, magnetic beads and pre-cleared chromatin from the double copy of *elav*-Gal4 (**Figure 3.4**). The best CAST-ChIP ratio was as high as using an anti-RPB3 antibody on extracts from wild type animals.



#	antibody	beads	preclearing
1	GFP(rabbit)	S protein A	yes
2	GFP(goat)	S protein A	yes
3	GFP(goat)	S protein G	yes
4	GFP(goat)	M protein G	no
5	GFP(goat)	M protein G	yes
6	GFP(goat)	M protein G	yes
7	RPB3 (rabbit)	M protein G	yes

1-5, *elav*-GAL4/+;UAS-GFP-*RPB3*/+
6, *elav*-GAL4/*elav*-GAL4;UAS-GFP-*RPB3*/UAS-GFP-*RPB3*
7, wild-type head RPB3

Figure 3.4 Optimization of CAST-ChIP.
ChIP enrichment at the *Fas2* gene over the control in qPCR analysis of CAST-ChIP using different antibodies (anti-GFP (rabbit and goat); anti-RPB3 (rabbit)), different bead composition [Sepharose (S) and magnetic beads (M)], and different bead conjugates (Protein A and Protein G).

To test why the ChIP is more efficient in the case of the double copy transgene compared to the single one, I performed western blot analysis on whole head extracts comparing the protein levels of GFP-RPB3. Although the transgene copy number is only doubled, GFP-RPB3 showed much higher expression from the double copy transgene (**Figure 3.5**). This is probably due to the multiplicator effects of Gal4. In comparison to the glia-specific driver (*repo*-Gal4), the double copy *elav*-Gal4 samples contained similar amounts of the tagged protein, despite the number of glia cells being much less than neurons (~5x) in the whole head. Thus, it seems that the amount of GFP-RPB3 transgene expressed in the head is crucial in order to provide optimal ChIP signals.

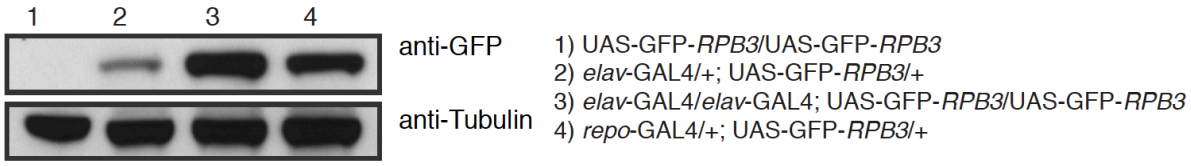


Figure 3.5 Expression of GFP-RPB3 in western blot analysis.

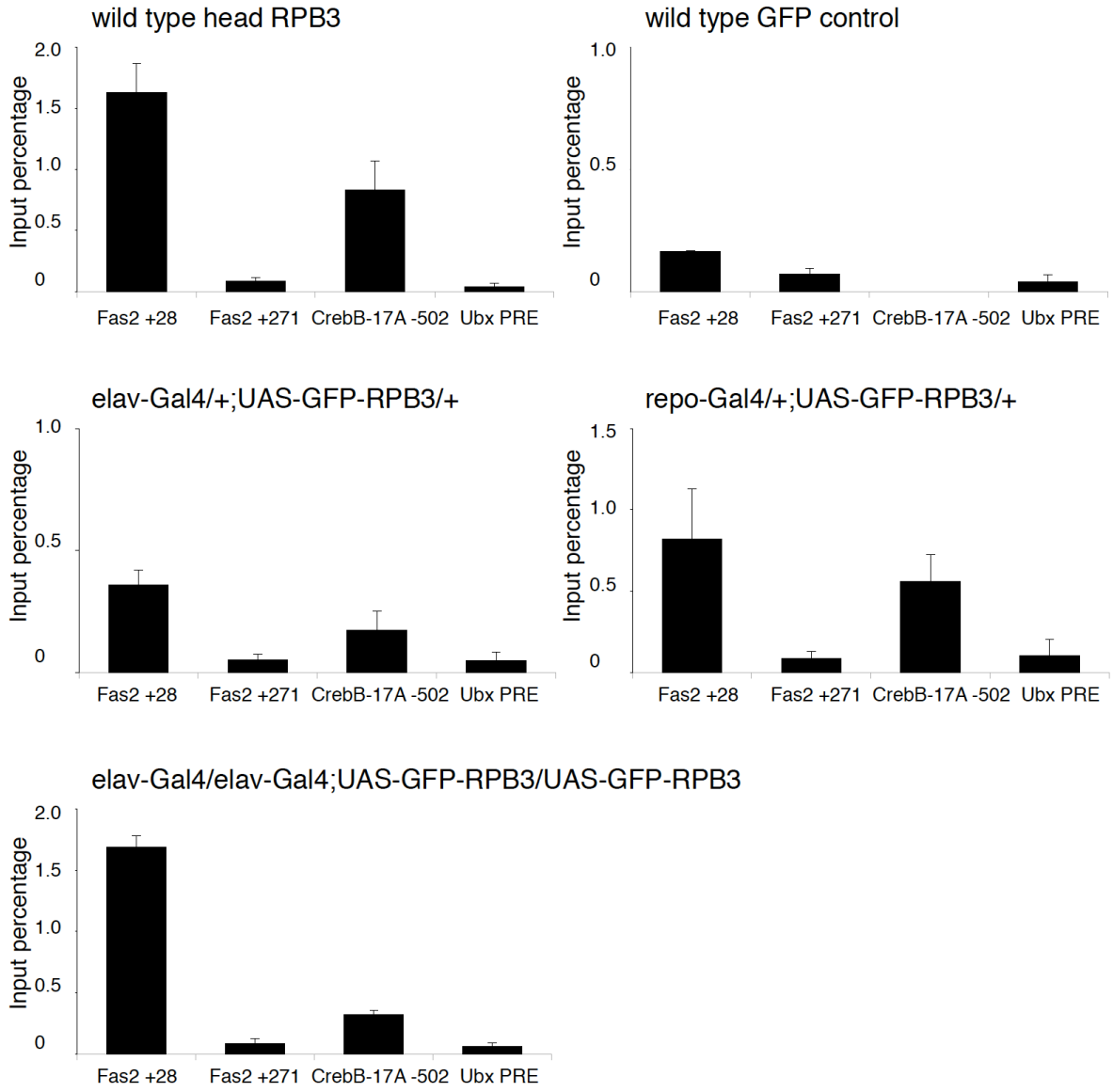
The GFP-tagged RNA polymerase II (Pol II) reporter (GFP-RPB3) 1) without the driver, 2) and 3) expressed in neurons (*elav*-GAL4; one and two copy transgene) and 4) in glia (*repo*-GAL4). An anti-Tubulin antibody served as loading control.

Once the CAST-ChIP protocol was optimized, I continued to create cell-type-specific ChIP profiles, which I tested in qPCR (**Figure 3.6**). As a control I used the same anti-GFP antibody on wild type animals lacking the GFP-tagged reporter. There was a small peak at the *Fas2* TSS in the GFP control, which can be explained by either the unspecific binding of the antibody or by the different shearing properties of the open TSS chromatin. However, the peak at the *Fas2* TSS (+28) compared to the gene body (+271) and the peak at the *CrebB* TSS (-502) are remarkably higher in the neuronal (*elav*, double copy) and glial (*repo*) CAST-ChIP. The region at the *Ubx* gene was chosen as control as a nucleosome-free region (Papp et al., 2006).

Figure 3.6 qPCR analysis of CAST-ChIP.

Regular ChIP was performed on wild type head using anti-RPB3 and anti-GFP antibody (top panels). CAST-ChIP on neuron- (single copy: elav-Gal4/+;UAS-GFP-RPB3/+; double copy: elav-Gal4/elav-Gal4;UAS-GFP-RPB3/UAS-GFP-RPB3) and glia-specific (repo-Gal4/+;UAS-GFP-RPB3/+) samples (middle and bottom panels). qPCR primers were designed for the Fas2 TSS (+28) and gene body (+271), CrebB-17A TSS (-502)(see Methods) as well as the PRE (polycomb response element) of the Ubx gene (F4 primer from Papp et al., 2006).

Figure 3.6



3.4.1.2 Cell-type-specific RNA polymerase II profiles

I performed Chromatin ImmunoPrecipitation combined with Illumina sequencing to map genome-wide Pol II binding in heads (ChIP RPB3), neurons and glia (CAST-ChIP). Wild type head anti-GFP ChIP and Input chromatin were also sequenced as controls (**Figure 3.7**). All sequencing runs were duplicated as biological replicates. To find Pol II enriched regions, we called the peaks using CCAT on each replicate and the anti-GFP ChIP as background (*data analysis in collaboration with Petra Schwalie*). To identify all the regions obtained in the head and the two cell types, we defined the region of interest (ROI) by merging the datasets (**Figure 3.7** and see Methods). To call differences in the

cell types, we performed differential expression analysis (DESeq; Anders et al., 2010) and determined neuronal, glial and invariant subsets (**Figure 3.7**).

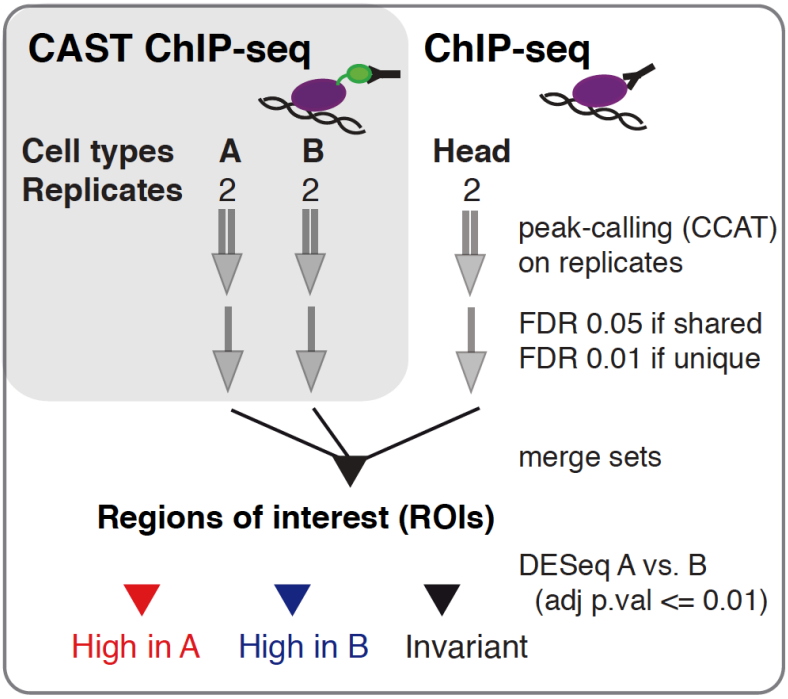


Figure 3.7 CAST-ChIP data analysis workflow.

The regions of interest (ROIs) were defined by merging the peak calls from the cell types and head. Cell-type-specific peaks were determined using DESeq (Anders et al., 2010).

There were in total about 7500 regions with significant Pol II enrichment. As shown in previous studies (see **section 2.2.1**), a high proportion (75%) of the peaks is located near transcription start sites (TSS; **Figure 3.8**). The rest of the peaks were found in protein coding regions indicating multiple start sites, or at a 5 kb distance from genes, suggesting un-annotated upstream start sites.

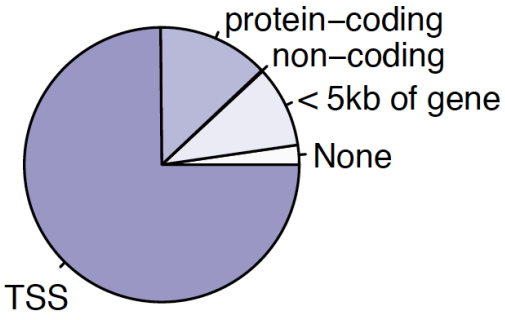


Figure 3.8 Pol II ROIs mainly localize to TSSs.

Proportion of Pol II peaks located at TSSs, protein coding, non-coding and within 5kb of genic regions.

Next, we compared the regions of interest (ROIs) in the head RPB3 dataset and in the cell-type-specific ones. We identified peaks found only in the neuronal or only in the glial datasets, as well as regions present in all sets (invariant; **Figure 3.9**). Quantifying the differential peaks, we found 986 neuronal and 937 glial peaks (see Methods). In

addition, there were peaks absent from these two cell types and only present in the whole head (*data not shown*). The invariant regions on the Venn diagram (**Figure 3.9**) contain the ROIs where the peaks are either indifferent between neurons and glia or those that have no peaks in this cell type but in the head dataset.

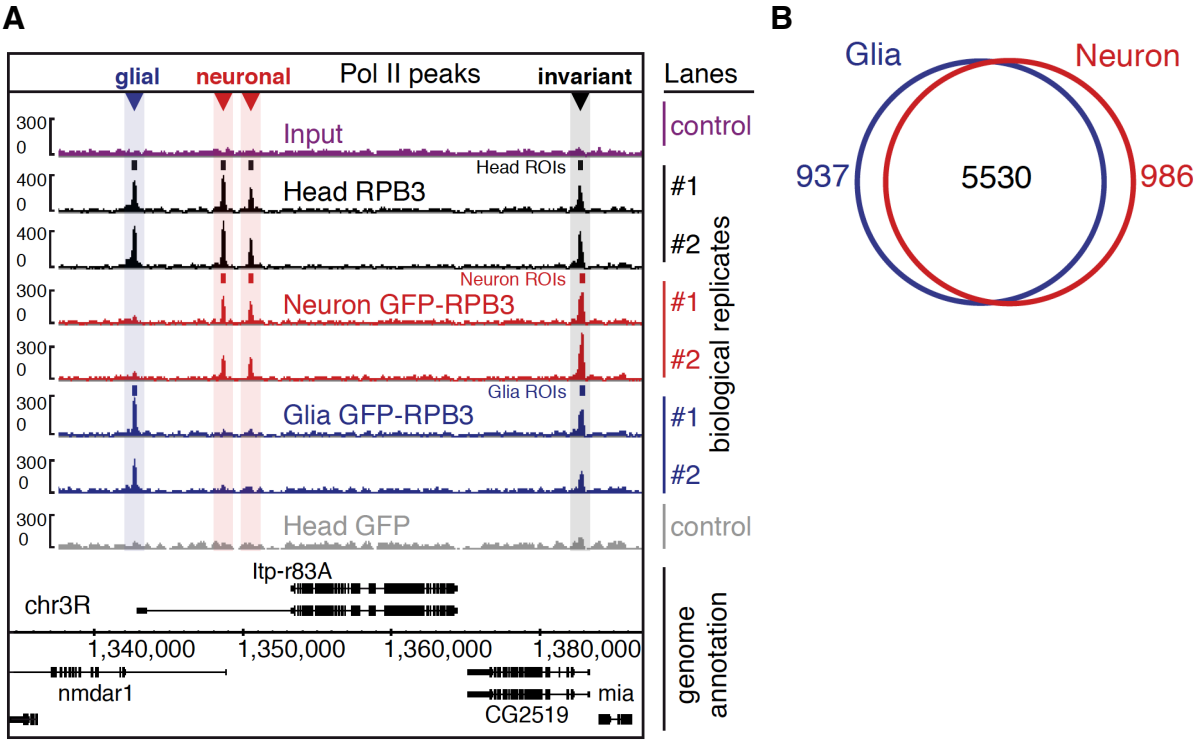


Figure 3.9 CAST-ChIP reveals cell-type-specific Pol II regions.

(**A)** Genome browser snapshot at the Nmdar1 gene showing Input (magenta); head RPB3 ChIP (black); neuronal (red), glial (blue) GFP-RPB3 CAST-ChIP and head GFP ChIP (grey) in biological replicates. Identified regions (ROIs) are indicated as rectangles above the tracks, differential regions as color-coded triangles (top track). **(B)** Venn diagram showing the quantification of neuron- and glia-specific Pol II ROIs.

Biological replicates correlated very well (R=0.96) as shown by statistical analysis at the ROI (**Figure 3.10**). In contrary, Spearman correlation of neuronal vs. glial CAST-ChIP peaks was relatively low (R=0.65), indicating significant differences between distinct cell types. Thus, I was confident that our rapid and efficient CAST-ChIP method generates reproducible results by profiling gene activity in neurons and glia.

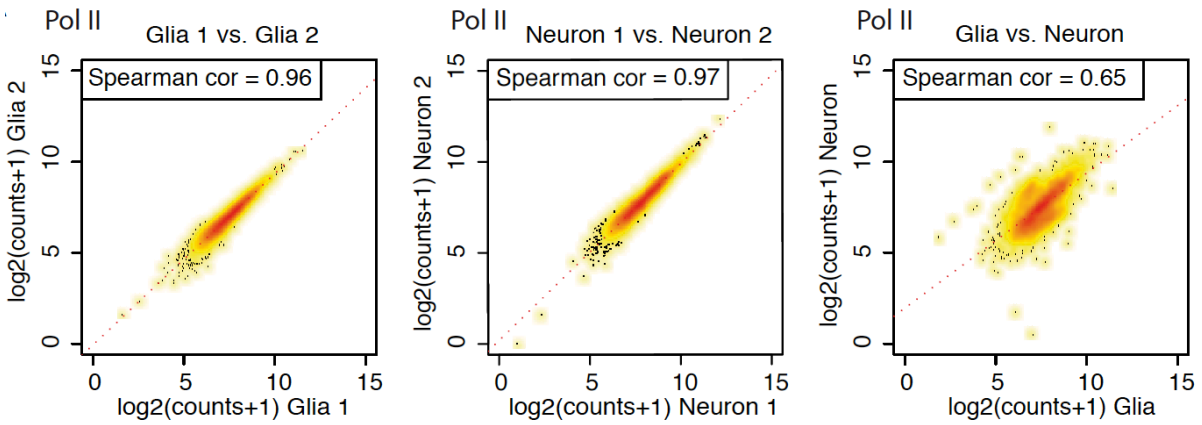


Figure 3.10 Pol II CAST-ChIP shows high correlation between replicates and low correlation between cell types.

Spearman correlation of glial and neuronal Pol II CAST-ChIP-seq profiles between biological replicates (R=0.96 and R=0.97) as well as between glia and neurons (R=0.65).

3.4.2 Validation of CAST-ChIP

CAST-ChIP identified about a thousand of neuronal and glial Pol II sites, respectively. In order to evaluate whether these Pol II regions are functional in the particular cell type, we performed 1) computational and 2) experimental validation. We compared my data to already existing datasets such as curated lists of central nervous system related genes (Pfeiffer et al., 2008) or using microarray data from dissected tissues (FlyAtlas; Chintapalli et al., 2007). In addition, we tested enrichment of gene ontology (GO) terms in our cell-type-specific datasets. Experimentally, I used the advantage of the available *Drosophila* enhancer trap lines, which carry an insertion of Gal4 transgenes showing locus specific expression pattern. Comparing spatial expression patterns of cell-type-specific loci identified by CAST-ChIP with cell type markers, we can test whether a cell-type-specific Pol II peak is characteristic of the cell type.

3.4.2.1 Computational validation

To compare the findings to published data, we first used a list of 925 selected genes, which showed expression (or at least were predicted to be expressed) in the *Drosophila* brain (Pfeiffer et al., 2008). The authors described these potentially functional genes as transcription factors, neuropeptides, ion channels, transporters, and receptors. We split genes according to CAST-ChIP results as neuron- or glia-specific, invariant or no Pol II binding categories and tested the representation in the Pfeiffer *et al.* dataset. All three head datasets having Pol II were significantly over-represented in the "Pfeiffer list", whereas genes without Pol II were not (**Figure 3.11**). However, the far most significant

class was the neuronal subset, suggesting that the Pol II bound genes specific for neurons have a neuronal function by matching the "brain gene-set" from Pfeiffer *et al*.

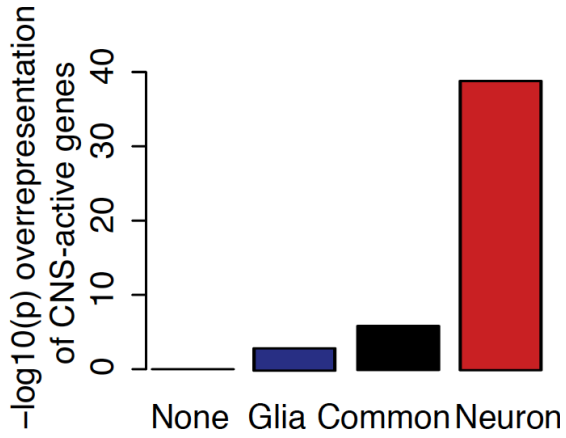


Figure 3.11 Representation of cell-type-specific Pol II-bound genes in the *Drosophila* brain-specific geneset.

Genes without Pol II, with glia, common and neuron-specific Pol II peaks are shown in comparison to brain-specific gene list (p-value is shown; Pfeiffer et al., 2008).

As a second test, we used the advantage of the published microarray data collection from the FlyAtlas (Chintapalli et al., 2007). FlyAtlas is a compendium of gene expression profiling from several dissected tissues and organs from different developmental stages. We tested the variation of gene expression values among FlyAtlas tissues in our neuronal, glial and common gene-sets. As expected, genes with cell-type-specific Pol II peaks, deviate significantly more compared to those with invariant Pol II ROIs (common) ($p < 10^{-16}$ using Wilcoxon rank sum test) (**Figure 3.12**). This confirms that we could distinguish between cell-type-specific and common (invariant) genes using CAST-ChIP on two cell types.

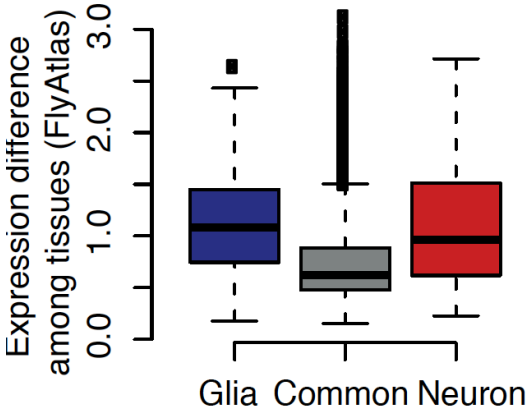


Figure 3.12 Cell-type-specific Pol II-associated genes show diverse RNA expression across tissues in the FlyAtlas.

Standard deviation of FlyAtlas probe-level values for genes carrying glia-specific, common or neuron-specific Pol II obtained by CAST-ChIP.
(FlyAtlas: Chintapalli et al., 2007)

Visualizing FlyAtlas expression values as heatmaps in the three subclasses (common, glial, neuronal), we observed that common genes are indifferent, whereas specific genes deviate among tissues (**Figure 3.13**). Note that neuron-specific genes show higher

expression in FlyAtlas tissues such as larval central nervous system (CNS), thoracicoabdominal (TA) ganglion and adult brain (**Figure 3.13, black rectangle**).

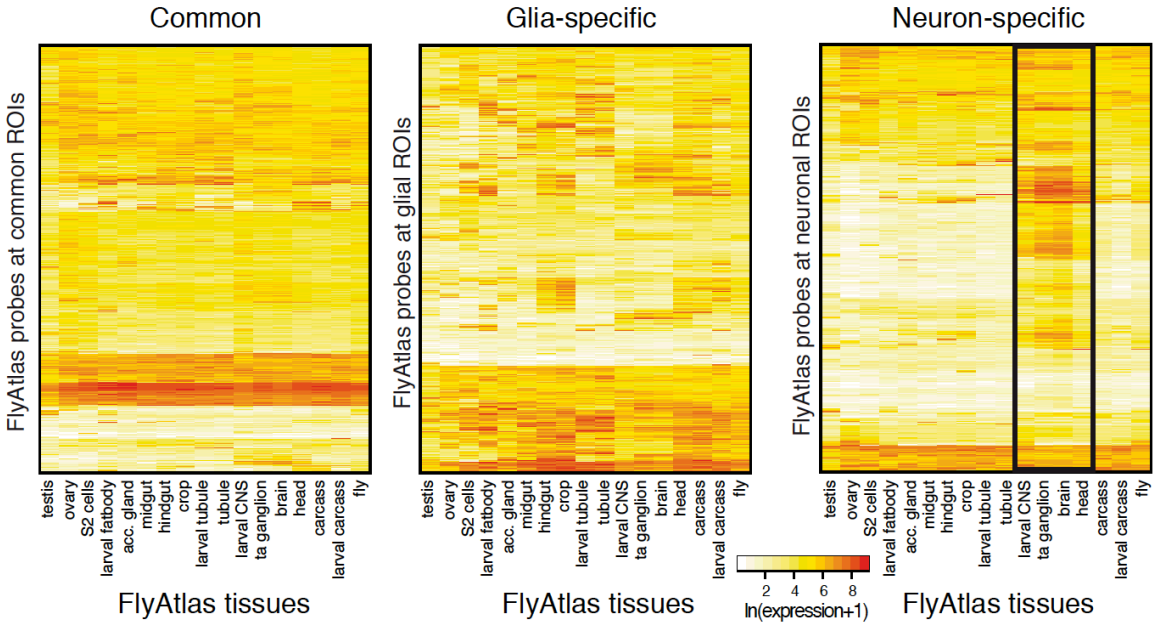


Figure 3.13 Comparison of FlyAtlas expression in common and cell-type-specific gene classes.

Heatmaps showing FlyAtlas expression values (color range) from dissected tissues (x- axis) for genes overlapping common, glial and neuronal Pol II ROIs (3 panels; y-axis: probes). Common genes do not differ extensively between different tissues (left panel), whereas genes marked by neuronal Pol II binding are highly expressed in tissues such as larval CNS, thoracicoabdominal ganglion (ta ganglion), brain and head (right panel; black rectangle).

The third computational test was to find gene ontology (GO) terms enriched in the cell-type-specific gene-sets. In the neuronal dataset, GO terms related to neuronal function (axonogenesis, nervous system development etc.) showed significant enrichment (**Table 3.1**). On the other hand, we did not find relevant terms enriched in the glial dataset. This might be due to the diverse function of glia cells (see chapter 2.1.2). However, the neuron-specific terms also confirmed that our specific Pol II peaks mark genes involved in specific cellular functions.

GO ID	Term	P-value
GO:0007409	**axonogenesis**	0.00153
GO:0007411	**axon guidance**	0.00352
GO:0040008	**regulation of growth**	0.00377
GO:0044267	**cellular protein metabolic process**	0.00606
GO:0007417	**central nervous system development**	0.00680

Table 3.1 Neuronal function-related GO terms are enriched in the neuronal dataset.
Terms related to neuronal function are enriched in the geneset overlapping neuronal Pol II.

Gene ontology analysis showed enrichment of neuronal terms in the neuronal dataset, however there are several genes with other special neuronal function that are absent from GO. Therefore, we took the top 30 significantly enriched genes that have a known neuron-related function based on FlyBase (**Table 3.2**). Beside terms what we already identified by GO analysis (e.g. Sdc; gogo etc. [axon guidance]), we found several genes with neurotransmitter receptor activity (e.g. nAcRbeta-96A; 5-HT1A Oamb; Glu-RI; Octbet3R etc.), as well as genes with cell adhesion function (e.g. klg [learning and memory]; see **section 2.1.2.1**).

FlyBase ID	Gene Name	Fold Change	adj. p-value	FlyBase terms
FBgn0053202	dpr11	**4.1**	3.54E-45	sensory perception of chemical stimulus
FBgn0259246	brp	**4.1**	1.69E-38	calcium channel activity
FBgn0004118	nAcRbeta-96A	**4.0**	7.41E-38	acetylcholine-activated cation-selective channel activity
FBgn0050361	mtt	**4.8**	2.23E-36	G-protein coupled receptor activity
FBgn0052183	Ccn	**4.0**	9.37E-36	neurogenesis
FBgn0004168	5-HT1A	**3.9**	7.09E-35	serotonin receptor activity
FBgn0053960	Sema-2b	**3.8**	1.15E-34	synaptic target attraction; dendrite guidance
FBgn0033058	CCHa2r	**3.6**	7.66E-34	neuropeptide receptor activity
FBgn0005775	Con	**3.7**	3.91E-33	homophilic cell adhesion; axonal fasciculation
FBgn0028433	Ggamma30A	**3.6**	6.23E-33	GTPase activity; phototransduction
FBgn0259225	Pde1c	**3.7**	1.53E-30	cAMP phosphodiesterase activity; male mating behavior
FBgn0086778	gfA	**3.6**	2.09E-30	acetylcholine-activated cation-selective channel activity
FBgn0053171	mp	**3.3**	6.13E-30	carbohydrate binding; motor axon guidance
FBgn0051190	Dscam3	**3.4**	7.83E-30	homophilic cell adhesion
FBgn0015721	king-tubby	**3.6**	9.47E-30	sensory perception of smell
FBgn0024944	Oamb	**3.3**	2.39E-29	octopamine receptor activity
FBgn0010415	Sdc	**3.6**	2.45E-29	axon guidance
FBgn0017590	klg	**4.0**	1.66E-28	learning or memory; homophilic cell adhesion
FBgn0052635	Neto	**3.7**	4.82E-27	neuromuscular synaptic transmission; locomotion
FBgn0013759	CASK	**3.0**	6.17E-27	regulation of neurotransmitter secretion; adult locomotory behavior
FBgn0025593	Glut1	**3.5**	6.22E-27	glucose transmembrane transporter activity
FBgn0032151	nAcRalpha-30D	**3.5**	7.16E-27	acetylcholine-activated cation-selective channel activity
FBgn0004619	Glu-RI	**4.0**	1.05E-26	extracellular ligand-gated ion channel activity, AMPA receptor
FBgn0085447	sif	**3.1**	1.36E-25	regulation of axonogenesis.
FBgn0052227	gogo	**3.5**	2.69E-25	axon guidance
FBgn0024963	GluClalpha	**3.0**	6.87E-25	extracellular-glutamate-gated chloride channel activity
FBgn0035092	Nplp1	**2.9**	1.65E-23	neuropeptide hormone activity
FBgn0250910	Octbeta3R	**3.3**	3.03E-23	octopamine receptor activity
FBgn0015774	NetB	**3.4**	3.06E-23	axon guidance
FBgn0013467	igl	**3.2**	6.44E-23	calmodulin binding

Table 3.2 Top 30 neuronal Pol II-bound genes with known neuronal function.
The table shows the FlyBase ID, the Gene Name, the Fold Change of neuron vs. glia comparison, the corresponding adjusted p-value and the functional term from FlyBase.

In summary, the computational validation confirmed that cell-type-specific genes obtained by Pol II CAST-ChIP:

1) overlap with a curated *Drosophila* brain-related gene list (Pfeiffer et al., 2008),

2) show higher variation in gene expression among dissected tissues (FlyAtlas) compared to common genes (Chintapalli et al., 2007),

3) show enrichment of GO terms relevant for the cellular function.

3.4.2.2 Experimental validation

Beside computational validation, I wanted to evaluate the data experimentally, by testing whether a gene identified by cell-type-specific Pol II shows a spatial expression pattern in the marked cell population. To do this, I performed co-staining of a nuclear localized GFP reporter (histone-GFP) and the nuclear cell markers ELAV (pan-neuronal) and REPO (pan-glial; see **section 2.1.2**). I chose the nuclear GFP reporter for several reasons: 1) to evaluate overlap with the nuclear cell-specific markers (ELAV and REPO), it is easier to use nuclear localized reporters compared to diffuse localization of RNA; 2) *in situ* hybridizations do not work as accurately as regular immuno-staining on such fragile tissues as the fly brain; and 3) the lack of antibodies against chosen cell-type-specific gene products argues for the GFP reporter using an anti-GFP antibody.

I expressed the nuclear GFP reporter using Gal4 enhancer trap lines that carry an insertion usually close to the TSS, ensuring transgene expression according to the enhancer activity of the locus. I chose regions with cell-type-specific Pol II peaks and checked for available Gal4 insertion lines specific for the locus. Out of 15 lines selected, 12 expressed the GFP transgene in viable and healthy offspring. Genes marked by Pol II only in neurons and without any Pol II enrichment in glia (*king-tubby, igloo, Oaz, CG6044* and *klingon*) showed clearly neuronal expression (**Figure 3.14 and 3.15**). The GFP reporter clearly overlapped with the neuronal ELAV but was excluded from the glial REPO staining. Glia-specific ROIs such as at *Mocs1, CG4666* and *tramtrack* genes showed the opposite pattern overlapping only with REPO marker. There were regions with significant enrichment in one of the cell types but carrying a smaller peak in the other (*bitesize, CG5835* and *Eaat1*, for example) expressing GFP in the cell type with the higher Pol II levels. *aPKC* has multiple Pol II peaks in both cell types and shows expression in both glia and neurons (**Figure 3.15**).

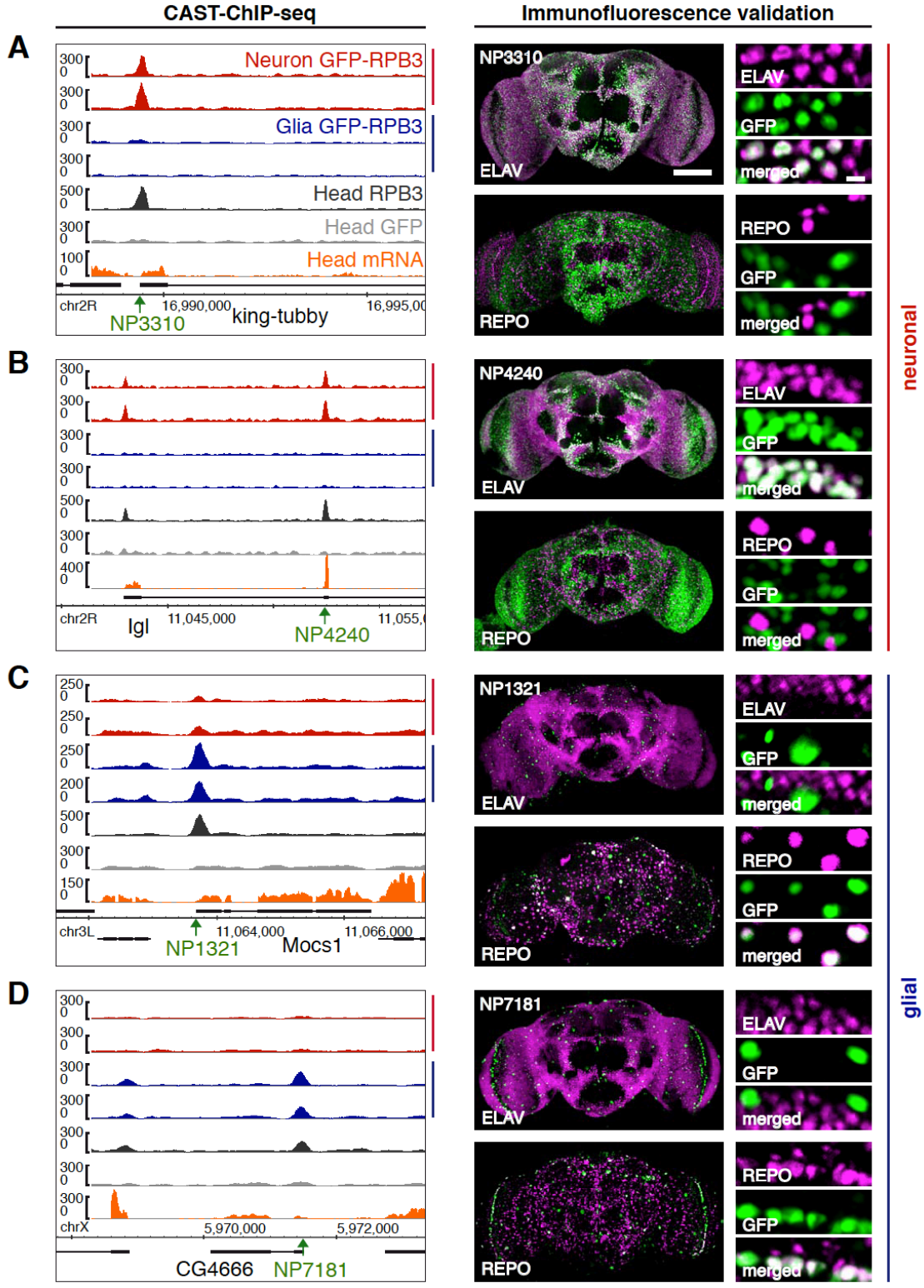


Figure 3.14 Experimental Validation of Neuronal and Glial Pol II Enrichment.

To validate neuron- and glia-specific Pol II peaks, we monitored the neuronal and glial specificity of enhancer-trap insertion lines driving nuclear GFP at (A) king-tubby, (B) Igl, (C) Mocs1 and (D) CG4666. Left panels: Pol II genomic profiles for neurons and glia, whole-head Pol II ChIP, GFP control and whole-head mRNA tracks, including gene annotation and location of the Gal4 driver (green arrow, insertion point). Right panels: Co-staining of nuclear GFP (green) expressed using Gal4 drivers located in proximity of the cell-type-specific Pol II peaks, with REPO or ELAV markers (magenta) to overlap expression patterns (white).

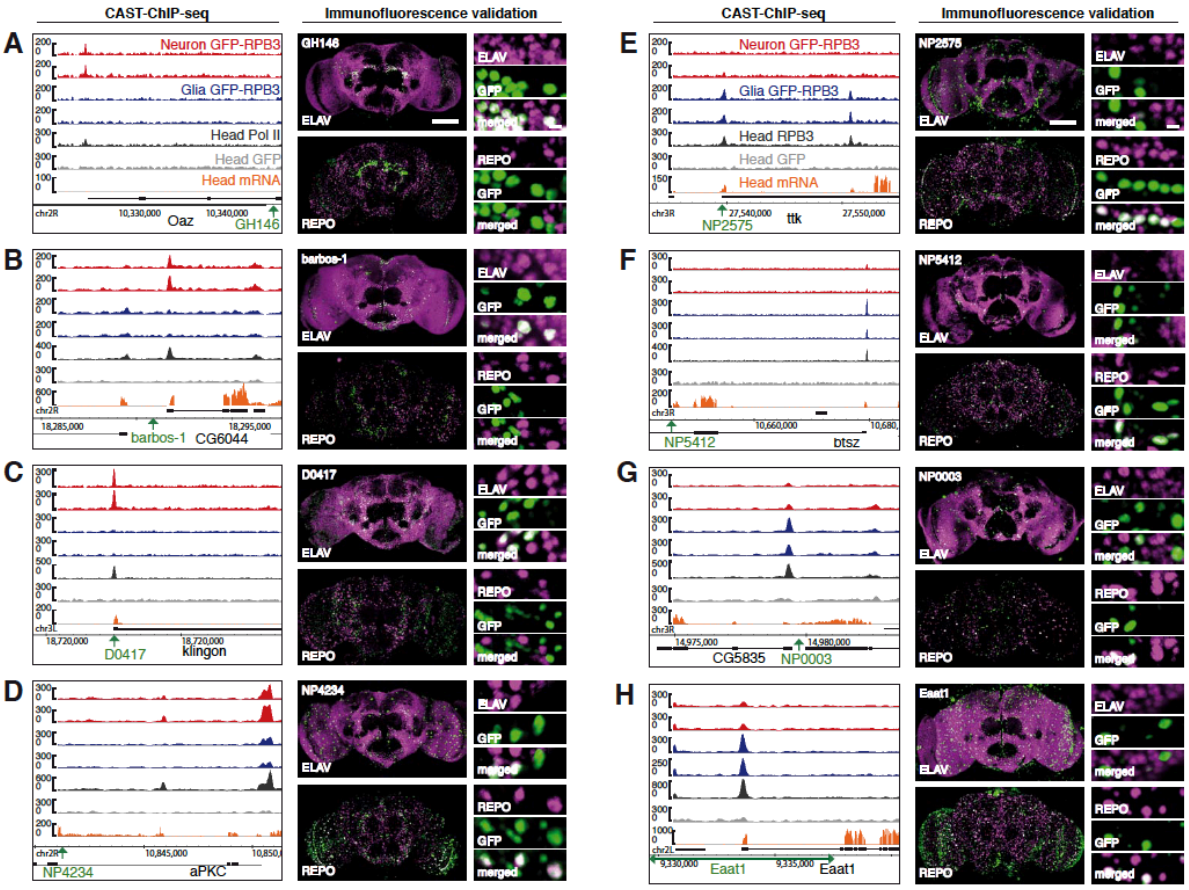


Figure 3.15 Experimental Validation of Neuronal and Glial Pol II Enrichment II.
Further validation at the genes at genes such as (A) *Oaz*, (B) *CG6044*, (C) *klingon*, (D) *aPKC*, (E) *ttk,* (F) *bitesize* (*btsz*), (G) *CG5835* and (H) *Eaat1*. Labels are the same as in **Figure 3.16**.

In summary, analysis of enhancer trap driven nuclear GFP verifies the CAST-ChIP data. Pol II binding qualitatively reflects the spatial expression pattern of the gene it localizes to. Thus, the method CAST-ChIP is able to biochemically enrich the chromatin-associated GFP-tagged Pol II expressed in a subset of specialized adult cell types within the *Drosophila* head.

3.5 Discussion

In this chapter, I introduced a novel method, CAST-ChIP. CAST-ChIP combines *Drosophila* genetics with biochemical enrichment of chromatin-associated protein from specific cell populations. CAST-ChIP is an optimized ChIP protocol and a rapid and sensitive method to obtain cell-type-specific RNA polymerase II-associated regions. Since CAST-ChIP works on formaldehyde cross-linked samples, it minimizes any transcriptional or stress disturbance during the experimental procedure. The experimental time is comparable to regular ChIP protocols without any additional steps (staining or FACS sorting). The amount of ChIPed DNA is enough for regular high throughput sequencing runs without amplification (although sometimes pooling of technical replicates is required). CAST-ChIP does not require any special equipment, such as FACS sorters, and can be carried out on a cooled bench environment, therefore being a simple, efficient approach.

Using CAST-ChIP, we obtained about 1500 neuron- and glia-specific RNA polymerase II enriched regions. Genes carrying cell-type-specific Pol II peaks have a function in the particular cell populations by comparing our data to curated neuronal gene-sets and to GO terms (**Figure 3.11 and Table 3.1**). Several genes have neuronal function among those that carry a neuronal peak with the highest enrichment (**Table 3.2**). There is evidence (based on Flybase) that these genes have real molecular function in neurons (see **section 2.1.2.1**), such as well-known receptors activated by neurotransmitter including acetylcholine, glutamate, serotonin and octopamine (**Table 3.2**). There are neuropeptides in the neuronal list involved in hormonal regulation (*Nplp1*). *Bruchpilot* (*brp*), a ubiquitous presynaptic active zone protein required for efficient vesicle release at synapses in general (Wagh et al., 2006) and for anesthesia-resistant memory in the mushroom body (Knapek et al., 2011). The cell adhesion molecule *klingon* (*klg*), required for long-term memory formation, carries also a highly significant neuronal Pol II peak (Matsuno et al., 2009). Thus, I am confident that the top neuronal hits are functional in neurons.

To validate CAST-ChIP, I chose several cell-type-specific Pol II regions and tested whether expression from its locus can drive a GFP reporter in the given cell population. All of the tested lines indicate that genes marked by cell-type-specific Pol II follow the expected expression pattern. Genes such as *igloo* (*Igl*) encodes a protein homologue of a

calmodulin binding protein, GAP43 important in developing and regenerating neurons (Neel et al., 1994). The gene *Eaat1* carries a significantly higher glia-specific Pol II peak and its reporter expression is also mainly overlapping the glial marker REPO. *Eaat1* encodes a glial excitatory amino-acid transporters that re-uptakes glutamate at synapses (see **section 2.1.2**; Rival et al., 2004). Interestingly, some of the insertions express GFP in a broad range of neuronal cells (e.g. *king-tubby*, *Igl*), some mark in a localized subset of cells (in the olfactory lobe: *Oaz*) or a small but more diverse neuronal populations (*klg*). Further characterization of these subsets by CAST-ChIP or other cell-type-specific methods might reveal what is common in these sub-anatomical cell populations, which genes are shared or unique to these cells.

CAST-ChIP is a suitable tool to identify genes required for cell-type-specific function. I discovered several neuronal and glial Pol II-associated genes including those that are expressed only at low mRNA levels (**Figure 3.9, 3.14** and *data not shown*). Any type of cell-type-specific approach based on RNA profiling would have failed to call these differences. Most probably these genes carry stalled Pol II at their promoter and there is very slow transcription elongation ongoing at these genes. I investigated naive flies that are not exposed to extreme environmental stimuli; therefore some of the neuro-receptors are not in actively transcribing status. Future experiments might reveal that actually these genes get activated upon neuronal induction and the stalled RNA polymerase II may turn to an elongating phase. Here, I reported the basal status of RNA polymerase marking the transcription start site of neuron-specific genes.

I found that cell-type-specific genes obtained by CAST-ChIP vary the most among dissected tissues of the FlyAtlas (**Figure 3.12 and 3.13**). In contrary, genes with shared Pol II ROIs in glia and neurons show ubiquitous expression. In addition, this group contains genes encoding ribosomal proteins (*data not shown*). That suggests that there are two classes of genes: specific ones, which are dynamically regulated during development of distinct cells; and ubiquitous genes, which are expressed independently of developmental stage and tissue. Some of these can be defined as housekeeping genes.

Pol II CAST-ChIP is a sensitive method to distinguish between specific and ubiquitous genes making possible to study how these genes are regulated. However, it remains unclear whether other chromatin-associated proteins, such as histone variants, could be used to either confirm these results or highlight other features of cell-type-specific gene regulators.

4 H2A.Z – an epigenetic mark for ubiquitous gene activity

This project was performed in experimental collaboration with Carla Margulies (EMBL, Heidelberg, Germany) and bioinformatics collaboration with Petra Schwalie (EBI, Hinxton, UK).

4.1 Summary

To monitor gene activity in distinct cell types, I sought to profile the incorporation of the histone variant H2A.Z using the established CAST-ChIP method (see **chapter 3**.). H2A.Z is known to mark active genes carrying Pol II (see **section 2.2.2**); therefore, its CAST-ChIP profiles may reveal insights into the interaction between Pol II and H2A.Z in distinct cell types.

First, I confirmed previous reports that H2A.Z is mainly enriched close to the transcription start site (TSS) of expressed genes and absent from inactive genes (see **section 2.2.2**). H2A.Z overlaps with a high proportion of RNA Polymerase II bound genes, however there are thousands of regions marked by H2A.Z only or only by Pol II. This suggests a role for H2A.Z that is at least partly unrelated to that of Pol II and delineates a class of Pol II associated genes that are marked by the absence of H2A.Z. Surprisingly, H2A.Z profiles obtained in neurons and glia are remarkably similar to each other, with only about hundred cell-type-specific H2A.Z regions. In contrast, H2A.Z is absent from thousands of genes that are marked by cell-type-specific Pol II; in contrary, H2A.Z is enriched at common, cell-type-invariant genes. Furthermore, we compared H2A.Z incorporation in the embryo and the adult head using regular ChIP methods and find that it is maintained across developmental stages.

By comparing our cell-type-specific H2A.Z maps with recently published data (see **section 4.4.1**), I found that H2A.Z associates with constitutively active chromatin, suggesting its role in maintaining the expression of ubiquitous genes. H2A.Z is present at genes with broad promoters and overlaps with housekeeping gene clusters. Further, H2A.Z is significantly enriched at the boundaries of chromatin domains and associates

with insulator binding proteins, suggesting a role of H2A.Z in marking the borders of ubiquitous domains.

In summary, CAST-ChIP reveals a novel, likely chromatin organizer function of H2A.Z in marking ubiquitous gene transcription.

4.2 Introduction

Nucleosome positioning helps to organize the genome, ensuring the correct regulation of gene expression. On the other hand, changes in gene expression influence the position of nucleosomes by reorganizing chromatin at the level of histone modifications and chromatin remodeling. In addition, histone variants including H2A.Z are incorporated in well-positioned nucleosomes flanking TSSs (Barski et al., 2007; Mavrich et al., 2008; Schones et al., 2008; Jiang et al., 2009; Bargaje et al., 2012). H2A.Z is enriched at the -1 and +1 nucleosome adjacent to the nucleosome-free region of promoters in yeast (Raisner et al., 2005) and in humans (Barski et al., 2007). However, *Drosophila* (and *Arabidopsis*) promoters generally lack H2A.Z nucleosomes at the -1 position, while it is present in nucleosomes downstream of the TSS (Mavrich et al., 2008; Zilberman et al., 2008), predominantly at the +1 position.

Homotypic H2A.Z-containing nucleosomes are enriched at active genes, with moderate to high expression and are depleted from inactive genes in *Drosophila* (Weber et al., 2010). A combination of histone H3.3-H2A.Z-containing, double variant nucleosomes mark active promoters, enhancers and insulator regions in human cells (Jin et al., 2009). The proximity of H2A.Z nucleosome to the TSS positively influences RNA polymerase II recruitment and gene expression levels, suggesting the importance of the well-defined position (Bargaje et al., 2012). In contrast, H2A.Z incorporation in gene bodies is associated with lower expression levels, but a higher responsiveness to environmental stimuli (ColemanDerr et al., 2012). In yeast, H2A.Z is necessary for the reactivation of repressed genes (e.g. INO1 and GAL1), suggesting a role in transcriptional memory (Brickner et al., 2007 and Brickner et al., 2009).

H2A.Z associates with active histone post-translational modifications, including H3K4me3, and also with bivalent domains that carry both active and repressive marks at the same site (i.e. H3K4me3 and H3K27me3) in embryonic stem cells (ESCs; (Ku et al., 2012 and Pandey et al., 2013). H2A.Z was found to be ubiquitylated by the human

PRC1 component Ring1B at bivalent domains (Sarcinella et al., 2007; Ku et al., 2012). However, the role of H2A.Z at these sites is not completely clear; H2A.Z may mark genes that are poised for activation. H2A.Z can also be acetylated at its lysines K3, K8, K10 and K14 by the yeast Gcn5 and Esa1 histone acetyl-transferases at active genes (Millar et al., 2006). In cancer cells, the increase of acetylated H2A.Z at the TSS co-occurs with a decrease of total H2A.Z upon oncogene activation (ValdesMora et al., 2012). In *Drosophila*, acetylation of H2A.Z increases at the *Hsp70* gene upon heat shock activation (Tanabe et al., 2008; Kotova et al., 2011).

In summary, H2A.Z is generally incorporated into the promoter regions of active genes. This enrichment is evolutionarily conserved. Monitoring H2A.Z occupancy in a genome-wide manner uncovers transcriptionally active regions. Whether regions marked by H2A.Z differ among cell types and whether H2A.Z is coupled to cell-type-specific Pol II binding is still unclear. Therefore, I decided to compare H2A.Z and Pol II enrichments in specific cell types in order to reveal whether cell-type-specific gene regulation may differ from the transcriptional regulation of ubiquitous genes.

4.3 Methods

4.3.1 Experimental procedures

Experiments in this chapter were performed by Tamás Schauer and Carla Margulies

4.3.1.1 Fly Stocks

Fly stocks were kept and used as described under **3.3.1.1**. Additionally, genomic-H2A.Z-GFP was used to compare untagged and GFP-tagged H2A.Z (gH2A.Z-GFP from Robert Saint, [Clarkson et al., 1999]). The UAS-H2A.Z-GFP strain was generated by Carla Margulies. Briefly, the H2A.Z-GFP ORF was PCR amplified from the gH2A.Z-GFP construct and inserted into pUAST vector to create transgenic lines.

4.3.1.2 Chromatin Immunoprecipitation and Sequencing

ChIP and CAST-ChIP experiments on adult, head-derived chromatin were carried out as described in **section 3.3.1.3**. ChIP and CAST-ChIP on H2A.Z were performed in

collaboration with Carla Margulies. In this chapter, the following antibodies were used for ChIP: anti-GFP (goat, Ladurner lab stock), anti-H2A.Z and anti-H2A (from Robert Glaser; Leach et al., 2000), anti-H3 (AbCam 1791).

The ChIP-qPCR analysis at *Hsp70* was performed as described (Adelman et al., 2006). In all qPCR experiments the Fast SYBR Green mix was used on an Applied Biosystems 7500 Fast Real-Time PCR System. Primer sequences were obtained from Karen Adelman.

ChIP on embryos was carried out as described previously (Mavrich et al., 2008; Whittle et al., 2008). Briefly, 0-6 hour embryos were collected and washed with PBST 0.1% Triton-X, dechorionated with 3% sodium hypochlorite, fixed with 2 % formaldehyde and heptane. After quenching the fixing with glycine (125 mM), the embryos were washed, frozen in liquid nitrogen and stored at -80°C. Chromatin preparation from 1 gram fixed, frozen embryos was continued with homogenization as described above.

4.3.1.3 RNA Isolation and Sequencing

25-50 fly heads (2-3 day old) were homogenized in Tri-Reagent (Sigma) and isolated according the manufacturer´s recommendations. 5-10 µg total RNA was obtained and sent for deep sequencing in 2 biological replicates. Sequencing library was prepared using Illumina polyA-mRNA library preparation methods with paired-end option. 72bp reads were obtained from the sequencer.

4.3.2 Data analysis

Data analysis was performed by Petra Schwalie (European Bioinformatics Institute, Hinxton, UK)

4.3.2.1 Sequence Alignment and Peak-calling

Reads were aligned to the reference genome and visualized in the genome browser as described in **section 3.3.2.1**. Regions of high ChIP enrichment were detected with CCAT 3.0 (Xu et al., 2010) on individual replicates using Input or ChIP against histone H3 as controls. Regions from different cell types were merged and processed further to identify differences similarly as in the case of Pol II (**section 3.3.2.1**).

4.3.2.2 Correlations and Profile Plots

Spearman correlations were calculated based on read counts inside corresponding ROI categories. Values shown in **Figures 4.9, 4.11 and 4.18** are means of the two replicates using the "Head-Neuron-Glia" ROI sets. To display neuron-glia and head-embryo ChIP enrichments in **Figures 4.12 and 4.17**, 150 bp extended ChIP-seq reads were summed over 5 bp bins in a 2 kb window centered around the summits of Pol II ROIs and divided by a normalization factor based on the total read number of the individual replicates. Counts were then visualized with Treeview (Saldanha et al., 2004). For simplicity, only one replicate was displayed, but both replicates show similar patterns. To visualize embryonic data, scaling factors estimated by DESeq were used for normalizing both Pol II and H2A.Z whole-head replicates and the single embryo sample. Processed tiling array data for H2A.Z and insulator binding proteins were obtained from the modENCODE project (GSE32729 and GSE32730). **Figure S6B** displays the whole-head endogenous, tagged, glial, neuronal and embryonic H2A.Z profiles at ChIP-chip enriched locations derived from modENCODE early (GSE32730) and late (GSE32729) H2A.Z embryonic data. 150 bp extended ChIP-seq reads were summed over 10 bp bins in a 5 kb window sorted based on the ChIP-chip peak score and divided by the size-factors estimated by DESeq when available.

4.3.2.3 Expression Analysis

Paired-end RNA-seq reads obtained from whole male fruit fly heads were aligned with Tophat v1.0.14 to Ensembl transcript annotations. Transcript level expression values were estimated using Cufflinks v0.9.0 (Trapnell et al., 2010). Only single-transcript genes were used in the analysis to avoid ambiguity in read assignment to different isoforms. The dataset was also split into six categories depending on the associated expression values: 0 - no expression and 1-5 based on expression quantiles, displayed in **Figure 4.2**. For these expression categories and the subset of genes located on the + strand, **Figure 4.2** displays the average normalized Pol II profiles in 5 bp bins and a 2 kb window around the TSS. Reads were normalized by dividing with the input and multiplying with a scaling factor based on the library size.

4.3.2.4 H2A.Z Domain Definitions

We determined uninterrupted domains of genes overlapping H2A.Z for each chromosome by constructing binary vectors of 0 and 1. We plotted the distribution of

numbers of consecutive H2A.Z-positive genes for both real and simulated data (1000 repetitions, obtained by shifting the H2A.Z ROIs with random distances drawn from a normal distribution; data not shown). The sizes of simulated H2A.Z clusters were significantly lower than the real H2A.Z data (t-test, p-value $<2*10^{-16}$). We fitted a normal distribution to the simulated data and determined a cluster size cutoff for each chromosome corresponding to a significance level of 0.05, based on which we selected the final H2A.Z domains (**Figure 4.19**).

4.3.2.5 Association with Chromatin Domains and Insulator Binding Proteins

Chromatin domains were publicly available under accession number GSE22069 (Filion et al., 2010). We used the five color domains provided, as well as transition areas between different colors in our analysis. For each category of Pol II/H2A.Z-bound regions, we asked what their distribution with respect to the different chromatin types was and plotted the top three domains in **Figure 4.21**. If a single region overlapped two different domains, this was considered a category as such, and was displayed in shading, accordingly: black-striped-yellow for regions present in both “yellow" and “black" chromatin and yellow-striped-red ROIs for both “red" and “yellow" regions. We plotted scaled (based on the library size) averaged whole-head Pol II, H2A.Z, Input and H3 read profiles in a 5 kb window centered on “yellow"-to-”black" and “red"-to-”blue" transitions in **Figure 4.22**.

Processed tiling array data for insulator binding proteins were available from the modENCODE project (Negre et al., 2010). We used the GSM409067, GSM409068, GSM409069, GSM409070, GSM409071, GSM409073, GSM409074 and GSM409077 insulator datasets for BEAF-32, CP190, CTCF, GAF, MDJ4 and Su(Hw). We calculated overlaps of the different insulator sets [Class I and class II insulators, as defined in (Negre et al., 2010), as well as BEAF-32, CP190 and CTCF-C) with the “Head-Embryo-Neuron-Glia" H2A.Z-only, H2A.Z and Pol II, Pol II-only regions and report the associated fractions in **Figures 4.23**.

4.4 Results

4.4.1 H2A.Z is an active mark in differentiated cell types of the fly CNS

4.4.1.1 Profiling H2A.Z in the fly head

Recently, H2A.Z has been mapped in *Drosophila* cell culture (i.e. S2 cells) and in developmental stages such as in embryos (Weber et al., 2010; Mavrich et al., 2008; ThemodENCODEConsortium et al., 2010). First, we wanted to confirm these findings in terminally differentiated cells. Therefore, we used an anti-H2A.Z antibody (Leach et al., 2000) to generate genome-wide profiles in the *Drosophila* head. As expected, more than three quarter of H2A.Z enriched regions (ROIs) localize to annotated transcription start sites **(Figure 4.1)**.

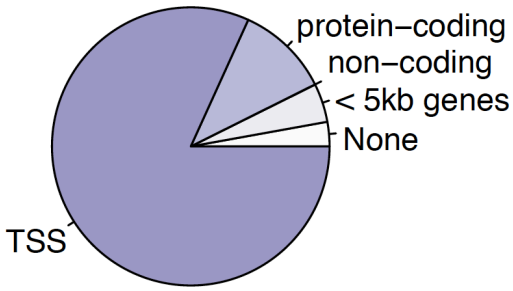


Figure 4.1 H2A.Z ROIs mainly localize to TSSs.

Proportion of H2A.Z regions located at TSSs, protein coding, non-coding and within 5kb of genic regions.

RNA polymerase II localizes about +50 bp downstream of the TSS, whereas H2A.Z peaks further downstream around +200 bp of the TSS **(Figure 4.2)**. This agrees with previous reports indicating that *Drosophila* H2A.Z is enriched on the +1 nucleosome and to some extent further downstream. Next, we split genes according their gene expression values derived from our RNA-Seq analysis, which we performed on *Drosophila* head samples. Gene expression positively correlates with the average Pol II peak maximum, as expected (**Figure 4.2**). H2A.Z is absent from genes with no detectable mRNA levels, but is present at expressed genes. However, the highest expression quantile carries only moderate levels of H2A.Z, whereas low to moderately expressed genes has high levels of H2A.Z. This can be due to the lower levels of nucleosomes at the highest gene expression subset. To test this, we compared H3 occupancy in the gene expression classes and confirmed that highly expressed genes carry less histone H3 in general (**Figure 4.2**).

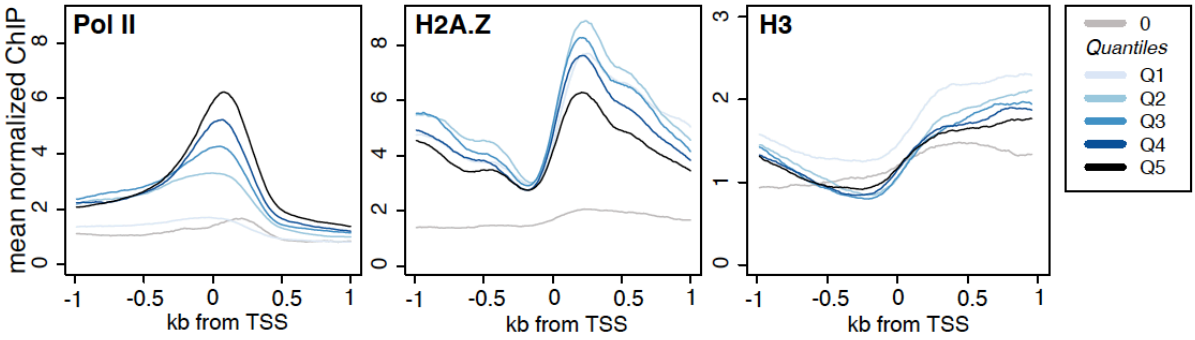


Figure 4.2 Pol II, H2A.Z and H3 average profiles centered on the TSS in RNA-seq expression quantiles.

Mean normalized ChIP enrichment of Pol II, H2A.Z and H3, respectively in a 2 kb region centered on the TSSs of genes with different RNA-seq expression levels (0, Q1<Q5).

To further investigate the level of H2A.Z on inducible genes, I performed heat shock experiments in adult *Drosophila*, as previously described for S2 cells (Adelman et al., 2006). I carried out ChIP against RPB3, H2A.Z and canonical histones such as H2A and H3 on heat shocked and control *Drosophila* head chromatin (**Figure 4.3**). Upon heat shock, Pol II occupied the entire Hsp70 gene and nucleosomes including H2A.Z are evicted from the transcriptional unit. Genes that were not induced by heat shock, do not show any changes in histone occupancy upon heat shock (**Figure 4.4**). In contrast, Pol II binding decreases at these genes, suggesting a genome-wide re-localization of Pol II towards heat shock genes (**Figure 4.4**).

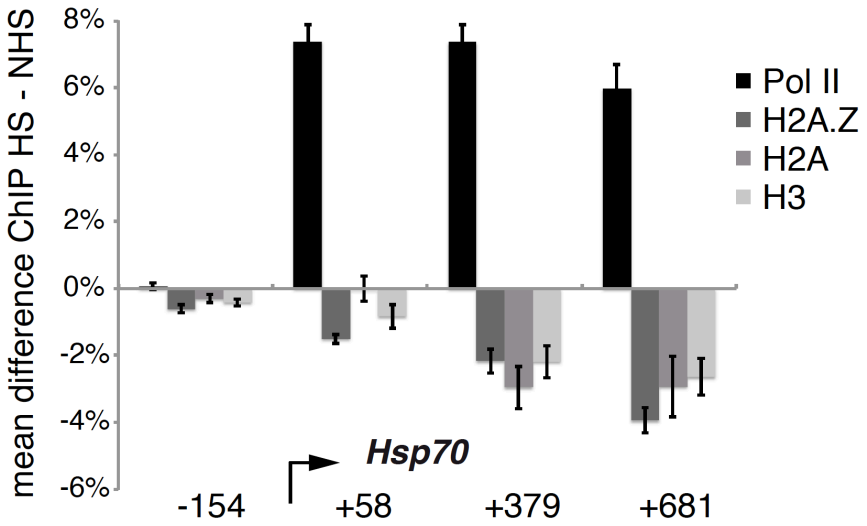


Figure 4.3 Heat shock induced accumulation of Pol II and depletion of histones at the *Hsp70* locus in fly heads.

The difference between ChIP enrichment (input percentage: heat shock, HS *vs.* no heat shock, NHS) is shown at 4 sites across the *Hsp70* gene (-154, +58, +379 and +681 relative to the TSS).

Upon heat shock, Pol II occupies the entire gene and nucleosomes, including H2A.Z-containing nucleosomes, are strongly depleted. The graph shows the mean of at least four biological replicates and the standard deviation is indicated with error bars.

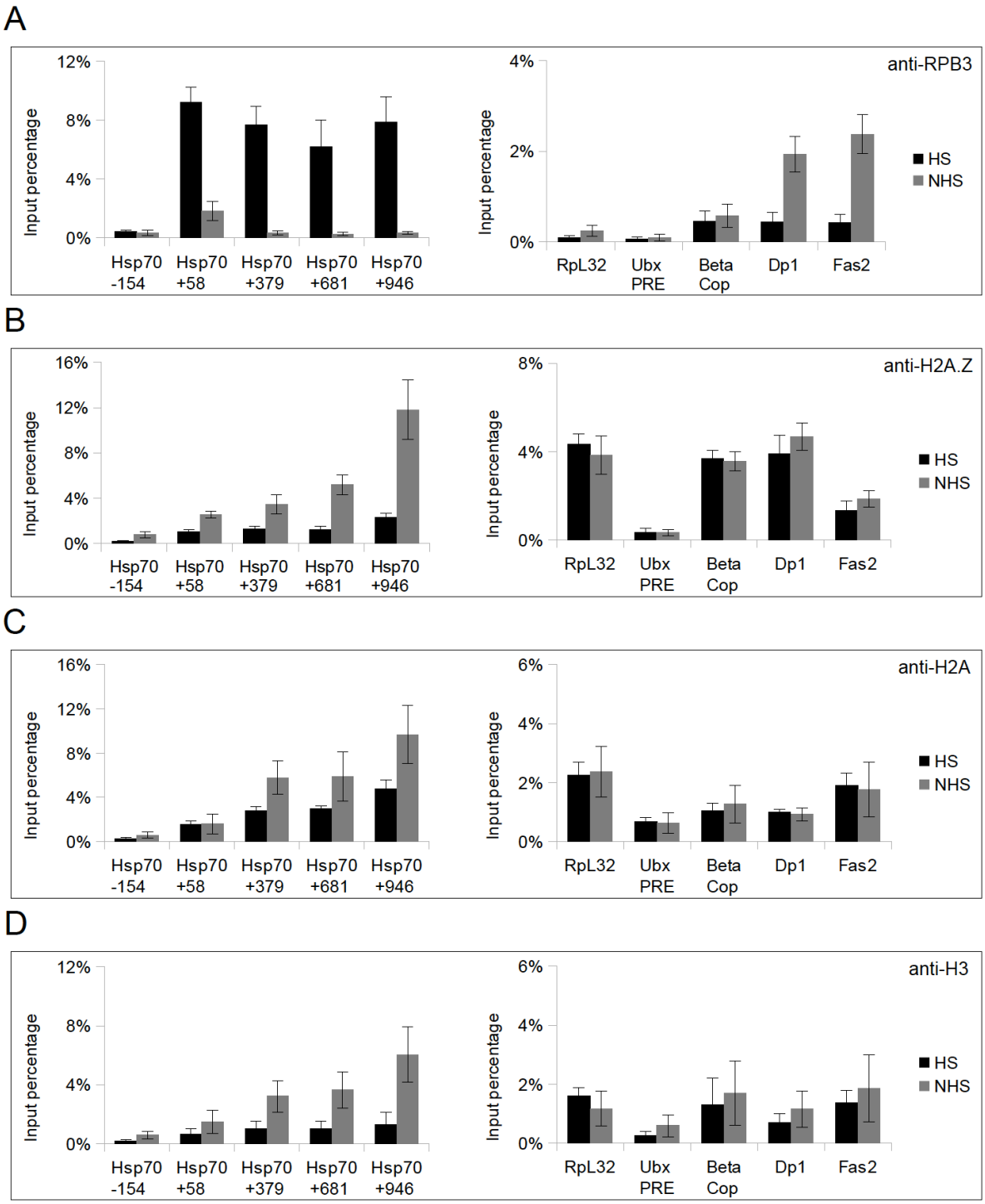


Figure 4.4 Heat shock induction results in increased Pol II and decreased histone (i.e. H2A.Z, H2A and H3) occupancy at the *Hsp70* locus, whereas decreased Pol II and unchanged histone occupancy at non-heat-shock genes

ChIP enrichment is shown as Input percentage against RPB3 (A), H2A.Z (B), H2A (C) and H3 (D) at the hsp70 gene (left panel) and at not heat shock inducible genes (right panel). RPB3 binds to

hsp70 and gets depleted from other genes (A). In contrary, nucleosomes (H2A.Z, H2A, H3) are removed from hsp70 and are stable at the control genes.

In summary, heat shock experiments on differentiated *Drosophila* tissues agree with previous findings from S2 cells (Adelman et al., 2006; Petesch et al., 2008). Therefore, the relatively lower binding of H2A.Z at highly expressed genes is due to lower nucleosome occupancy

Although H2A.Z is found at active genes, not all H2A.Z regions overlap with Pol II peaks (**Figure 4.5**). We find 1493 H2A.Z-only and 2547 Pol II-only regions in the head, suggesting that the H2A.Z might have a Pol II independent role. On the other hand, Pol II-bound genes without H2A.Z might have a specific function. To analyze some of the connections between Pol II and H2A.Z bioinformatically, we first tested the expression differences between genes carrying H2A.Z-only, H2A.Z+Pol II, Pol II only and with none of the two factors (**Figure 4.5**). In general, Pol II-associated genes show higher expression, while genes lacking Pol II and H2A.Z show the lowest expression. H2A.Z-only genes, in contrast to H2A.Z+Pol II and Pol II-only, have a moderate level of expression.

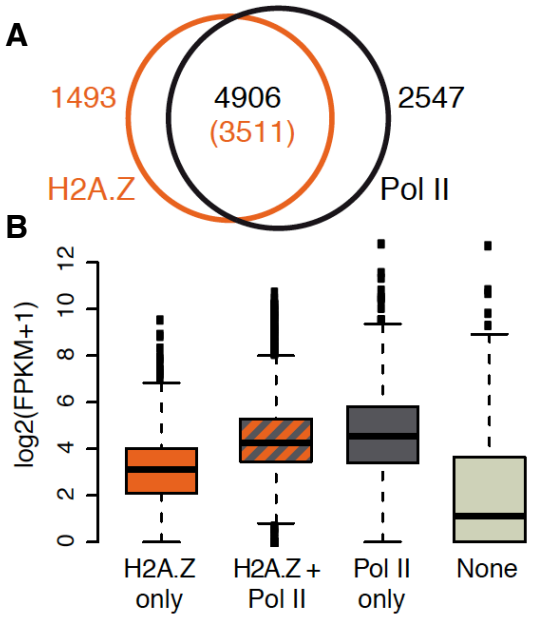


Figure 4.5 Comparison of H2A.Z- and Pol II-bound regions in the head data.

A) Venn diagram showing overlap of H2A.Z ROI (orange) and Pol II ROI (black). **B)** Expression values (FPKM) of genes associated with ROIs of H2A.Z-only, H2A.Z+Pol II, Pol II-only and None subclasses.

Second, we compared our gene classes with or without H2A.Z and/or Pol II to other existing gene expression profiles such as the FlyAtlas (**Figure 4.6**; Chintapalli et al., 2007). Expression values of the FlyAtlas confirm our findings from the head, where Pol II associated genes show higher RNA levels and genes bound by H2A.Z-only have a lower expression. Interestingly, genes bound by H2A.Z look very similar among FlyAtlas

tissues, regardless whether they exhibit Pol II peaks or not (**Figure 4.6**). Standard deviation of expression values among these tissues is significantly lower at genes associated with H2A.Z. In contrast, Pol II-only genes deviate more, suggesting these genes are more specifically regulated (**Figure 4.6B**).

Genome-wide analysis of H2A.Z and Pol II in the *Drosophila* head thus reveals a potential novel function of H2A.Z. Genes carrying Pol II having H2A.Z incorporated can be distinguished from genes with Pol II lacking H2A.Z. I hypothesize that these latter ones are specifically regulated, whereas H2A.Z-associated genes show a ubiquitous, cel-type-independent function.

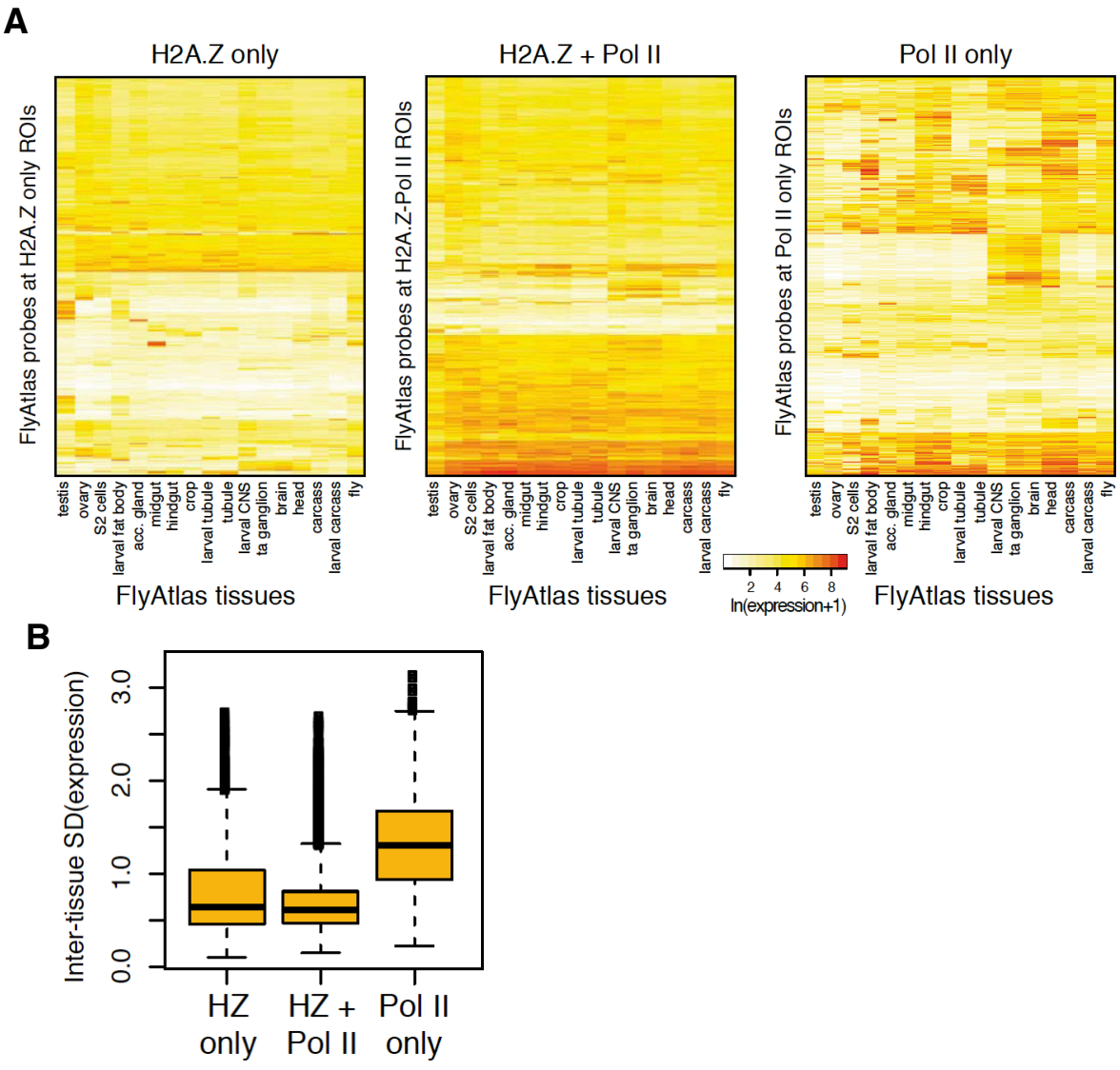


Figure 4.6 H2A.Z-bound ROIs show similar, whereas Pol II-only ROIs diverse RNA expression in the FlyAtlas.

A) Heatmaps showing expression levels (*color range*) in FlyAtlas tissues (Chintapalli et al., 2007). Genes bound by Pol II show higher expression (*second and third panel*) compared to

H2A.Z-only genes (*first panel*). Genes not associated with H2A.Z ROIs (*third panel*) show larger differences in expression among tissues.
B) Standard deviations (SD) of expression values across FlyAtlas tissues (Chintapalli et al., 2007) are shown in H2A.Z-only (HZ), H2A.Z+Pol II (HZ+Pol II) and in Pol II- only gene subclasses. Genes associated Pol II without H2A.Z deviate the most.

4.4.1.2 CAST-ChIP reveals that H2A.Z marks cell-type-invariant genes

To test the hypothesis that histone H2A.Z may have a role in the regulation of or correlate with cell-type-independent gene regulation, we sought to compare cell-type-specific H2A.Z maps using CAST-ChIP (see **chapter 3**). First, we tested whether a GFP-tagged H2A.Z construct, expressed under its genomic promoter (Clarkson et al., 1999), correctly incorporates to genomic sites obtained by an antibody against endogenous H2A.Z (Leach et al., 2000). ChIP profiles of H2A.Z and H2A.Z-GFP (using the same GFP antibody as for Pol II CAST-ChIP) were very similar with an overlap of 96% (**Figure 4.7**) and with a Spearman correlation of R=0.95 (**Figure 4.7B**). This high correspondence indicates that an H2A.Z-GFP transgene construct driven by a cell-type-specific promoter would profile and faithfully report on the incorporation of H2A.Z in distinct, fully-differentiated cell populations within the *Drosophila* brain.

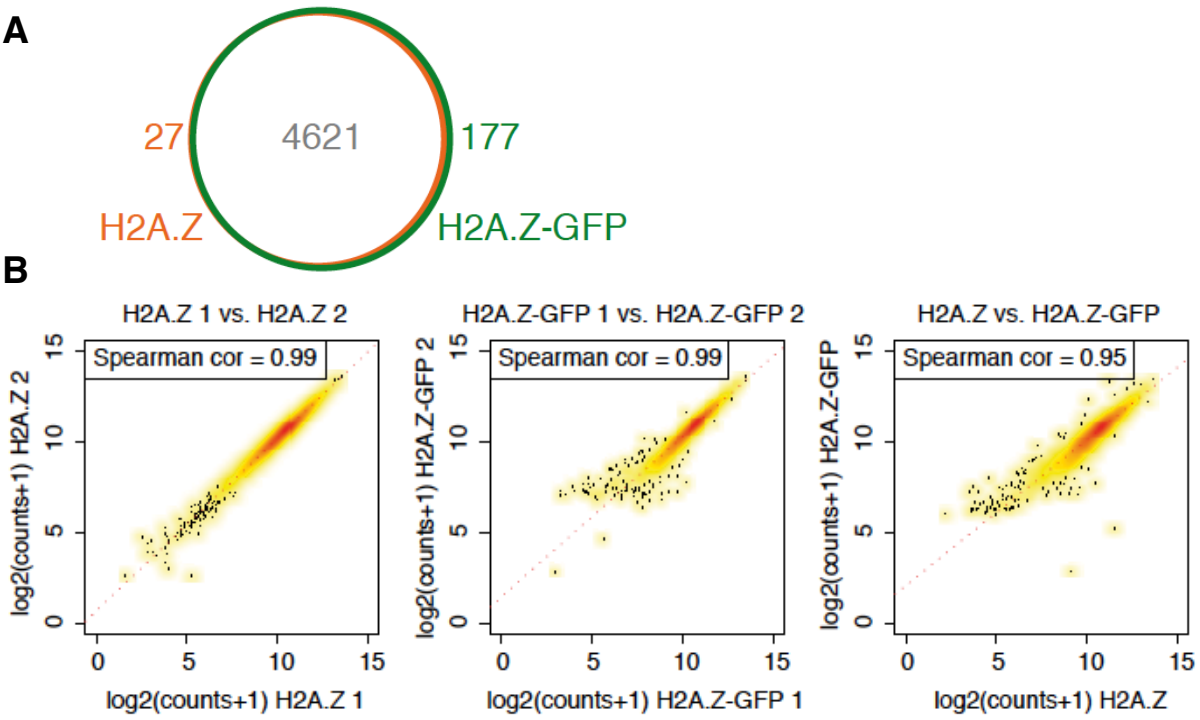


Figure 4.7 Endogenous H2A.Z binding correlates well with tagged H2A.Z-GFP.

A) Venn diagram showing the overlap between endogenous H2A.Z (orange) and tagged H2A.Z-GFP (green) ROIs.
B) Spearman correlation of endogenous H2A.Z and tagged H2A.Z-GFP replicates (R=0.99; R=0.99, respectively) as well as endogenous and GFP-tagged H2A.Z samples (R=0.95).

To generate cell-type-specific H2A.Z profiles, we expressed the C-terminally GFP-tagged reporter in neurons and glia cells, similarly as described for Pol II (see **chapter 3**). CAST-ChIP comparison of the two cell types surprisingly reveals that they are almost identical (**Figure 4.8**). We find only few hundred minor differences between neurons and glia cells using the same DESeq cutoffs as for Pol II (**Figure 4.8B**; for a comparison see **Figure 3.9**). Spearman´s correlation at the H2A.Z ROI across neuronal and glial samples is almost as high (R=0.96) as for biological replicates (R>0.98; **Figure 4.9**), suggesting that H2A.Z binding is essentially identical in these two distinct cell populations.

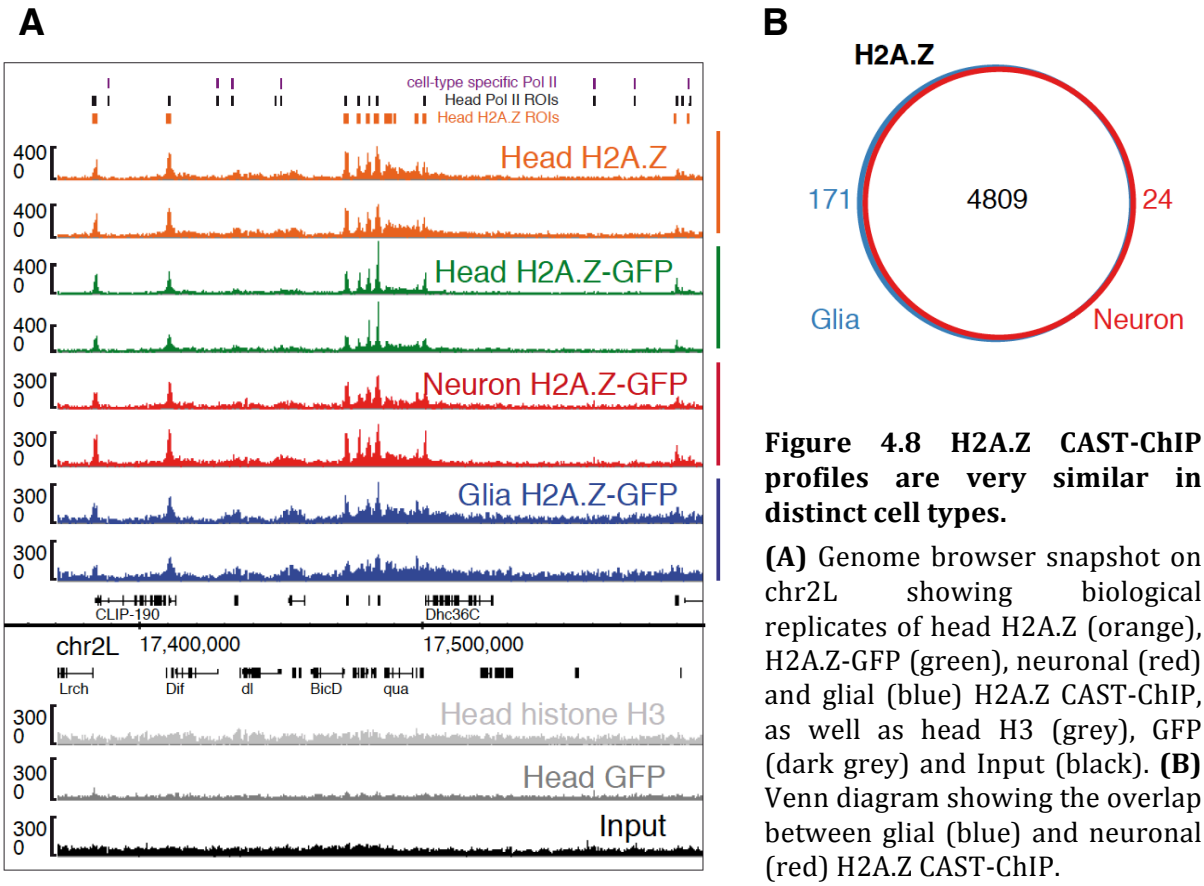


Figure 4.8 H2A.Z CAST-ChIP profiles are very similar in distinct cell types.

(A) Genome browser snapshot on chr2L showing biological replicates of head H2A.Z (orange), H2A.Z-GFP (green), neuronal (red) and glial (blue) H2A.Z CAST-ChIP, as well as head H3 (grey), GFP (dark grey) and Input (black). **(B)** Venn diagram showing the overlap between glial (blue) and neuronal (red) H2A.Z CAST-ChIP.

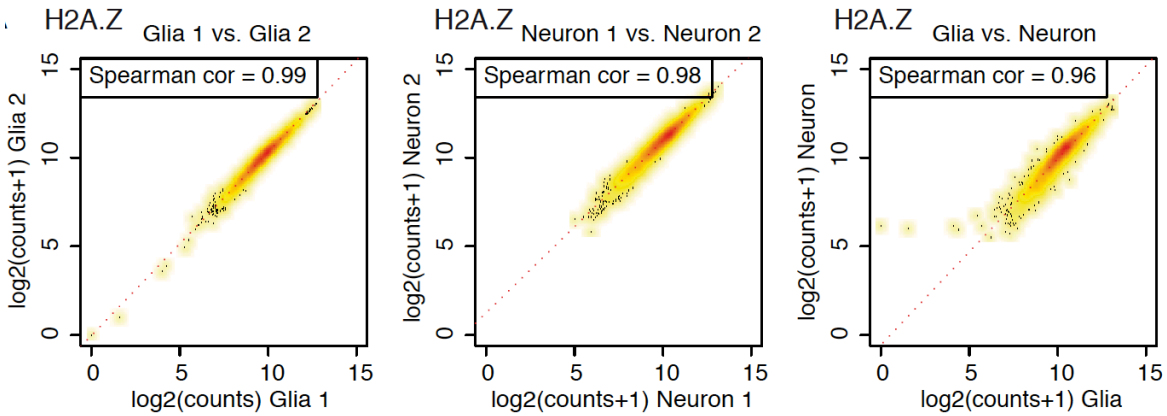


Figure 4.9 H2A.Z CAST-ChIP profiles show high correlation between neurons and glia.
Spearman´s correlation is shown between biological replicates of glial, neuronal as well as between glial and neuronal H2A.Z CAST-ChIP data.

Interestingly, cell-type-specific genomic sites marked by Pol II are depleted from H2A.Z. In contrast, Pol II-invariant regions are highly bound by H2A.Z (**Figure 4.10**). Heatmaps, where peak regions were ranked according to the difference between glial and neuronal Pol II occupancy, clearly show that common, cell-type-invariant regions carry H2A.Z, while cell-type-specific Pol II regions lack H2A.Z (**Figure 4.10**).

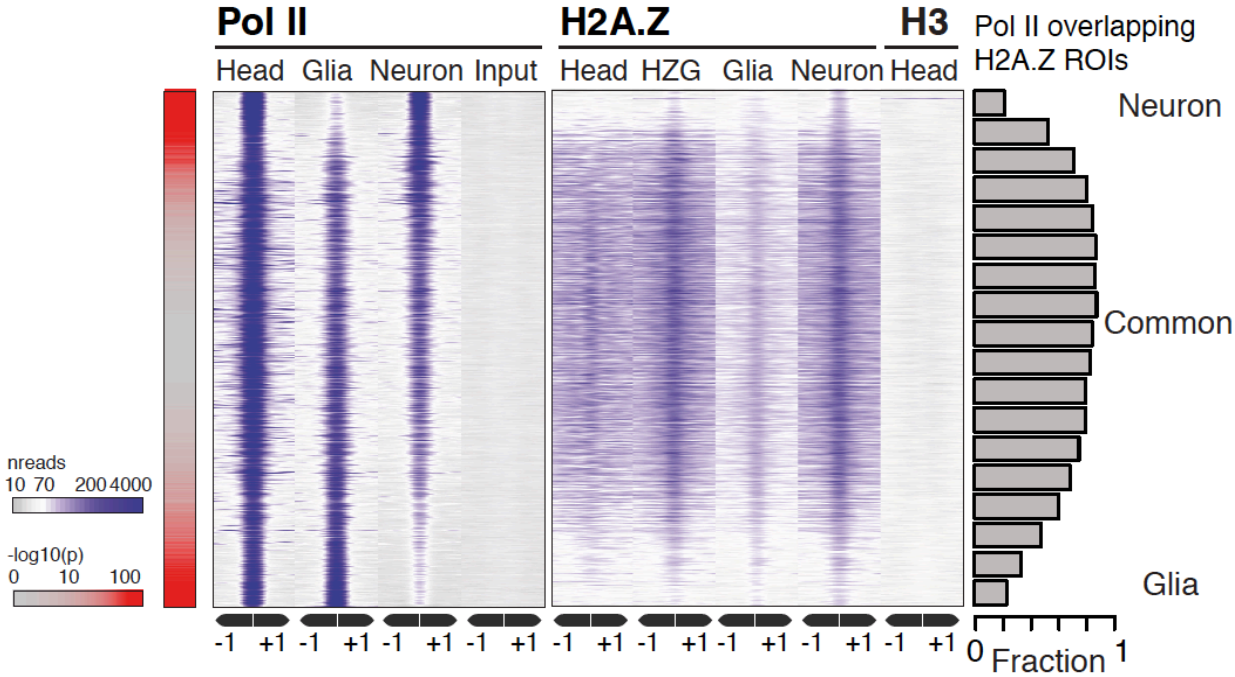


Figure 4.10 H2A.Z is enriched at cell-type-invariant and absent from cell-type-specific Pol II ROIs.
Heatmaps showing head, glial and neuronal Pol II, head Input and head, H2A.Z-GFP (HZG), glial and neuronal H2A.Z enrichments, as well as head histone H3. Regions are centered to the Pol II peak maxima and sorted from lowest to highest signal in glia (p-values are shown in red on the left). The fraction of H2A.Z ROIs overlapping Pol II ROIs is shown as barplots on the right.

Comparison of ROIs at the gene level confirm these results. 90% of the genes overlapping Pol II in the presence of H2A.Z ROIs are cell-type-invariant (**Figure 4.11A**, left plot, grey fraction) and only about 10% is cell-type-specific (**Figure 4.11A**, left plot, red fraction). In contrast, more than half of the genes bound by Pol II lacking H2A.Z are cell-type-specific (**Figure 4.11A**, right plot, red fraction). From the H2A.Z point of view, genes overlapping H2A.Z are almost always cell-type-invariant (**Figure 4.11B,** grey

fraction), independently of the presence or absence of Pol II (**Figure 4.11B**, left and right plot, respectively).

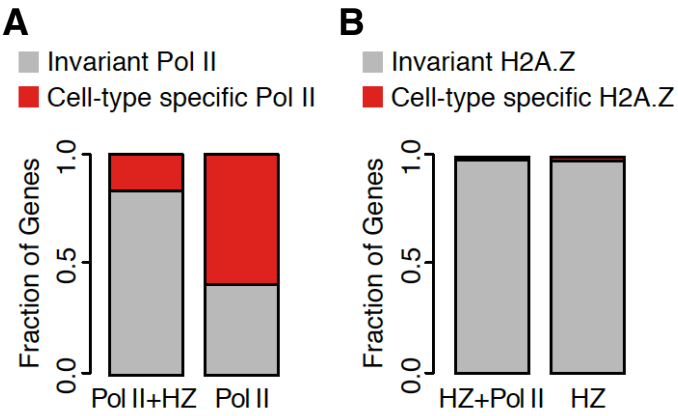


Figure 4.11 Association of genes with cell-type-specific and invariant Pol II and H2A.Z ROIs.

A) Barplots showing the fraction of genes overlapping cell-type-specific (red) and cell-type invariant (grey) Pol II ROIs in the presence (HZ+Pol II) or absence of H2A.Z (Pol II-only).
B) Same for H2A.Z: fraction of genes overlapping cell-type-specific (red) and cell-type-invariant (grey) H2A.Z ROIs in the presence (HZ+Pol II) or absence of Pol II (HZ-only).

In summary, comparison of neuronal and glial CAST-ChIP profiles suggests that H2A.Z marks genes that are potentially active in both cell types. Further, our data indicate that cell-type specific genes are regulated in an H2A.Z independent manner. H2A.Z thus appears to have distinct functions at cell-type-specific and at ubiquitous genes.

4.4.1.3 H2A.Z is maintained across cell types of different developmental origin

Neurons and glia are distinct cell types within the fly nervous system, but are both derived from the ectoderm (see **chapter 2.1**). The similarity in H2A.Z profiles between neurons and glia could thus be the result of the shared developmental history for neurons and glia. To check whether H2A.Z enrichment may more broadly correlate with ubiquitous genes, consistent with a role in ubiquitous gene regulation, we profiled a mesodermal head tissue, the fat body, using CAST-ChIP. The fat body has analogous function to the mammalian liver and adipocytes (see **section 2.1.2.3**). Tripartite comparison of glia, neurons and fat body cells, identifies significant differences in the genomic binding sites for Pol II (**Figure 4.12A**). However, consistent with our data in neurons and glia, we observe only minor changes in H2A.Z distribution (**Figure 4.12B**). We find 4725 Pol II-enriched regions in all cell types (cell-type-invariant), from which more than 80% associate with H2A.Z (**Figure 4.12C**). Out of 1540 cell-type-specific regions, only 30% show overlap with H2A.Z ROIs (**Figure 4.12C**). This suggests that H2A.Z marking invariant genes is not restricted to ectodermal cells, but may rather be a ubiquitous function of this histone variant.

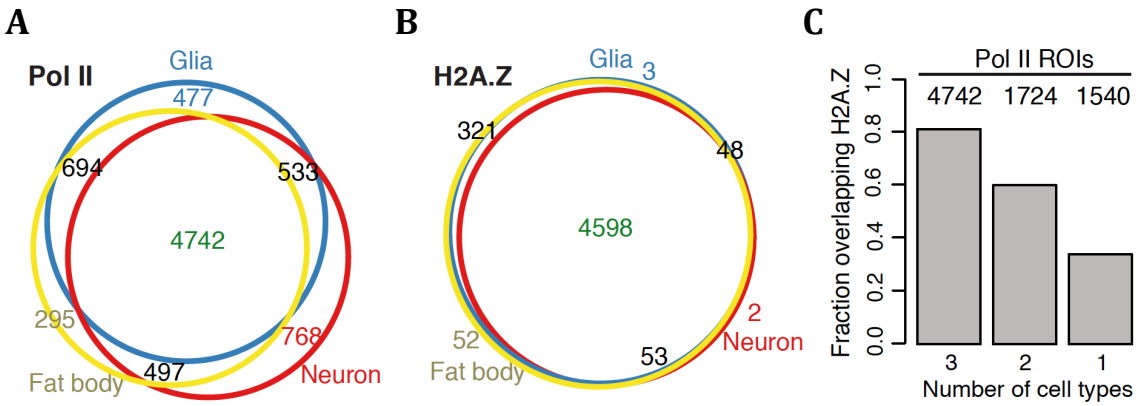


Figure 4.12 Pol II ROIs are more different, whereas H2A.Z ROIs are similar in three cell types with distinct developmental origin.

Venn diagram showing the overlap across glia- (red), neuron- (blue) and fat body-specific (yellow) ROIs of Pol II **(A)** and H2A.Z **(B)**. **(C)** Barplot indicating the number of Pol II ROIs found in 3, 2 and 1 cell types and fraction of these ROIs overlapping H2A.Z.

As an example, I chose cell-type-specific genes that have a neuronal function based on FlyBase (**see section 3.4.2**, **Table 3.2** and **section 2.1.2.1**). These selected genes carry a sharp peak in neurons and in total head, but lack Pol II in glia and the fat body (**Figure 4.13**). None of them show an enrichment of H2A.Z close to their promoter. They usually have a long first intron and are long in length.

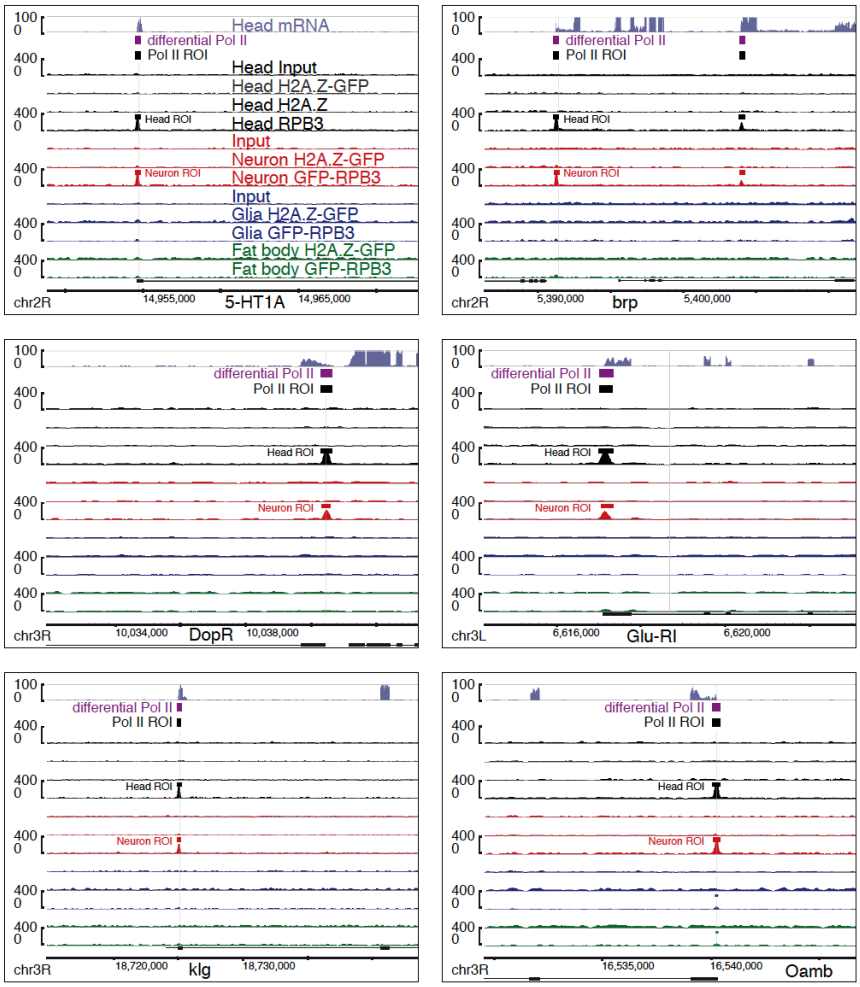


Figure 4.13 Selected neuronal genes with neuronal Pol II peak lacking H2A.Z.
Genome browser snapshots of genes with known neuronal function (based on Flybase) including *Serotonin receptor 1A* (*5-HT1A*), *bruchpilot* (*brp*), *Dopamine receptor* (*DopR*), *Glutamate receptor I* (*Glu-RI*), *klingon* (*klg*) and *Octopamine receptor in mushroom bodies* (*Oamb*). The tracks show in *black/grey*: Head mRNA, Head Input, Head H2A.Z-GFP, Head H2A.Z, Head RPB3; in *red*: head Input (genotype: elav-GAl4; UAS-H2A.Z-GFP), Neuron H2A.Z-GFP, Neuron GFP-RPB3; in *blue*: head Input (genotype: repo-GAl4; UAS-H2A.Z-GFP), Glia H2A.Z-GFP, Glia GFP-RPB3 and in *green*: Fat body H2A.Z-GFP and Fat body GFP-RPB3, respectively in each snapshot. Significantly different Pol II regions (*purple*), total Pol II ROIs (*black*), and track-specific ROIs are indicated as small rectangles above the tracks.

The three cell types, we investigated, separate during gastrulation, which is a relatively early step of embryonic development (**section 2.1.1**). If H2A.Z sites are conserved among these cell types, I hypothesized that H2A.Z may already be established in the embryo and be maintained to the adult. To test this, I generated H2A.Z profiles from 0-6 hour embryos and compared these profiles to those of the adult head (**Figure 4.14**). Here, I used antibodies against endogenous H2A.Z and RPB3 (Pol II). We find thousands of differential Pol II regions between embryo and adult head (**Figure 4.14B**). In contrast, H2A.Z differs only by hundred genomic sites using the same cutoff in DESeq (**Figure 4.14B**). To further analyze the relationship between Pol II and H2A.Z, we ranked the ROIs according the Pol II differences between embryo and head; H2A.Z is present at stage-invariant Pol II regions, but is absent from stage-specific Pol II regions (**Figure 4.15**). Spearman´s correlations across replicates and embryo *vs*. head show the same trend; Pol II is extensively different at different developmental stages, while H2A.Z is essentially the same across all stages (**Figure 4.16**).

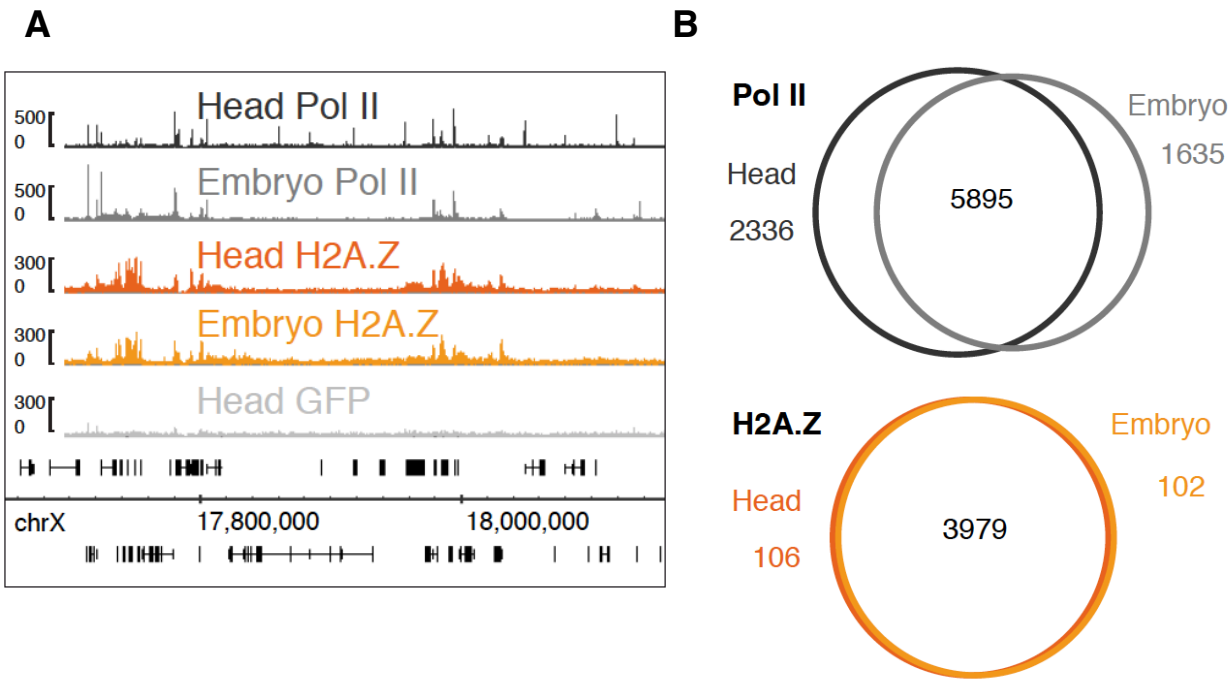


Figure 4.14 Comparison of adult head and embryo on the level of Pol II and H2A.Z.

(A) Genome browser snapshot from the X chromosome showing profiles of head (black) and embryo (grey) Pol II, as well as head (dark orange) and embryo (light orange) H2A.Z and head GFP control (light grey). **(B)** The Venn diagram shows the overlap between ChIP-enriched regions of the adult head and 0-6 hour embryos using antibodies against RPB3 (Pol II; upper panel) and H2A.Z (bottom panel).

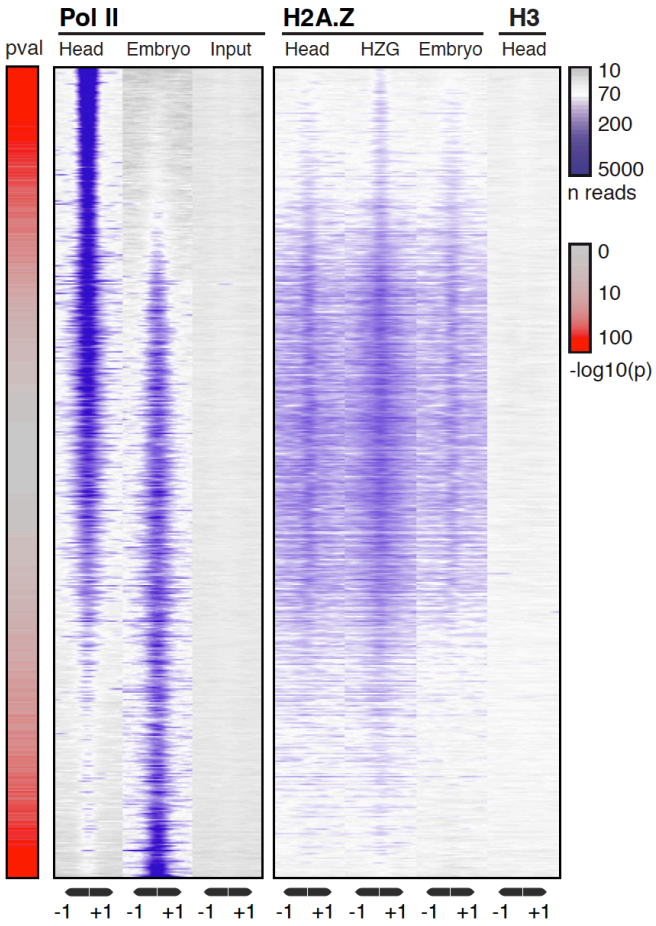


Figure 4.15 H2A.Z is enriched at stage-invariant and absent from stage-specific Pol II ROIs.

Heatmaps showing head, embryo Pol II, Input and head, embryo H2A.Z, H2A.Z-GFP (HZG) as well as head H3 ROIs. Regions are centered and ranked according the difference between head and embryo Pol II.

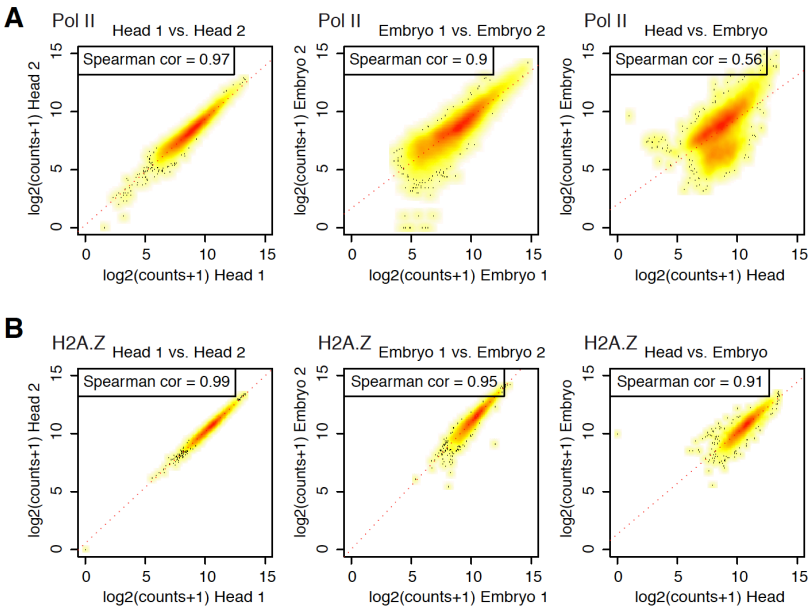


Figure 4.16 Pol II ChIP profiles show low, whereas H2A.Z high correlation across embryo and head.

Spearman´s correlation is shown between biological replicates of embryo, head as well as between embryonic and head Pol II **(A)** and H2A.Z **(B)** ChIP data.

4.4.2 H2A.Z associates with chromatin domains that display ubiquitous gene expression

Chromatin is organized into higher order structures (see **section 2.2.2.3**). Co-regulated genes tend to cluster together and are separated from specifically regulated genes (Lercher et al., 2002; Weber et al., 2011). Weber et al. determined so-called "Tau-clusters" of genes that are located next to each other and are co-regulated. They defined co-regulated genes as "low-specificity" genes based on a Tau factor derived from the FlyAtlas expression values (Chintapalli et al., 2007). These genes can be called ubiquitous genes, since they are expressed independently of developmental stage and cell type. We performed a similar type of clustering analysis for genes carrying the histone variant H2A.Z in the fly head (see **Methods 4.3.2.4**). We found that H2A.Z-associated genes cluster together and these clusters overlap with the Tau cluster well (**Figure 4.17**). In addition, the fraction of regions carrying both H2A.Z and Pol II is associated significantly more highly with Tau clusters compared to genes marked by only Pol II (**Figure 4.17B**). Thus, H2A.Z marks clusters of ubiquitous genes.

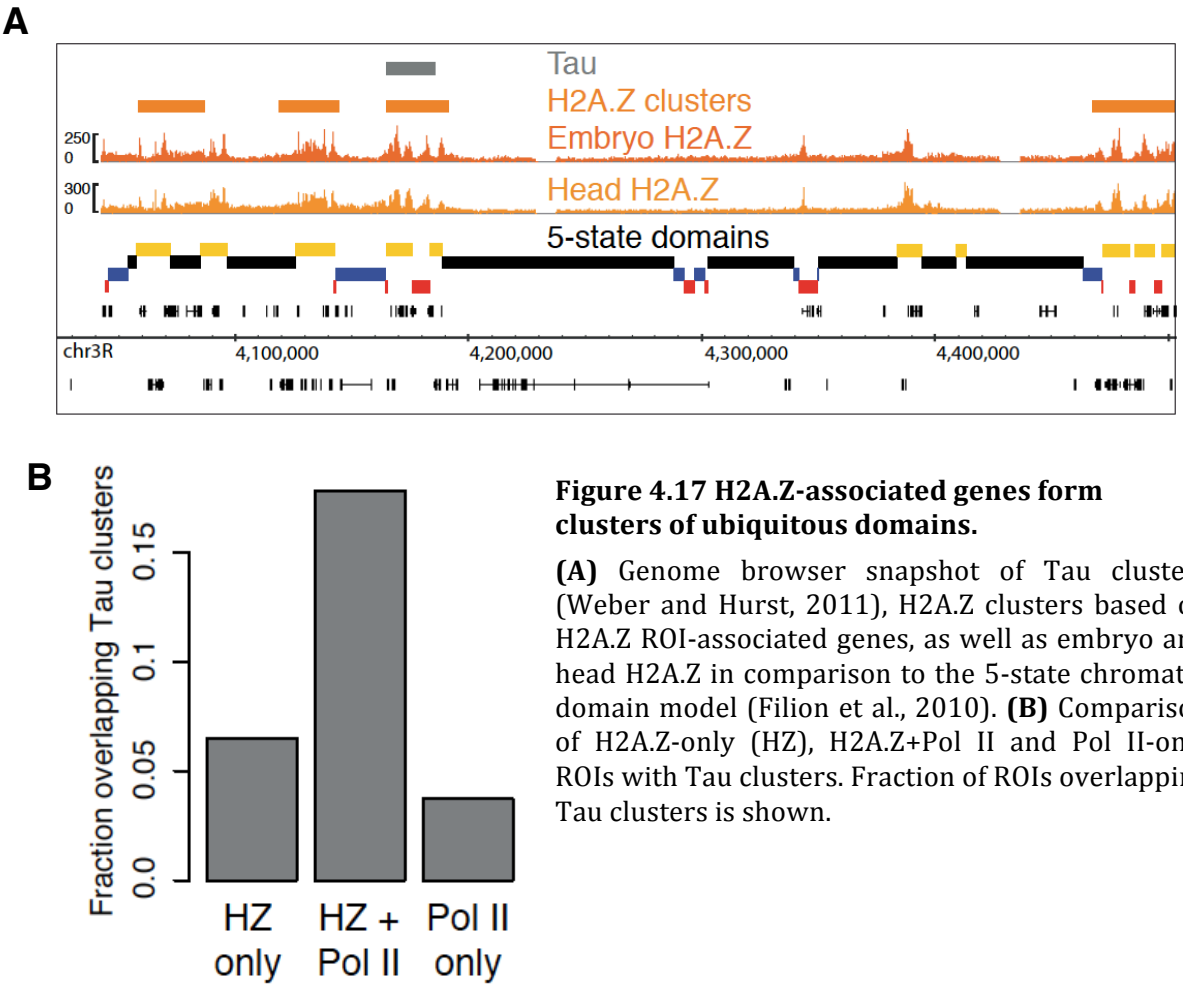


Figure 4.17 H2A.Z-associated genes form clusters of ubiquitous domains.

(A) Genome browser snapshot of Tau clusters (Weber and Hurst, 2011), H2A.Z clusters based on H2A.Z ROI-associated genes, as well as embryo and head H2A.Z in comparison to the 5-state chromatin domain model (Filion et al., 2010). **(B)** Comparison of H2A.Z-only (HZ), H2A.Z+Pol II and Pol II-only ROIs with Tau clusters. Fraction of ROIs overlapping Tau clusters is shown.

Ubiquitous genes have an altered promoter structure compared to specifically regulated genes (see **section 2.2.1.1**). The former ones carry broad promoters, whereas the latter ones have narrow, peaked start sites (Hoskins et al., 2011). Finding H2A.Z at ubiquitous genes, I was curious whether H2A.Z-bound genes have also distinct promoters characteristics. Indeed, about 70% of H2A.Z marked promoters are broad with and without Pol II enrichment. In contrast, Pol II-only genes associate mainly with unclassified and sharp TSSs (**Figure 4.18**). Thus, H2A.Z-bound genes also share promoter features of ubiquitous genes.

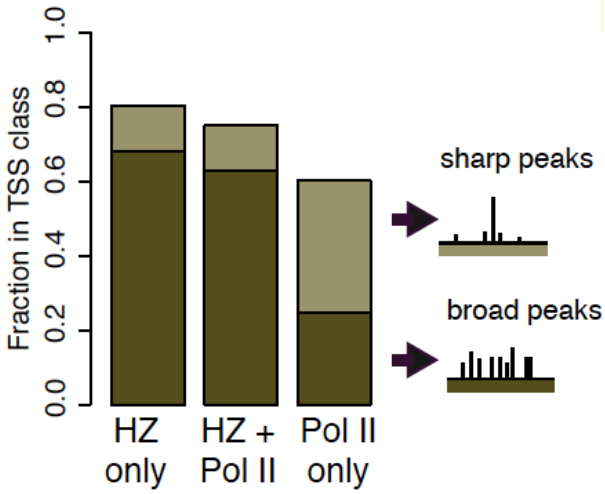


Figure 4.18 H2A.Z ROIs (with or without Pol II) mainly associate with broad promoters.

Fraction of H2A.Z-only (HZ), H2A.Z+Pol II and Pol II-only ROIs in promoter (TSS) classes such as sharp (light), broad (dark) and unclassified (remaining fraction).

Chromatin can be further divided into epigenetic domains, which carry similarities in protein occupancy and/or histone modification (see **section 2.2.2.3**). Filion et al. determined a 5-state model of chromatin, where they split the genome into 5 domains based on the Dam-ID maps of 53 chromatin-associated proteins (Filion et al., 2010). These domains include the constitutively active (YELLOW) and dynamically active (RED) chromatin. In contrast, the BLACK chromatin is in general silent, but also dynamically regulated during development. The other two classes are heterochromatic (see **section 2.2.2.3**). I wondered whether H2A.Z-associated regions could be integrated to these domain classes. Genome browser snapshots already suggested that clusters of H2A.Z associate with the constitutive YELLOW chromatin (**Figure 4.17A**). Genome-wide comparison of H2A.Z ROIs and the 5-state domains confirmed this observation (**Figure 4.19**). H2A.Z- and Pol II-bound regions are mainly enriched in the YELLOW chromatin, H2A.Z-only regions in the YELLOW and BLACK, whereas Pol II-only

regions in RED and BLACK chromatin (**Figure 4.21**). Taken together, H2A.Z is embedded into domains that have characteristic of ubiquitous (or constitutive) gene expression.

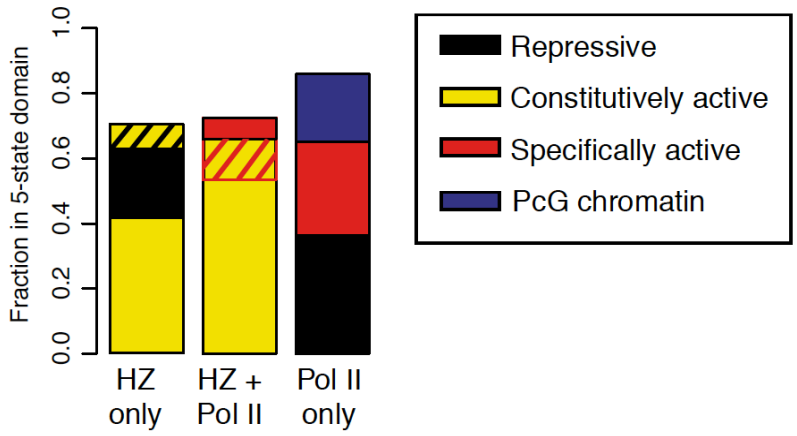


Figure 4.19 H2A.Z ROIs (with or without Pol II) mainly associate with constitutively active chromatin.

Fraction of H2A.Z-only (HZ); H2A.Z+Pol II and Pol II-only ROIs in the 5-state model of chromatin domains (see legend). Boundaries between two domains are shown as two-colored stripes.

In addition, H2A.Z ROIs are often found at the border of two domains such as YELLOW-BLACK (H2A.Z-only) and YELLOW-RED (H2A.Z+Pol II) (**Figure 4.19**). Transition profiles from one domain to the other clearly show that H2A.Z is enriched in the YELLOW chromatin, as well as at the boundary between YELLOW and BLACK domains (**Figure 4.20**). As a control, profiles centered to the RED-BLUE boundary indicate that H2A.Z is neither enriched at these domains, nor at their boundary. In contrast, Pol II profiles drop in both cases towards the silent domains (**Figure 4.20**).

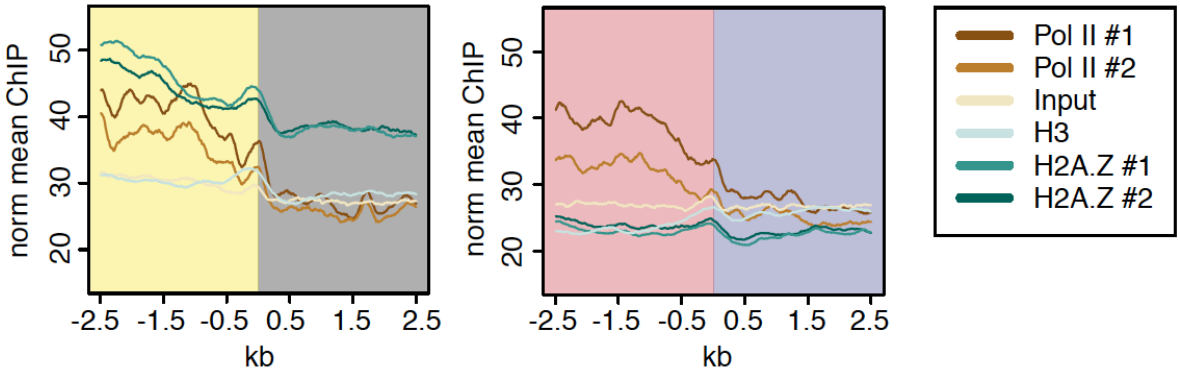


Figure 4.20 H2A.Z is present at the transition between chromatin domains.

Pol II (brown), H2A.Z (green), Input (light brown) and H3 (light green) profiles centered to the boundary of YELLOW-BLACK (left panel) as well as RED-BLUE (right panel) chromatin.

Epigenetic domains of chromatin match topological domains revealed by recent high-resolution chromosome conformation capture data (Sexton et al., 2012; Dixon et al., 2012). These and other reports (see **section 2.2.2.3**) confirmed that insulator-binding proteins are enriched at the boundaries of such domains maintaining the higher order structure (Maeda et al., 2007; Bushey et al., 2009). H2A.Z found at the boundaries of chromatin domains indicates that it co-localizes to insulator-binding sites. To test the association of H2A.Z with insulator proteins, we matched H2A.Z regions with the data of class I and class II insulators obtained in *Drosophila* embryos (Negre et al., 2010). Regions carrying H2A.Z and Pol II almost always overlap with class I insulators, such as BEAF-32, CP190 and CTCF (especially the first two; **Figure 4.21**). The association of H2A.Z-only and Pol II-only with insulator proteins is less than 20%, indicating that both H2A.Z and Pol II are present at these sites. In contrast, there is no (or very little) overlap with the known class II insulator Su(Hw) (**Figure 4.21**).

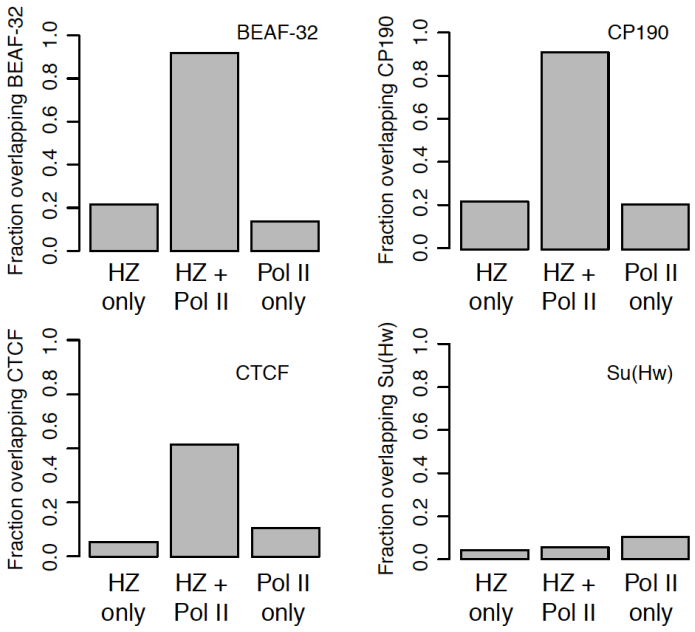


Figure 4.21 H2A.Z and Pol II shared ROIs associate with class I insulator-binding proteins.

Fraction of H2A.Z-only (HZ); H2A.Z+Pol II and Pol II-only ROIs overlapping class I (BEAF-32, CP190 and CTCF) and class II (Su(Hw)) insulators (see legend).

Taken together, our comparative analysis of H2A.Z and Pol II in association with genes that are ubiquitously expressed confirms previous findings. Specifically that 1) ubiquitous genes form clusters, 2) have broad promoter architecture, 3) are present in constitutive epigenetic domains and 4) boundaries of these domains are separated by insulator binding proteins. Our established method, CAST-ChIP confirms these studies by uncovering a novel feature of H2A.Z incorporation with genes that are ubiquitously regulated among highly distinct cell types.

4.5 Discussion

In this chapter, I described new features of cell-type-specific gene regulation, a potential mechanism that distinguishes between ubiquitous and cell-type-specific genes. I extended the analysis of mapping H2A.Z in the fly head in terminally differentiated cells by using CAST-ChIP in three distinct cell types. CAST-ChIP uncovered a novel function of H2A.Z, marking ubiquitous genes that are independent of their developmental fate. Comparison of differentiated cells to developing embryonic cells also showed the cell-type- and stage-invariant feature of H2A.Z.

In early reports, H2A.Z was found to be non-randomly distributed on polytene chromosomes, both in euchromatic and heterochromatic regions (Leach et al., 2000) and to be involved in Polycomb-type heterochromatin formation (Swaminathan et al., 2005). This seemingly contradicts novel genome-wide studies, where H2A.Z was mapped close to the promoter regions of active genes in *Drosophila* (Mavrich et al., 2008; Henikoff et al., 2009; Weber et al., 2010). These studies also showed that H2A.Z containing nucleosomes are well positioned, peaking on the +1 nucleosome downstream of the TSS and gradually decreasing towards the gene body. Similarly, the positioning of H2A.Z nucleosomes is intrinsically present in heterochromatic regions, including DNA transposons (Zhang et al., 2011). This suggests a role of H2A.Z independent from ongoing transcription. Further, H2A.Z is not only present at active genes, but also at genes that are poised for activation.

H2A.Z is present at active genes and absent from inactive genes but its level does not show linear correlation with the gene expression level obtained by microarrays or RNA-seq (Barski et al., 2007; Jin et al., 2009; Weber et al., 2010). Genes with the highest expression usually carry less H2A.Z compared to middle and lower gene expression classes (Weber et al., 2010; and see **Figure 4.2**). This phenomenon is caused by the loss of nucleosomes at very highly expressed genes or in case of robust gene activation upon induction (e.g. *Hsp70*; Petesch et al., 2008 and see **Figure 4.3**). H2A.Z is present at genes with lower expression levels that can get rapidly activated, such as the INO1 and GAL1 genes in yeast (Brickner et al., 2007). H2A.Z enrichment at the gene body associates with lower expression but higher responsiveness upon stress induction in *Arabidopsis* (ColemanDerr et al., 2012). Thus, H2A.Z may play a role in "transcriptional memory" at inducible genes, ensuring their rapid re-activation upon environmental change.

Our analysis with specific cell types from the *Drosophila* head also revealed several sites carrying H2A.Z without Pol II. *Vice versa*, there are several regions with Pol II that lack H2A.Z. In this chapter, I show that the expression of Pol II-bound genes in the absence of H2A.Z is highly variable among tissues and developmental stages of the FlyAtlas (Chintapalli et al., 2007). Therefore, this set of genes is regulated extensively in a cell- and development-specific manner. These specifically-regulated genes are located in chromatin domains marked by activating proteins (RED in Filion et al., 2010), but lack H3K36me3, as well as in a chromatin domain that is in general silent but shows highly diverse gene expression in developmental stages and tissues (BLACK in Filion et al., 2010). In contrast, H2A.Z is present at genes that share features of H3K36me3 (an elongation mark) and associated proteins such as MRG15 (Zhang et al., 2006; YELLOW chromatin in Filion et al., 2010). These genes include for example ribosomal genes with universal cellular function (*data not shown*). Therefore, I suggest that H2A.Z could be a chromatin mark for ubiquitous, such as housekeeping genes, and also broadly inducible genes such as heat shock genes.

I found H2A.Z is present at the borders of chromatin domains and associates with class I insulator-binding proteins such as CP190, BEAF-32 and CTCF (see **Figure 4.22-4.23**). H2A.Z nucleosomes are enriched at CTCF sites also in mammals (Jin et al., 2009). Insulator sites bound by class I insulator proteins demarcate gene boundaries and are enriched between differentially expressed promoters (Negre et al., 2010). These insulators are able to restrict the spread of Polycomb-type heterochromatin (Schwartz et al., 2012) and are involved in the maintenance of Polycomb domains marked by H3K27me3 (VanBortle et al., 2012). In embryonic stem cells (ESCs), the dually modified (acetylated and ubiquitinated) H2A.Z was also found at bivalent domains carrying both H3K4me3 and H3K27me3 (Ku et al., 2012). However, *Drosophila* chromatin lacks bivalent domains (Schuettengruber et al., 2009), therefore, gene regulation across *Drosophila* development might be different compared to ESCs cells. The role of H2A.Z at bivalent domains may be to mark genes that are poised for getting activated. In that aspect, H2A.Z-only-bound genes, lacking Pol II enrichment in the Drosophila head ChIP, might stay in a similar poised state (**Figure 4.5** and **4.6**).

Genome-wide profiles of H2A.Z, especially in *Drosophila*, have to date been generated from cell cultures and embryos. I sought to identify H2A.Z incorporated regions in differentiated cells within the intact *Drosophila* head. I applied the cell-type

specific CAST-ChIP method (described in **chapter 3**) and found minor differences of H2A.Z among cell types (see **Figure 4.8-4.12**), but marked differences in Pol II association between cell types, reflecting distinct gene activity between cell types. To confirm the robustness of CAST-ChIP, I used antibodies against the endogenous protein on two distinct developmental stages. Surprisingly, H2A.Z was present at the same genomic sites in both embryos and the adult head. Taken together, our data indicate that *Drosophila* genes marked by the histone variant H2A.Z share a common chromatin feature in all cells, specifically the enrichment with chromatin domains that are ubiquitously regulated.

5 TRAP – translating ribosome affinity purification for cell-type-specific translatome profiling

This project was performed in a bioinformatics collaboration with Petra Schwalie (EBI, Hinxton, UK).

5.1 Summary

Cell-type-specific chromatin profiling reveals which enhancers and genes are active in a particular cell population (see **chapter 3** and Bonn et al., 2012). However, downstream regulation steps including mRNA processing or ribosome binding also play important roles in determining cell-type-specific gene expression patterns.

Here, I developed and applied an approach that profiles cell-type-specific mRNAs. The Translating Ribosome Affinity Purification method (TRAP; Heiman et al., 2008 and Doyle et al., 2008) was developed in mice and has recently been combined with the UAS/Gal4 system in *Drosophila* (Thomas et al., 2012). TRAP is a powerful method to find cell-type-specific differences at the level of the translatome (Dougherty et al., 2010). In parallel to now published efforts, I developed and implemented the method for the fly and compared distinct head cell types and identified cell-type-specific transcript classes with neuronal (e.g. receptor-, neuropeptide- or hormone activity) or glial function (e.g. transporter activity). Neuronal TRAP genes are over-represented in the brain, larval CNS and thoracicoabdominal ganglion (Chintapalli et al., 2007). Using cell-type-to-cell-type comparisons (e.g. neurons vs. glia), instead of a given cell population to the total (e.g. neurons vs. head), the differences could be identified with greater resolution. TRAP uncovered more neuronal genes compared to our neuronal RNA polymerase II data (CAST-ChIP). Thus, TRAP data confirm the importance of post-transcriptional regulation in defining cell identity.

In summary, TRAP is one of the best methods to reveal differential "omics" data among distinct cell types by profiling ribosome-bound mRNAs. TRAP is a promising tool to reveal cell-type-specific transcriptional and translational changes in a perturbed environment.

5.2 Introduction

Gene expression is regulated via complex regulatory networks in a spatio-temporal manner (see **section 2.13** and **2.2**). At the transcriptional level, gene expression is determined by enhancer and promoter activity and by chromatin modifications, which modulate the environment of genes and organize the genome into large domains (see **section 2.2.2**). The binding of RNA polymerase II to enhancers and promoters defines the local activity of these single sites (see **section 2.2.1** and Bonn et al., 2012). On the other hand, RNA polymerase II binding does not necessarily lead to active elongating polymerase; Pol II often stays in a paused state producing short RNAs (Nechaev et al., 2010). The classical view of gene expression suggests a forward flow from genes towards proteins. However, recent studies reveal uncoupling of transcription from translation (see **section 2.2.3** and Tebaldi et al., 2012). In addition, mild stress conditions do not affect the total transcriptome but affect the ribosome-associated translatome in yeast (Halbeisen et al., 2009). Therefore, each step of gene expression gives additional complexity to the network. To understand gene regulatory networks, profiling protein-DNA interactions by ChIP is not sufficient to uncover all aspects of gene expression. Indeed, in cell types such as neurons, translational control plays a role in synaptic plasticity (CostaMattioli et al., 2009). A complementary approach to defining gene activity by protein-DNA interaction (e.g. Pol II binding) is thus to map mRNA (transcriptome) or ribosome-bound mRNAs (translatome).

Here, I applied a cell-type-specific RNA profiling method that relies on the affinity purification of ribosome-associated mRNA. TRAP (Translating Ribosome Affinity Purification) has been developed in mice to map the translatome of several neuronal cell populations (Heiman et al., 2008; Doyle et al., 2008 and Dougherty et al., 2010). TRAP identifies cell-type-specific, ribosome-bound transcripts that are enriched over the total RNA pool of the given organ (e.g. Purkinje cells vs. cerebellum), having a high IP/total tissue ratio. These transcripts are basically both cell-type-specific and highly translated messenger RNA molecules. To obtain a better resolution and more differences among cell types, instead of comparing the cell-type-specific IP to the total tissue, several distinct cell populations have to be compared to each other, especially in cases where the cell type of interest is highly abundant in the tissue (Dougherty et al., 2010). In order to distinguish between cell-type-specific and ubiquitous genes, mRNA

profiles from distinct cell types have to be compared. Therefore, I chose three main terminally differentiated cell types of the *Drosophila* head with distinct developmental origin: ectodermal neurons and glia cells as well as the mesodermal fat body (see **section 2.1.2**). A tripartite comparison of larger cell populations with diverse function probably uncovers genes required to fulfill these functions (e.g. neurons: neurotransmitter receptors; fat body: metabolic enzymes).

I adapted TRAP to *Drosophila* in combination with the UAS/Gal4 system, as did a competing group (Thomas et al., 2012). In the competitor's publication, TRAP identified hundreds of transcripts specific to *Drosophila* neurons compared to the head. TRAP was able to enrich mRNA from only about 200 neurosecretory cells by detecting insulin-like peptide (ILP)-encoding transcripts that are specific to these cells (Thomas et al., 2012). However, the authors´ analysis was limited to IP vs. total tissue comparisons, which as described above does not reveal all the mRNA differences among distinct cell types.

In this chapter, I use TRAP translational profiling of distinct cell types of the *Drosophila* head to purify cell-type-specific ribosome-associated mRNAs. Gene expression profiling of cell types at the level of mRNAs gives a higher dynamic range than chromatin-based approach, and is therefore a good tool to identify which genes are expressed in a given cell population. Furthermore, TRAP maps the cell-type-specific translatome, making it a suitable tool to study the dynamic change upon even mild environmental perturbations.

5.3 Methods

5.3.1 Experimental procedures

Experiments in this chapter were performed by Tamás Schauer.

5.3.1.1 Fly Stocks

Flies were kept on standard media at 25 °C. 1-3 days old flies were collected and frozen in liquid nitrogen at the same time of the day. Frozen flies were stored at -80 °C until used for ribosome affinity purification. UAS-GFP-L10A flies were generated in the 2202U background by cloning the PCR amplified *RpL10Ab* (*CG7283*) cDNA fragment into the pUAST-GFP vector. The transgene was expressed under the control of *elav*-

GAL4 (Bloomington stock no. 458), *repo*-GAL4 (Sepp et al., 2001) or *take-out*-GAL4 (Dauwalder et al., 2002).

5.3.1.2 Western blot and immunohistochemistry

Proteins were extracted from fly heads ground in 2x Laemmli buffer (4% SDS, 20% glycerol, 200 mM DTT, 120 mM Tris pH 6.8, 0.0025% w/v bromophenol blue) and boiled for 5 minutes. The tissue debris was removed by centrifugation with maximum speed for 10 minutes at room temperature. Three head equivalents of extract was loaded into each lane. Proteins were electroblotted onto nitrocellulose membrane and incubated in TBST (10 mM Tris HCl pH 6.8, 150 mM NaCl, 0.05% Tween-20) with 5% milk powder. The following primary antibodies were used in TBST+5% milk: anti-GFP (TP401, 1:5000), anti-Tubulin (Sigma T9026, 1:10000) and anti-RPL10A antibody (Abcam ab55544, 1:5000). The membrane was extensively washed with TBST at least four times for 5 minutes. The following secondary antibodies were used in 1:10000 TBST+5% milk: anti-rabbit-HRP (Bio-Rad 172-1019) and anti-mouse-HRP (Bio-Rad 170-6516). The membrane was developed using Immobilon Western Chemiluminescent HRP Substrate (Millipore) and Fuji medical X-ray film (Super RX).

Brain staining was performed as described under **3.3.1.2.**

5.3.1.3 Ribosome affinity purification

Ribosome affinity purification was adapted from Heiman et al., 2008. About ~1000 fly heads were homogenized in ~1 ml ice-cold ribosome extraction buffer (REB: freshly prepared from stock; 10 mM HEPES, 150 mM KCl, 5 mM $MgCl_2$, protease inhibitor cocktail tablet (PIC, Roche), 0.5 mM DTT, 100 µg/ml cycloheximide (CHX), 100 U/ml RNasIn (Promega)). Nuclei were collected by centrifugation at 2000G for 10 minutes at 4°C. The supernatant was cleared with NP-40 (final concentration 1%) and DHPC (1,2-Diheptanoyl-sn-Glycero-3-Phosphocholine; final concentration 30 mM; Avanti Polar Lipids) for 5 minutes on ice with gentle pipetting (the volume was set back to ~1 ml if needed). The rest of the debris was pelleted with 13000G for 10 minutes at 4°C. The supernatant (~1ml) was split into four portions of 250 µl and 10% (25 µl) of each tube was used as Input. The samples were diluted to 500 µl with REB containing 1% NP-40. 2 µl anti-GFP antibody (goat, lab stock) was added to each sample and incubated for 10 minutes at 4°C. The samples were added to Sepharose protein G beads (slurry volume: 25 µl/IP, previously equilibrated in REB + 1% NP-40 [GE Healthcare]) and further

incubated for 50 minutes at 4°C. The beads were washed 3x quickly with ribosome wash buffer (RWB: freshly prepared from stock; 10 mM HEPES, 350 mM KCl, 5 mM $MgCl_2$, 1% NP-40, protease inhibitor cocktail tablet (PIC, Roche), 0.5 mM DTT, 100 µg/ml CHX, 100 U/ml RNasin (Promega)). Total RNA was isolated from the beads using RNeasy Micro Kit (Qiagen) according to the manufacturer´s instructions. Beads from the four parallel IPs were pooled into one purification column. Purified RNA was eluted with 15 µl RNase-free water. Three replicates of immunoprecipitated total RNA was pooled into one (~1 µg required) for high-throughput sequencing.

5.3.1.4 Quantitative RT-PCR

100 ng of TRAPed and purified total RNA was used for reverse transcription (RT) using Superscript III reverse transcriptase and random primers (Invitrogen). qPCR was performed on an AB Fast real-time PCR system using fast SYBR Green Master Mix (Applied Biosystems). Ct values were normalized to the control gene *Rp49* (*RpL32*). ΔCt values were calculated as IP vs. Input and IP vs. IP comparisons. The following primer pairs were used in this study:

dilp2 (AGCAAGCCTTTGTCCTTCATCTC - ACACCATACTCAGCACCTCGTTG);
elav (CAACCGAAGTAACCATAACTGGA - TCCTTGCTCTCTGCTTCGAT);
Hsp70 (ATCGCCAGCGAATAACCTC - CCTGCTTCACATTGAAGACGTA);
igl (GTCCACTTTCCGTGGTCATT - TAAGCTCGGGATCGGTTAAA);
Obp99b (TTCGATGTCCACAAGATCCA - TAGACCTTGACGCTGTGCTG);
repo (ATCCCAATGGCATCAAGAAG - ACACGGGATTCGCTCAGAT);
Rp49 (CGGATCGATATGCTAAGCTGT - GCGCTTGTTCGATCCGTA);
18S rRNA (GACCAATTGGAGGGCAAGT - TACGCTAGTGGAGCTGGAATTA);
28S rRNA (AACGAGATTCCTACTGTCCCTATC - AATTATTCCAAGCCCGTTCC).

5.3.1.5 RNA sequencing

Approximately 1 µg of Input or TRAPed RNA was used for RNA sequencing. Sequencing library was prepared using Illumina polyA-mRNA library preparation methods with paired-end option. 72bp reads were obtained from the sequencer.

5.3.2 Data analysis

Data analysis was performed by Petra Schwalie (European Bioinformatics Institute, Hinxton, UK) and by Tamás Schauer.

5.3.2.1 Sequence alignment and differential expression analysis

Paired-end RNA-seq reads obtained from neuronal and glial TRAP IP and Input were aligned with Tophat v1.0.14 (Trapnell et al., 2009) to Ensembl transcript annotations. Gene annotation-based expression values were estimated using Cufflinks v0.9.0 (Trapnell et al., 2010). For further analysis the gene-based FPKM (fragment per kilobase per million mapped reads) values were used. Due to the lack of replicates, stringent fold change (FC) cutoffs were made to define differential expression in neuronal-glial TRAP IP vs. IP and IP vs. Input comparisons. Genes with greater than 2 fold change FPKM between Inputs were excluded. The data were subset into categories according to the fold change between neuron vs. glia TRAP IPs: 1) FPKM values greater than 8 FC were considered as very specific neuronal, 2) less than 8 FC and greater than 4 FC as specific neuronal, 3) less than 4 FC and greater than 2 FC as depleted in glia, 4) less than 2 FC and greater than 0.5 FC as invariant and 5) less than 0.5 as specific glial genes. Downstream analysis was performed based on these subsets.

Scatterplots and correlation coefficients (Pearson and Spearman) were generated and calculated using R (R Development Core Team, 2010).

5.3.2.2 FlyAtlas and Gene Ontology analysis using FlyMine

Neuronal, glial and invariant gene lists (according to FC cutoffs, see section 5.3.2.1) were compared to FlyAtlas (Chintapalli et al., 2007) using FlyMine (Lyne et al., 2007). Up- and down-regulated genes in FlyAtlas tissues were plotted as percentage of total gene number in the given category obtained by TRAP.

Gene ontology (GO) term enrichment analysis was performed using FlyMine with Holm-Bonferroni test correction and a maximal p-value of 0.05 (Lyne et al., 2007).

5.3.2.3 Comparison to CAST-ChIP Pol II data

Neuronal (FC>4), glial (0.5>FC) and invariant (2>FC>0.5) TRAP datasets were compared to neuronal, glial and invariant Pol II CAST-ChIP data (see **3.4.1.2**). Note that invariant Pol II peaks contain all head Pol II regions indifferent between neurons and glia, including those that do not carry Pol II in neurons and glia but in other head tissues. Venn diagrams in this section were generated using the online BioInfoRx tool (http://apps.bioinforx.com/bxaf6/tools/app_overlap.php).

5.4 Results

5.4.1 Establishing TRAP in *Drosophila*

5.4.1.1 Generating GFP-L10A flies

Translating ribosome affinity purification (TRAP) was originally developed in mice by expressing a tagged version of the ribosomal large subunit protein L10A in specific cells within the central nervous system. This was established using bacterial artificial chromosome (BAC) transgenic mice (Heiman et al., 2008; Doyle et al., 2008). In *Drosophila*, one of the most commonly used expression system is the UAS/GAL4 system. Therefore, I decided to express the tagged fly homologue of L10A under the control of well-characterized GAL4 drivers in order to develop the TRAP systems to profile distinct *Drosophila* cell types. ClustalW2 alignment of the mouse ribosomal protein L10A (GenBank: AAH83346.1) and fly RpL10Ab (*CG7283*) show high homology (Larkin et al., 2007; **Figure 5.1**).

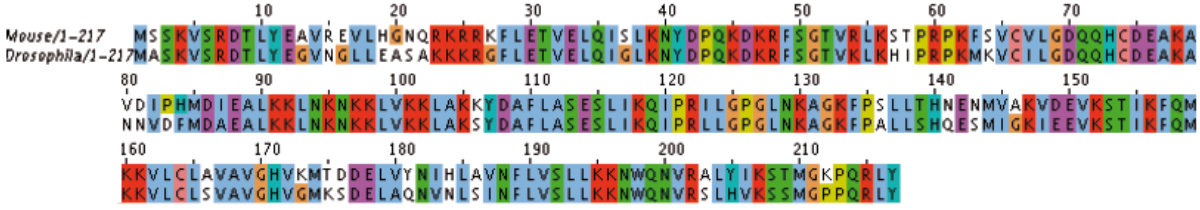


Figure 5.1 Alignment of the ribosomal protein L10A.
ClustalW2 alignment of the 217 amino acids long RpL10A protein from mouse and *Drosophila*.

In order to generate transgenic flies, I cloned *RpL10Ab* into the pUAST-GFP vector, which allows expression in a GAL4-dependent manner. I obtained 11 independent insertion lines, from which I chose one second chromosomal line for all later experiments. To test the expression of GFP-L10A, I generated flies carrying the GFP-L10A insertion together with a GAL4 under the control of cell-type-specific promoters, such as neuronal *elav*-Gal4, glial *repo*-Gal4 and fat body *take-out*-Gal4. Since animals carrying a homozygous double-copy of *repo*-Gal4 are not healthy, I used *repo*-Gal4 in heterozygous form, whereas *elav*-Gal4 and *take-out*-Gal4 give healthy animals even with

two copies of the transgene. Western blot analysis on fly heads showed that the expression levels in the case of *elav*-Gal4 and *take-out*-Gal4 is very similar and somewhat less in the case of *repo*-Gal4 (**Figure 5.2**). This can be explained with the smaller number of glia cells compared to neurons and fat body cells within the adult fly head (see **section 2.1.2**).

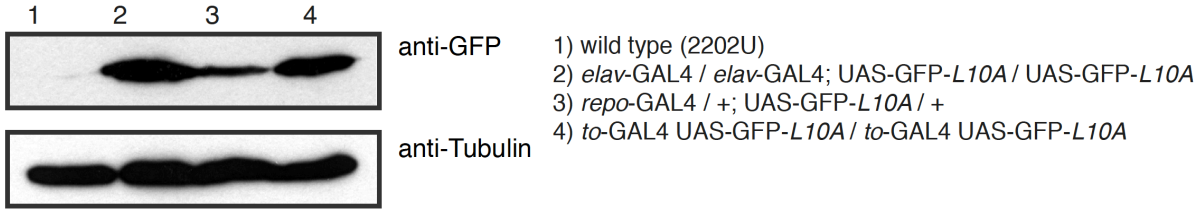


Figure 5.2 Expression levels of GFP-L10A in various cell types in the fly head.
Western blot using anti-GFP and anti-Tubulin (control) antibodies on wild type (2202U) and flies expressing GFP-L10A in neurons (elav-Gal4), glia cells (*repo*-Gal4) and fat body (to-Gal4) respectively.

To test whether GFP-L10A is expressed correctly in the cell type of interest, I analyzed its expression pattern by GFP fluorescence and immunostaining (**Figure 5.3**). Fluorescence of GFP-L10A in neurons (*elav*-Gal4), glia cells (*repo*-Gal4) and fat body (*take-out*-Gal4) showed clearly distinct patterns localized to the brain (neurons [*elav*]) and to adipose tissues (fat body [*take-out*]), respectively (**Figure 5.3A**). GFP-L10A expressed in glia was relatively weak in the whole head, but definitely stronger than auto-fluorescence in the wild type head. The weak signal of *repo*-marked cells was probably due to the smaller number of glia cells compared to neurons in the brain. Co-staining experiments of dissected and fixed brains with an anti-GFP antibody against GFP-L10A expressed in neurons (*elav*) showed a clear overlap with the neuronal marker ELAV and expressed in glia (*repo*) with glial marker (REPO), respectively (**Figure 5.3B**). *Takeout*-driven GFP-L10A was found outside of the brain in surrounding tissues that morphologically and anatomically share characteristics with the fat body. GFP-L10A without a driver did not show GFP staining with the same imaging settings. Zoom-in images of the same staining confirmed the overall pictures. As shown previously, *elav*-Gal4 driven GFP-L10A preferentially localizes to the cytoplasm of cells with ELAV positive nuclei (Thomas et al., 2012; **Figure 5.3C**). The glial nuclear marker, REPO, shows almost complete overlap with the *repo*-Gal4 controlled GFP-L10A (without distinction of nucleus and cytoplasm), which could be explained by the different cell

morphology of glia cells having larger nuclei compared to neurons. Thus, these data show that the TRAP reporter (i.e. GFP-L10A) is expressed in the cell type of interest using the UAS/Gal4 system.

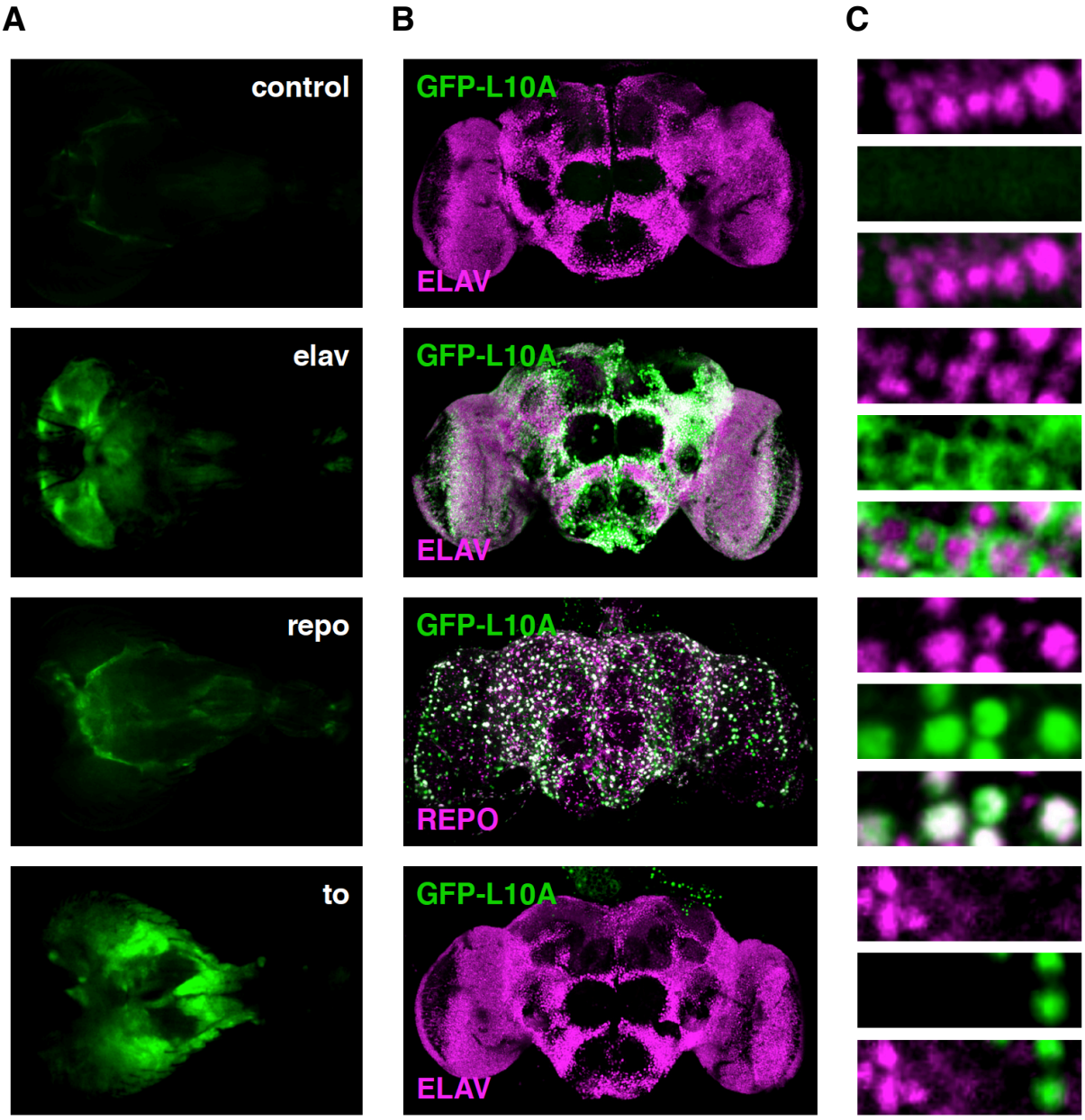


Figure 5.3 Expression pattern of GFP-L10A in various cell types of the fly head.

(A) GFP fluorescence of wild type control (first row), neuronal (*elav*; second row), glial (*repo*; third row) and fat body (*take-out* [*to*]; fourth row) GFP-L10A in whole heads. **(B)** Fluorescence immunostaining of GFP-L10A with an anti-GFP (green) antibody in fly brains carrying GFP-L10A without driver (first row) co-stained with anti-ELAV antibody (magenta); expressed in neurons with *elav*-Gal4 (second row) co-stained with ELAV (magenta), in glia with *repo*-Gal4 (third row) co-stained with REPO (magenta) and in fat body with *take-out*-Gal4 (fourth row) co-stained with an anti-ELAV antibody (magenta). **(C)** Zoom in image of staining described in (B) with separate channels for ELAV or REPO (magenta, top), GFP (green; middle) and merged image (bottom) in each case.

5.4.1.2 Optimizing TRAP

To test whether GFP-L10A can be pulled-down from *Drosophila* head extracts, I expressed GFP-L10A with the ubiquitous tubulin-Gal4 driver. I adapted the published protocol from Heiman *et al.* for *Drosophila* heads (see Methods; Heiman et al., 2008). First, I carried out western blot analysis on the Input and on the TRAPed samples with different amounts of protein G beads using anti-GFP antibody and normal rabbit serum as control (mock IP). The western blot was developed using an anti-RpL10A antibody (see Methods; **Figure 5.4A**). I could detect a 55 kDa band, corresponding to GFP-L10A, in the Input and the IP samples. In contrast, there was no clear band in the mock IP, as expected. Further, increased amount of beads pulled down increased amounts of GFP-L10A (**Figure 5A**; 1x vs. 2x beads).

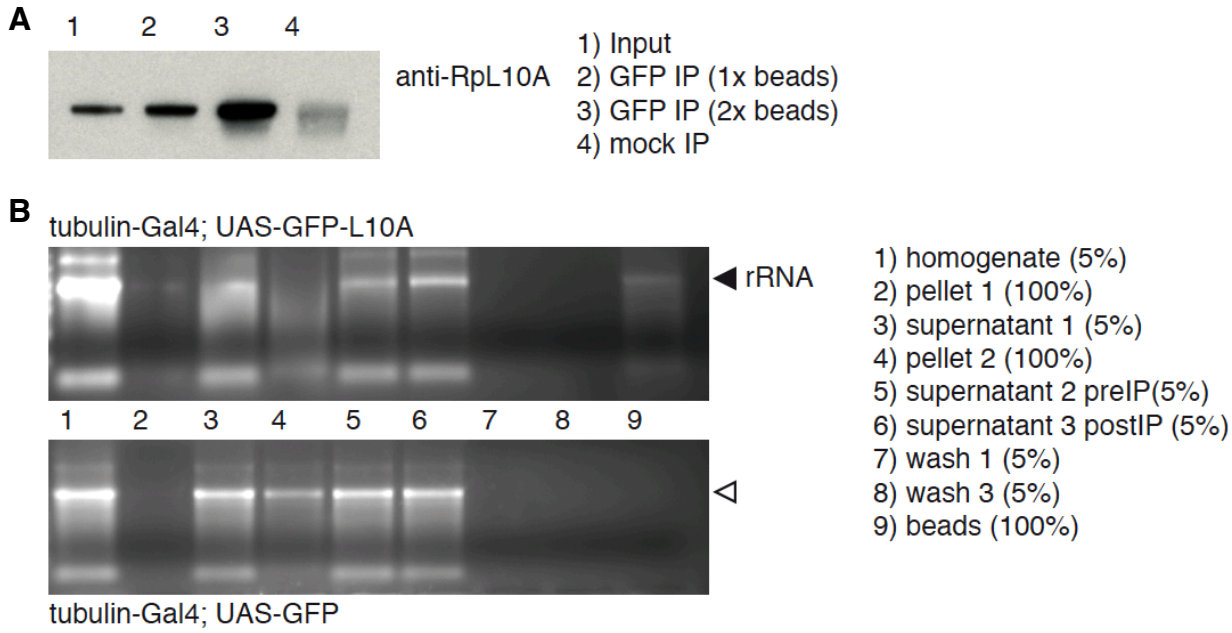


Figure 5.4 Affinity purification of GFP-L10A.

(A) Western blot analysis of GFP-L10A present in Input and ImmunoPrecipitation (IP) with anti-GFP antibody (lab stock; 1x and 2x amount of beads). In contrast, no GFP-L10A is present in mock IPs with normal rabbit serum. GFP-L10A was detected with a commercial anti-RpL10A antibody. **(B)** Agarose gel analysis of purified RNA taken from each step of the purification protocol in samples expressing GFP-L10A or GFP only (with tubulin-Gal4 driver). The black triangle indicates the rRNA band, whereas an empty triangle shows the absence of rRNAs.

Next, I analyzed the amount of purified total RNA obtained by GFP-L10A IPs compared to IPs of GFP only (**Figure 5.4B**). I took samples after each step of the experimental procedure and ran it on an agarose gel (see **Figure 5.4B** legend and Methods). A clear band of ribosomal RNAs and a smear of mRNAs are visible throughout the experiment (except in the washes). RNA bound to the beads is detected only in the case of the GFP-

L10A pull-down and not of GFP. Thus, I am confident that GFP-L10A can be specifically immunoprecipitated and that it co-purifies rRNAs and mRNAs.

To quantify mRNA and rRNA levels pulled-down by TRAP from the whole head, I performed RT-qPCR and determined the percentage of rRNA and a few head specific mRNAs compared to the Input (total unbound). Affinity purification of GFP-L10A co-purified the 28S ribosomal RNA and candidate mRNA transcripts I tested (**Figure 5.5**). Input percentage of the *Obp99b* (odorant binding protein 99b) mRNA, for example, was as high as for the ribosomal RNA, indicating that this transcript is mainly bound to ribosomes. *Elav* and *repo* showed smaller enrichment than Obp99b, suggesting that these transcripts are less associated with ribosomes. This highlights the feature of TRAP enriching transcripts that are likely to be undergoing translation. In contrast, immunoprecipitating GFP that was expressed in the whole head, did not co-purify rRNAs and mRNAs (**Figure 5.5**). Taken together, GFP-L10A specifically co-immunoprecipitates rRNA and head specific mRNA. This makes *Drosophila* TRAP as useful tool to investigate the cell-type-specific translatome.

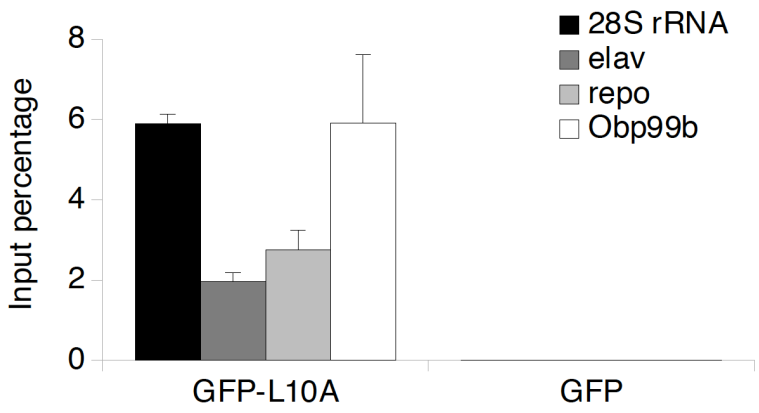


Figure 5.5 Detecting specific TRAPed RNA in RT-qPCR.
Input percentage (bound/total) of 28S rRNA as well as *elav*, *repo* and *Obp99b* mRNAs. The transcripts were co-purified with GFP-L10A and absent in GFP control. GFP-L10A or GFP were expressed in the whole head using *tubulin*-Gal4.

To optimize TRAP for cell-type-specific pull-downs, instead of comparing IP to Input (as Input percentage), I wanted to find the best conditions of comparing neuron- and glia-specific IPs. I previously found an optimal Sepharose bead:antibody ratio (see **section 3.4.1.1**). I therefore titrated the amount of the lysate added to the beads (**Figure 5.6**). Lysates were always prepared from about 1000 fly heads and 12.5% (125 head equivalent), 25% (250 head equivalent) and 50% (500 head equivalent) was used for each IP (see Methods). In all conditions, I obtained enrichments of the neuronal gene *igloo* (*igl*) in neuron-TRAP (**Figure 5.6 positive values**) and the glial gene *repo* in glia-

TRAP (**Figure 5.6 negative values**). I found the best enrichment in the case where I used 12.5% (125 head equivalent) of the lysate. However, the amount of purified transcripts was relatively low (Ct values higher than 30). Therefore, I continued with IPs using 25% (250 head equivalent) of the lysate. Ribosomal RNA (28S and 18S) and a control mRNA (*Hsp70*) did not show a difference between neuron- and glia-TRAP. Furthermore, the amount of transcripts was similar among head Inputs, excluding bias coming from the different transgenes. In summary, I found the best TRAP conditions to enrich transcripts such as *igloo* (a gene carrying a neuronal RNA pol II peak; using CAST-ChIP, see **Figure 3.16**) and *repo*, a known glial marker (see **chapter 2.1.2**).

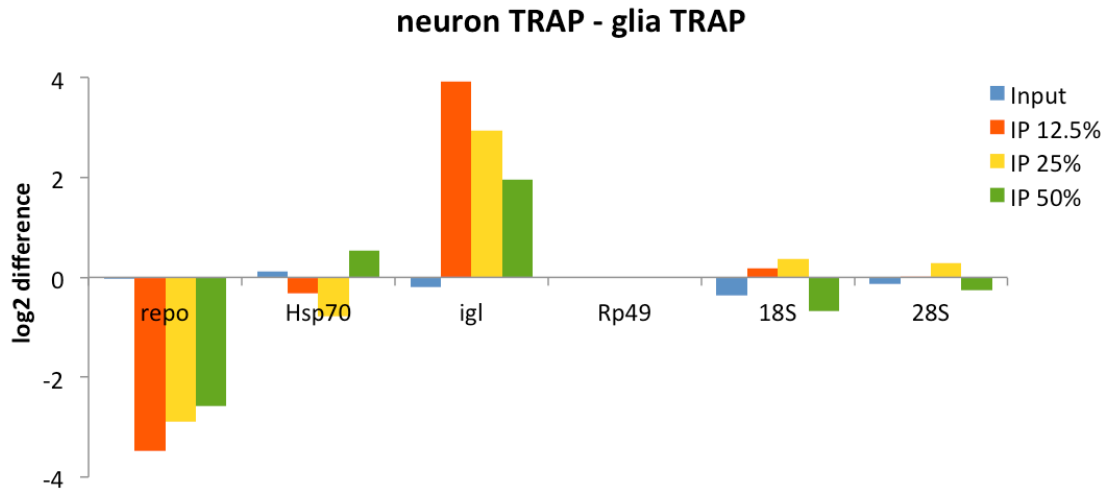


Figure 5.6 Optimizing TRAP by comparing neuron vs. glia IPs.

Log2 difference of neuronal and glial TRAP is shown in Input (blue) and IPs using 12.5% (orange), 25% (yellow) and 50% (green) of lysates prepared from ~1000 fly heads. *Igloo* (igl) shows neuronal enrichment (positive values), whereas *repo* glial (negative values). Control targets such as *Hsp70*, 18S and 28S rRNA are identical between cell types. Ct values were normalized to *Rp49*.

5.4.1.3 Establishing TRAP in three distinct cell types

To test whether TRAP results are reproducible and applicable to different cell types, I expressed GFP-L10A in neurons (*elav*-Gal4), glia (*repo*-Gal4) and fat body (*take-out*-Gal4), respectively (**Figure 5.2 and 5.3**), and performed TRAP using several biological replicates (n=6; **Figure 5.7**). First, I compared cell-type-specific TRAPs to the corresponding Input samples, which represent the total head transcripts having the same genetic background (**Figure 5.7 left column**). I found high enrichment over Input of an insulin-like peptide transcript (*Dilp2*, **Figure 5.7 red**) which is expressed in neurons of the *pars intercerebralis* (Rulifson et al., 2002). The glial transcript *repo* was depleted and the odorant binding protein gene (*Obp99b*) was invariant in the neuronal

TRAP compared to total. Glia-TRAP shows the opposite results, *repo* is over-represented, *Dilp2* and *Obp99b* depleted in the glial ribosome-bound fraction over the head Input (**Figure 5.7 dark blue**). TRAPing ribosome-bound RNAs from the fat body revealed the enrichment of *Obp99b* compared to the total (**Figure 5.7 green**).

Comparing highly abundant cell types (i.e. neurons) to Input (i.e. total head) makes it difficult to obtain high enrichment values using TRAP (see **section 2.3.3; 5.2** and Dougherty et al., 2010). Therefore, I performed pair-wise comparisons of the three cell types (neurons *vs*. glia, glia *vs*. fat body and neurons *vs*. fat body) using the same TRAP qPCR data (**Figure 5.7**). As expected from previous studies (Dougherty et al., 2010), the difference in specific transcript levels among cell types (TRAP-TRAP comparison, **Figure 5.7 right column**) was greater than between cell type and Input (TRAP-Input comparison; **Figure 5.7 left column**, note the different scale). For example, *Dilp2* was 64 fold (log2=6) enriched in neurons compared to glia (**Figure 5.7 purple**), but less than ~16 fold (log2=4) enriched compared to Input (**Figure 5.7 red**). The neurons vs. glia comparison also revealed that the ribosome-bound fraction of *Obp99b* is higher in neurons compared to glia (**Figure 5.7 purple**), which was not clear from the TRAP-Input comparison (**Figure 5.7 red)**. As expected, *repo* showed high enrichment in glia compared to neurons and also to the fat body. *Obp99b* is highly abundant in the fat body compared to glia (**Figure 5.7 light blue**), whereas *Dilp2* was invariant among glia and fat body. The last comparison (neuron-TRAP vs. fat body-TRAP) strengthens these results: *Dilp2* is very specific to neurons, *Obp99b* is enriched in fat body and *repo* is invariant between neurons and fat body (**Figure 5.7 orange**).

Taken together, tripartite comparison of three distinct cell types using TRAP gives higher resolution of specificity than comparing parts to the whole or only two cell types to each other. Careful qPCR analysis of TRAP revealed that 1) *Dilp2* is very specific to neurons and depleted from glia and fat body, 2) *repo* is very specific to glia and depleted from neurons and fat body and 3) *Obp99b* is preferentially bound to ribosomes in the fat body, to some extent in neurons and is depleted from glia.

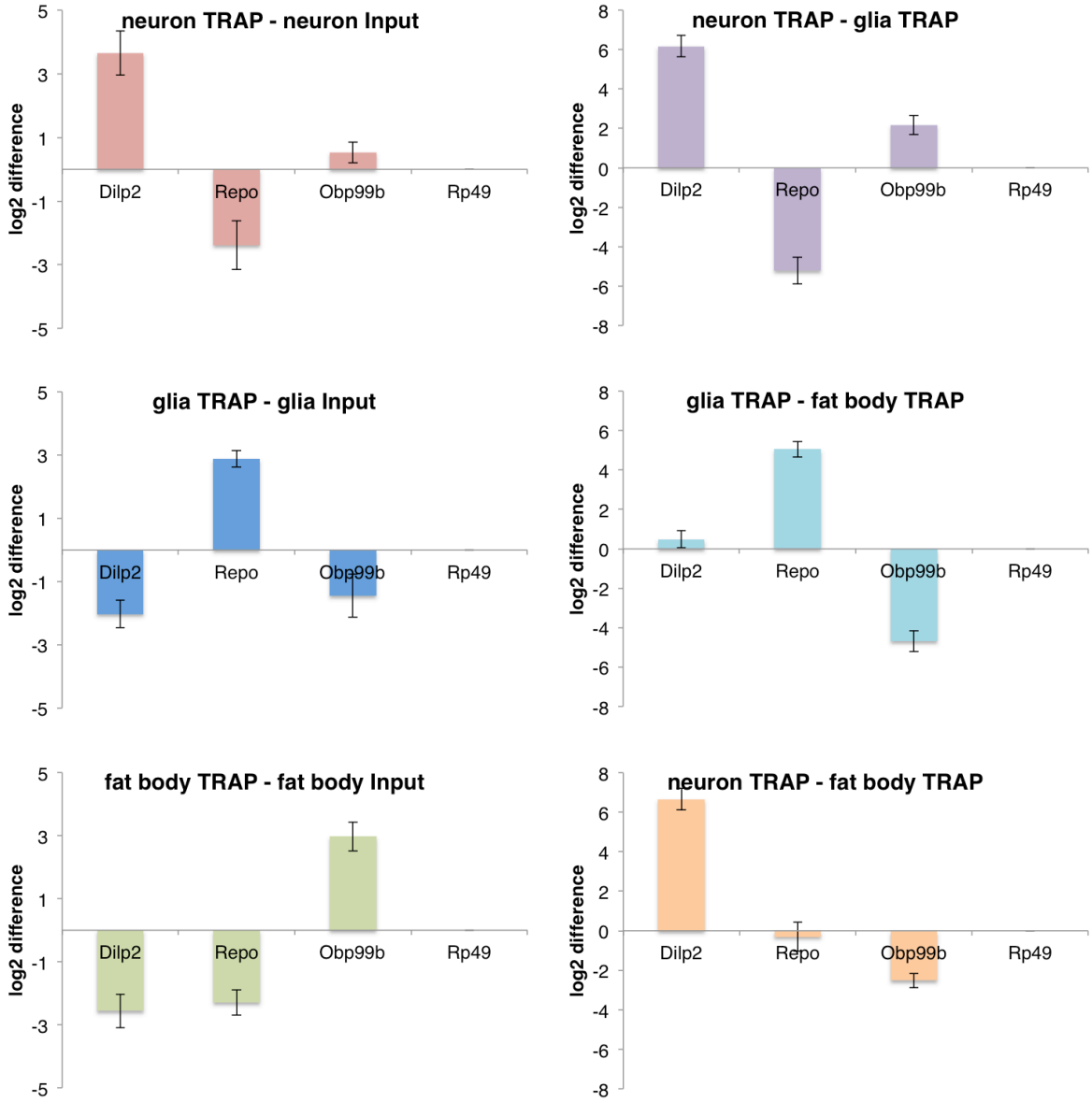


Figure 5.7 Tripartite comparison of TRAP in neurons, glia and fat body.

The left column shows the log2 difference between the cell-type-specific TRAPs and the corresponding total head Inputs in neurons (top, red), in glia (middle, dark blue) and fat body (bottom, green) using RT-qPCR analysis for *Dilp2*, *Repo* and *Obp99b* normalized to *Rp49*. The right column shows the log2 difference of the same transcripts between the same cell types in a pair-wise comparison (neuron-glia: top, purple; glia-fat body: middle, light blue; neuron-fat-body: bottom, orange). Error bars represent the standard deviation, with an n=6. Note the different scale in the two columns.

5.4.2 Genome-wide TRAP profiling

5.4.2.1 Analyzing TRAP by RNA-Seq

In order to profile cell-type-specific ribosome-bound mRNA in a genome-wide manner, I performed RNA-Seq experiments on RNAs obtained by TRAP. The required quantity for standard Illumina polyA protocols is at least 1 µg total RNA (optimal 5-10 µg). The amount of purified total RNA from one TRAP experiment (from 1000 heads in 4 IP replicates) was in the range of 250-500 ng. To reach the required lower limit (1 µg) for the standard sequencing protocol, I therefore pooled several TRAP replicates. The samples could have been amplified before sequencing libraries are made; however these types of pre-amplification are notorious for introducing bias (Tariq et al., 2011). To test the sequencing with relatively low amounts of RNA, one replicate of neuronal and glial TRAP was sequenced. As a control, the corresponding Inputs were also sequenced. These experiments I consider as trials to see genome-wide trends between two cell types. For future experiments, biological replicates and further optimization might be required.

I performed the sequence alignment using Tophat and defined the gene-based normalized read counts (FPKM - fragment per kilobase gene per million mapped fragments) using Cufflinks (Trapnell et al., 2009; Trapnell et al., 2010 and see Methods). The lack of replicates led me to analyze the data with stringent fold change (FC) cutoffs due to the unreliability of DESeq without replicates (Anders et al., 2010). First, I filtered out genes that are different in Inputs (FC>2 and 0.5>FC) to exclude differences derived from the two lines (*elav*-Gal4 *vs*. *repo*-Gal4). I also excluded genes with zero FPKM values since in that case no FC can be determined. Therefore, after filtering, I used 8019 genes for further analysis.

As shown previously (**Figure 5.7**), the best procedure to find differences in TRAP data is to compare cell types to each other and also the cell types to total Input. First, I determined the fold change FPKM between neuron TRAP and glia TRAP. Genes with higher than two FC are potential neuronal and less than 0.5 potential glial gene candidates. To dissect what cutoff gives specific or invariant subsets, I split the data into five categories according to the FC FPKM (**Figure 5.8A**). To characterize these FC categories, I tested the relationship of cell-type-specific TRAP and the corresponding

Inputs within these groups (**Figure 5.8B**). I averaged the fold change FPKM of IP vs. Input (e.g. neuron TRAP/neuron Input) in each category derived from IP vs. IP comparison (neuron TRAP/glia TRAP). Genes with a fold change higher than four (FC>4) are enriched and higher than eight (FC>8) are highly enriched in neuron TRAP and depleted in glia TRAP, compared to the corresponding Input (**Figure 5.8B**). In contrast, genes with a fold change less than 0.5 are enriched in glia TRAP and depleted in neuron TRAP over Input. The middle category (4>FC>2) did not show enrichment in neurons but depletion in glia compared to Input, therefore, this group cannot be considered neither as neuronal nor invariant but rather as glia-depleted class. The other middle category is invariant among cell types (2>FC>0.5; grey in **Figure 5.8A**) and is depleted in both TRAP data compared to Input, suggesting that invariant transcripts are less likely to be bound to ribosomes.

Taken together, using a careful fold change cutoff, we defined very specific neuronal genes (FC>8), specific neuronal genes (FC>4), genes depleted from glia (4>FC>2), invariant genes (2>FC>0.5) and glial genes (0.5>FC).

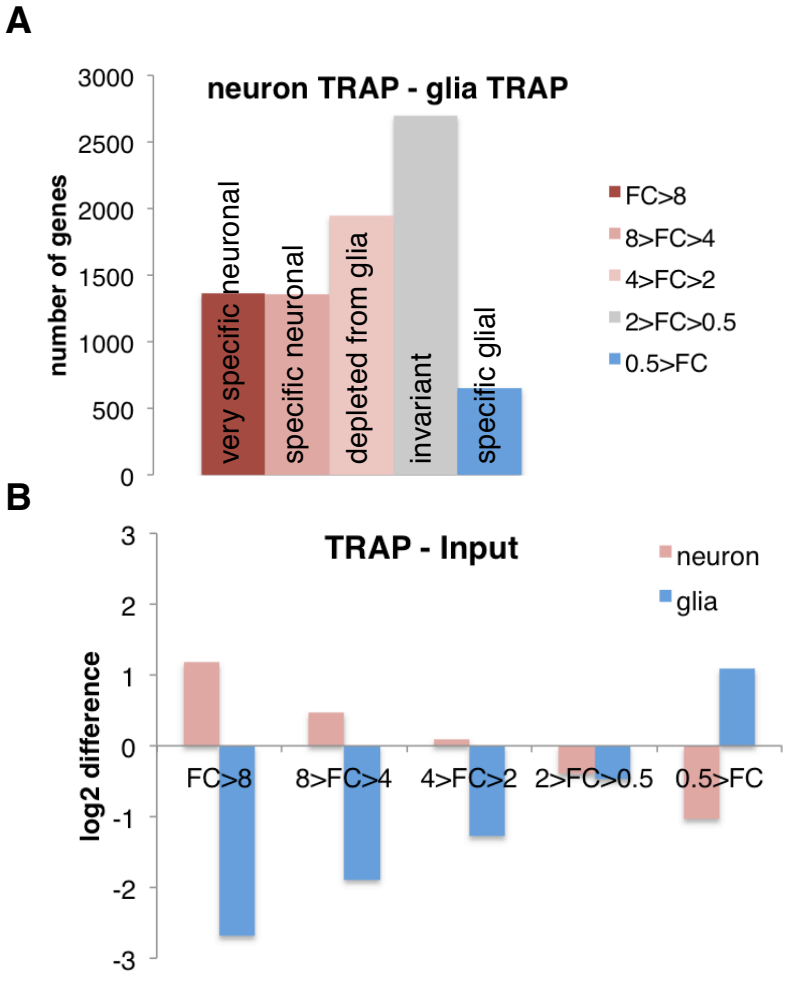
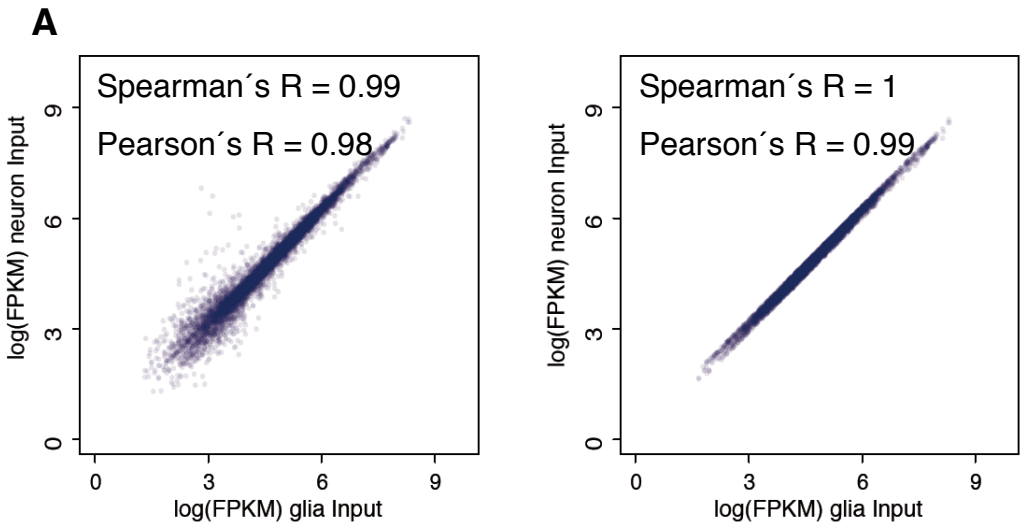


Figure 5.8 Comparison of neuronal and glial TRAP.

(A) Number of genes is shown in five categories based on fold change FPKM (FC) between neuron and glia TRAP (see legend). **(B)** Log2 difference of neuron TRAP compared to neuron head Input (red) and glia TRAP compared to glia head Input (blue) in the five categories from (A).

To further analyze the correlations among neuron and glia TRAP as well as Input, I plotted the FPKM values in scatterplots and determined Pearson´s and Spearman´s correlation coefficients (**Figure 5.9**). As before, I used only those genes that do not differ among Inputs (**Figure 5.9A**). The FPKM values are more scattered in the case of neuron TRAP - glia TRAP comparison (**Figure 5.9C**) compared to the TRAP vs. Input (**Figure 5.9B**), indicating a better distinction of differential genes in a cell-type-based comparison. Spearman´s correlations are very similar in TRAP-Input comparisons (R=0.94) and lower in TRAP-TRAP comparison (R=0.84). Pearson´s correlations differ from Spearman´s rho mainly in the case of TRAP-Input relations, suggesting that this is rather a monotonic correlation.

Scatterplots and Spearman´s correlations suggest that the dynamic range of differences in TRAP data is larger in TRAP/TRAP than in TRAP/Input comparisons. To test this, I plotted all FC values (in a log2 scale) in both cases as boxplots (**Figure 5.10**). FC values between neuron- and glia-TRAP are more spread out compared to neuron-TRAP/neuron-Input. This confirms that TRAP-TRAP comparisons have a better resolution, as shown by my previous observations with qPCR (**Figure 5.7**) and by published mouse data (Dougherty et al., 2010). Note that the median of TRAP-TRAP FC is shifted towards the positive values (**Figure 5.10 right**) in accordance with the previous observation that genes with a FC of two are not neuronal but rather genes depleted in glia (**Figure 5.8** and **5.10**).



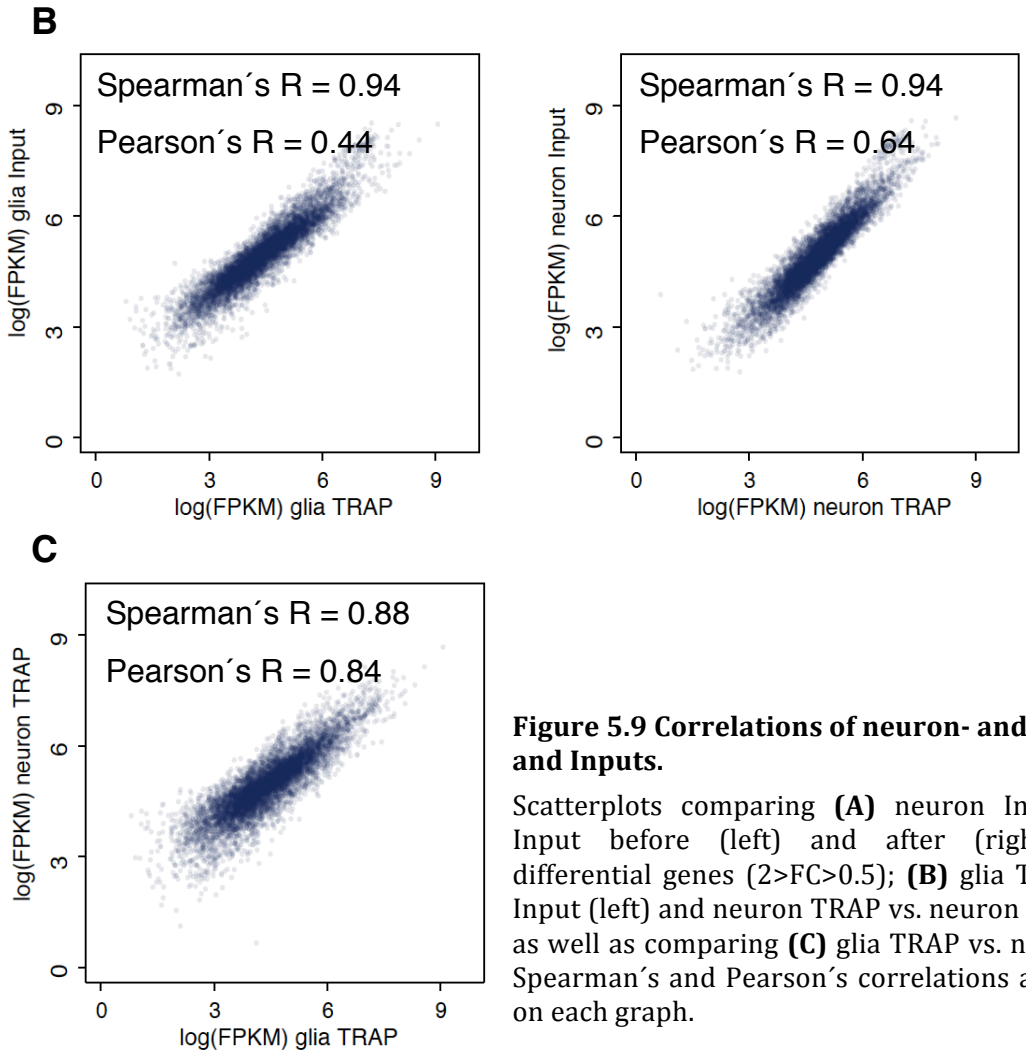


Figure 5.9 Correlations of neuron- and glia-TRAP and Inputs.

Scatterplots comparing **(A)** neuron Input vs. glia Input before (left) and after (right) filtering differential genes (2>FC>0.5); **(B)** glia TRAP vs. glia Input (left) and neuron TRAP vs. neuron Input (right) as well as comparing **(C)** glia TRAP vs. neuron TRAP. Spearman´s and Pearson´s correlations are indicated on each graph.

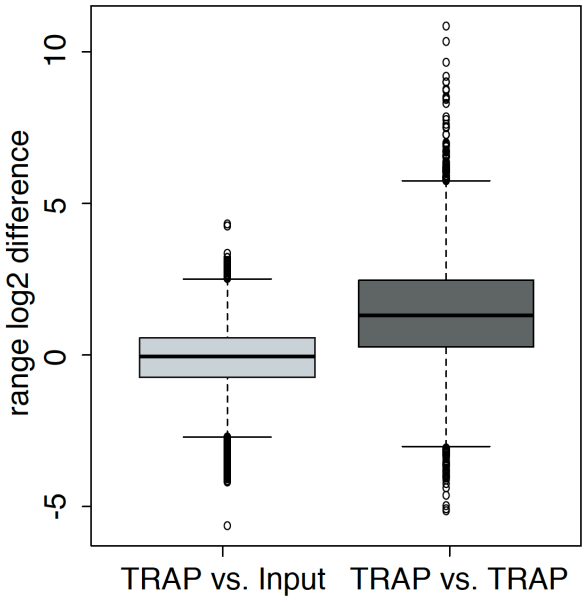


Figure 5.10 Dynamic range of TRAP-Input and TRAP-TRAP comparisons.

Boxplots showing log2 fold change values between neuron TRAP and neuron head Input (left, light grey) as well as neuron TRAP vs. glia TRAP (right, dark grey)

5.4.2.2 Validating TRAP

To validate the results obtained by cell-type-specific TRAP, I used biased and unbiased computational approaches. First, I compared neuronal genes, identified by TRAP with a FC higher than four, to genes with known neuronal function and expressed in neurons obtained by other studies (TRAP: Thomas et al., 2012; INTACT: Henry et al., 2012). I visualized a few examples of these on the scatter plot from **Figure 5.9C** on **Figure 5.11**. I found the neuronal marker *n-syb* (neuronal synaptobrevin), neurotransmitter receptors (*nAcRalpha-96Ab*, *5-HT1A*), the channel binding protein *Slob* (Ma et al., 2011; Sheldon et al., 2011) and the insulin-like peptide *Ilp2* (or *Dilp2*) among neuron-specific genes (**Figure 5.11**; **Figure 5.12** and see **section 2.1.2**). I also indicated neuron-TRAP-enriched genes that carry a neuronal RNA polymerase II peak obtained by CAST-ChIP (e.g. *king-tubby* and *igloo*; for a comparison see **Figure 3.16** and **section 5.4.2.3**).

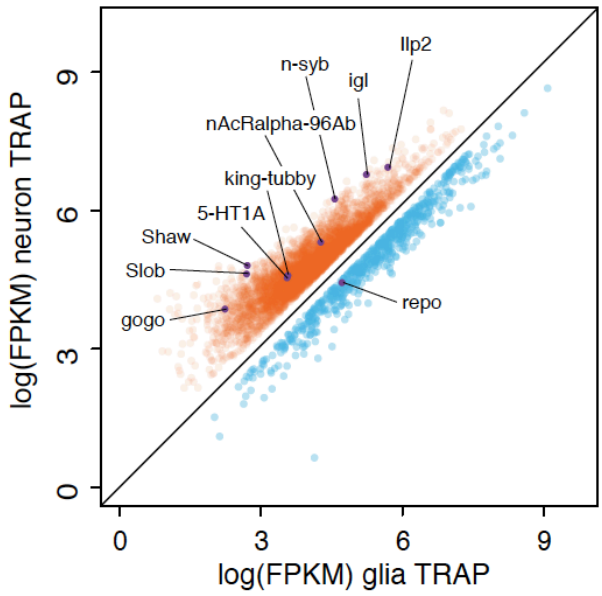


Figure 5.11 Scatterplot with marked neuronal genes.

Scatterplot compares neuron TRAP to glia TRAP (same as Figure 5.9C) showing genes with FC>2 in red and FC>0.5 in blue (right panel). Selected genes with known neuronal function are labeled.

Selected genes with known cell-type-specific function were also enriched in the cell type where they are functional, as shown by genome browser snapshots (**Figure 5.12**). The cathepsin encoding gene *crammer* (Comas et al., 2004), the cell adhesion molecule *midline fascilin* (*mfas*; Jacobs et al., 2000) and the *gcm* target *wrapper* (Egger et al., 2002) have a known glial function (also see **section 2.1.2**) and are clearly enriched in glia-TRAP (**Figure 5.12 left column**). Interestingly, I found 11 odorant binding proteins (e.g. *Obp99d*) and an insulin-like peptide (*Dilp6*; FC=0.56 just around threshold) among glia-specific genes, potentially indicating the role of glia in mediating the crosstalk between metabolism and behavior.

To further characterize genes enriched in cell-type-specific TRAP, I used the advantage of the online tool FlyMine (Lyne et al., 2007). I compared the TRAP datasets to FlyAtlas, a microarray collection of dissected tissues from different developmental stages (Chintapalli et al., 2007). I chose the most specific neuronal (1239; FC>8), glial (639; 0.5>FC) and also the invariant (2589; 2>FC>0.5) gene sets (defined in **Figure 5.8)** and determined how many of these are up- or down-regulated in FlyAtlas (**Figure 5.13**).

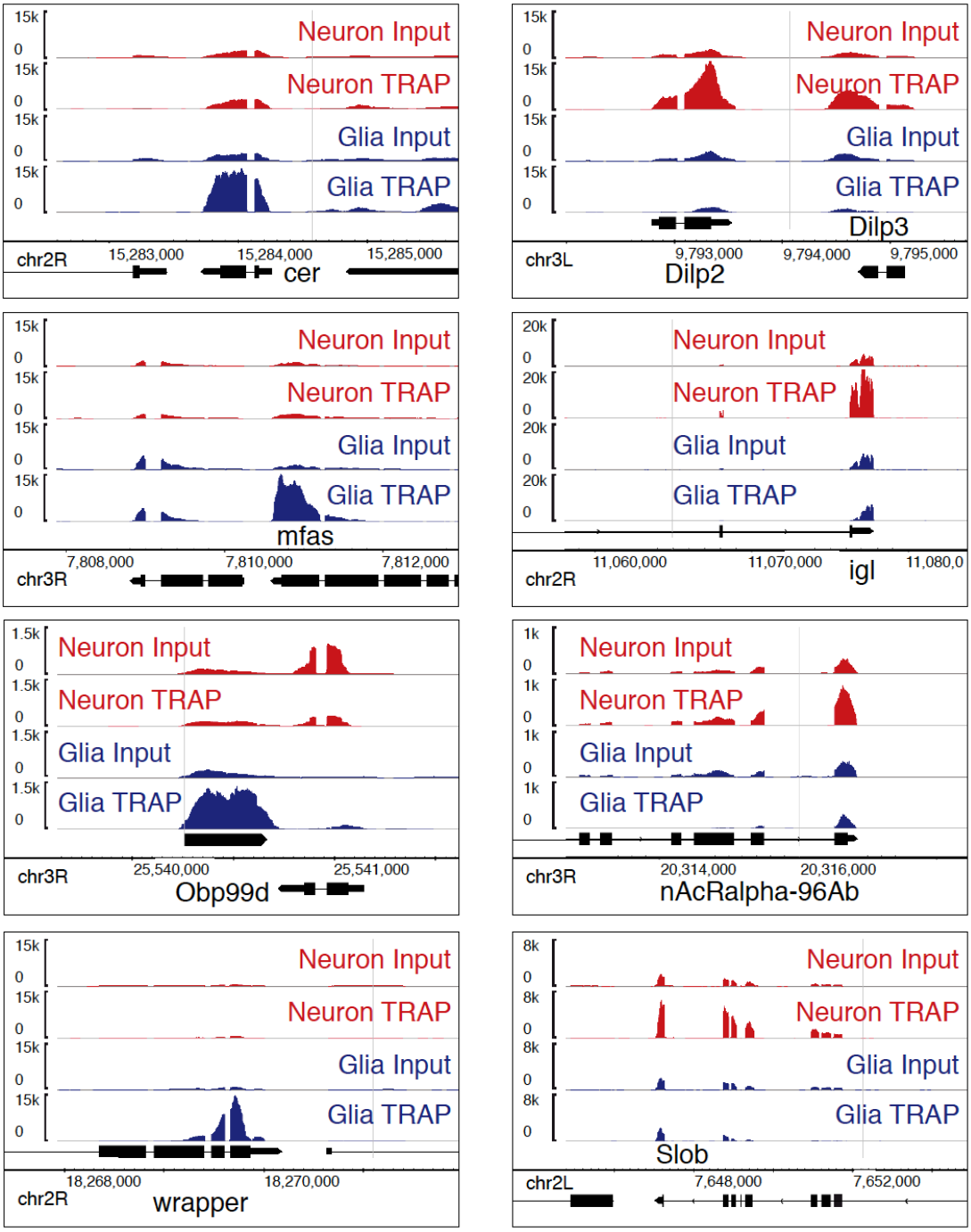


Figure 5.12 Genome browser view of selected glial and neuronal genes.

Left column shows glial genes enriched in glia TRAP (*cer*, *mfas*, *Obp99d* and *wrapper*) and right column shows neuronal genes enriched in neuron TRAP (*Dilp2*, *igl*, *nAcRapha-96Ab* and *Slob*).

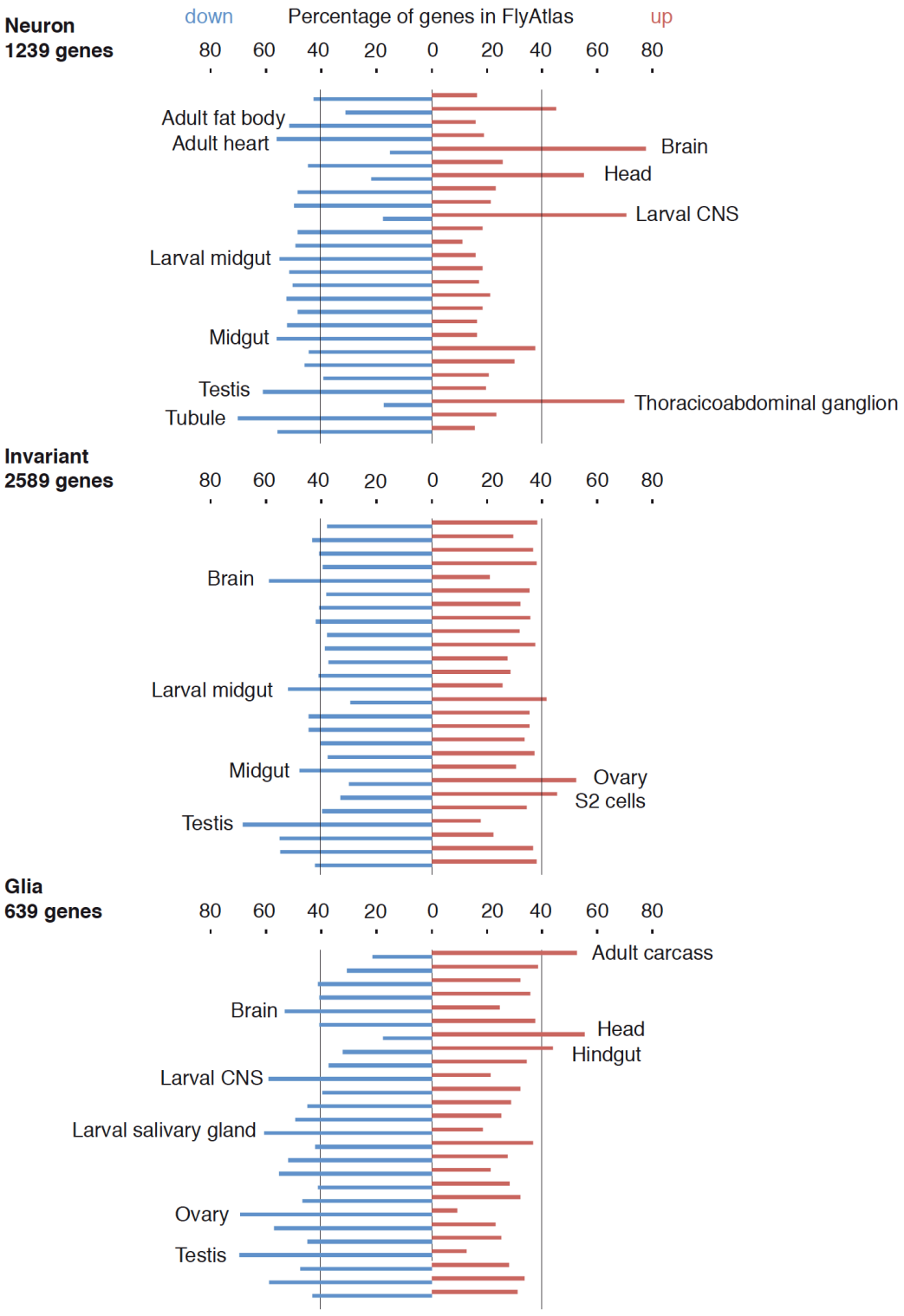


Figure 5.13 Comparison of cell-type-specific TRAP and FlyAtlas.

Percentage of genes up- (red) or down- (blue) regulated in FlyAtlas tissues in TRAP datasets of neuron- (top), glia- (bottom) specific and invariant (middle) genes.

Genes enriched in neuron-TRAP are over-represented among genes up-regulated in FlyAtlas tissues such as brain (76.8%), larval CNS (69.9%) and the thoracicoabdominal ganglion (69.1%; **Figure 5.13 top**). These genes are mainly down-regulated in tissues including tubule, testis, heart and the fat body. In general, invariant genes from the TRAP data are not up-regulated in any tissues except the ovary. In contrary, they are down-regulated in the brain, suggesting that this category lacks central nervous system related functions (**Figure 5.12 middle**). Specific genes in the glia TRAP dataset are up-regulated in the head but down-regulated in the brain and larval CNS (**Figure 5.13 bottom**), pointing to the important role of glia also in the periphery, i.e. outside of the brain (see **section 2.1.2**).

To test whether genes enriched in neuron- or glia-TRAP have a relevant cell-type-specific function, I performed GO analysis on the same specific or invariant sets (**Table 5.1**). GO terms obtained by FlyMine clearly demonstrate the importance of neuron-TRAP genes in biological processes, such as nervous system development, response to stimulus and axonogenesis. In addition, neuron-TRAP genes are part of neuronal cellular components, such as the synapse or projection (**Table 5.1**). Neuron-TRAP genes encode proteins having molecular functions, such as receptor or neuropeptide activity. Thus, I confirmed that neuronal genes discovered by TRAP carry real neuronal function (see **section 2.1.2.1**). Glial genes revealed by TRAP do not have such specific functions as neuronal ones (**Table 5.1**). This can be explained by the diverse function and type of glia cells (see **section 2.1.2.2**). Furthermore, I found house-keeping functions (e.g. catalytic activity or part of the ribosome) for genes in the invariant category (**Table 5.1).** This category contains genes that are invariant between glia and neurons, including genes that are potentially specific for other tissues such as the fat body. GO terms such as metabolic process and lipid particle support this idea.

Taken together, TRAP identifies neuronal and glial genes playing an important role in maintaining specific cellular functions. To reveal a more accurate picture, biological replicates and other tissues have to be involved in future analysis.

Neuronal genes (1239)

GO term	p-value	matches
biological process		
generation of neurons	8.4E-16	122
nervous system development	5.3E-15	208
neuron differentiation	7.2E-15	112
response to stimulus	1.9E-14	324
neuron projection morphogenesis	1.1E-11	83
axonogenesis	1.0E-08	61
axon guidance	5.6E-08	50
cellular component		
plasma membrane	2.3E-13	121
synapse	6.0E-12	48
axon	1.6E-09	28
neuron projection	4.7E-08	32
molecular function		
receptor activity	8.2E-08	82
neuropeptide hormone activity	1.2E-07	18
hormone activity	2.8E-06	21

Glial genes (639)

GO term	p-value	matches
biological process		
mitotic spindle elongation	2.17E-07	20
cellular component		
extracellular region	1.15E-16	79
ribosomal subunit	2.85E-10	30
molecular function		
glutathione transferase activity	6.95E-06	13
peptidase regulator activity	4.66E-05	20
active transmembrane transporter activity	5.88E-04	39

Invariant genes (2589)

GO term	p-value	matches
biological process		
oxidation-reduction process	1.25E-06	170
metabolic process	1.53E-05	1177
cellular component		
ribosomal subunit	5.84E-10	71
lipid particle	9.55E-10	93
molecular function		
catalytic activity	3.53E-22	954
oxidoreductase activity	9.01E-09	192

Table 5.1 GO analysis of neuron, glia TRAP and invariant genes.

GO terms obtained by FlyMine in gene sets of neuron TRAP (top), glia TRAP (middle) and invariant genes (bottom). GO term, p-value and the number of matching genes is indicated.

5.4.2.3 Comparing TRAP and Pol II CAST-ChIP

To analyze the correlation of RNA polymerase II binding and ribosome-bound RNAs in the distinct *Drosophila* cell types, I compared neuronal and glial CAST-ChIP results with

TRAP (see **chapter 3**). First, I took a list of the top 30 genes (from **Table 3.2**) carrying neuronal Pol II and having a known neuronal function. Most of these genes are enriched in the neuron TRAP data having a FC FPKM greater than 4 (**Table 5.2**). This biased approach indicates that genes with the most specific neuronal Pol II peak are not only transcribed, but are also translated in neurons.

FlyBase ID	Gene Name	FC Pol II	FC TRAP
FBgn0053202	dpr11	4.1	**92.2**
FBgn0005775	Con	3.7	**78.4**
FBgn0015774	NetB	3.4	**59.7**
FBgn0259246	brp	4.1	**40.6**
FBgn0085447	sif	3.1	**37.5**
FBgn0052227	gogo	3.5	**37.1**
FBgn0035092	Nplp1	2.9	**35.5**
FBgn0053171	mp	3.3	**27.0**
FBgn0013467	igl	3.2	**26.2**
FBgn0053960	Sema-2b	3.8	**23.0**
FBgn0259225	Pde1c	3.7	**22.7**
FBgn0010415	Sdc	3.6	**17.5**
FBgn0052635	Neto	3.7	**17.4**
FBgn0013759	CASK	3.0	**15.4**
FBgn0024944	Oamb	3.3	**14.3**
FBgn0086778	gfA	3.6	**14.3**
FBgn0032151	nAcRalpha-30D	3.5	**12.8**
FBgn0052183	Ccn	4.0	**12.7**
FBgn0051190	Dscam3	3.4	**11.6**
FBgn0028433	Ggamma30A	3.6	**10.1**
FBgn0050361	mtt	4.8	**9.4**
FBgn0033058	CCHa2r	3.6	**9.2**
FBgn0004118	nAcRbeta-96A	4.0	**8.7**
FBgn0015721	king-tubby	3.6	**8.6**
FBgn0250910	Octbeta3R	3.3	**8.0**
FBgn0004619	Glu-RI	4.0	**8.0**
FBgn0004168	5-HT1A	3.9	**7.9**
FBgn0017590	klg	4.0	**7.2**
FBgn0024963	GluClalpha	3.0	**3.2**
FBgn0025593	Glut1	3.5	**1.7**

Table 5.2 Comparison of top neuronal Pol II genes and TRAP.

Table contains the top 30 neuronal genes carrying Pol II in neurons from **Table 3.2**. Flybase ID, gene name, fold change of Pol II CAST-ChIP (neuron-glia) and fold change of TRAP (neuron-glia) are shown.

To match genes obtained by Pol II CAST-ChIP and TRAP, I compared all genes carrying a significant cell-type-specific Pol II peak (from **Figure 3.9**) and genes enriched in neuron TRAP (**Figure 5.8**; FC>4), in glia TRAP (FC>0.5) as well as invariant TRAP (2>FC>0.5). As expected, cell-type-specific Pol II binding and cell-type-specific ribosome-bound RNAs do not overlap completely (**Figure 5.14**, see **section 2.2.3**). More than half of

neuronal Pol II peaks match neuron-TRAP transcripts but there are about a thousand transcripts bound to ribosomes in neurons that lack discernible Pol II peaks. In contrast, only about one sixth of glia-specific Pol II genes are enriched in neuron-TRAP. The third Pol II category (invariant) is diverse. There are genes carrying shared Pol II between neurons and glia but also genes that have no Pol II in these two cell types (see **section 3.4.1**). Therefore, the small overlap of one-third invariant Pol II genes with invariant-TRAP genes is not surprising. Probably these genes are not active in any of the cell types. As expected, glia-TRAP and neuron-specific Pol II do not match, with only a few genes showing overlap (**Figure 5.14 second row**). Surprisingly, there is a relatively small proportion of glia-TRAP genes associated with glial Pol II peaks. This suggests that Pol II peaks in glia are not actively transcribing or glial transcripts associated with ribosomes are stably maintained without new transcription. Genes found in the invariant TRAP dataset are excluded from the neuronal category, confirming that these genes are not neuronal (**Figure 5.14 third row**). However, more than one third of glial Pol II genes overlap with invariant TRAP genes, corroborating the previous observation that glial peaks do not necessarily produce cell-type-specific mRNA transcripts.

These results reveal that TRAP identifies more than twice as many neuronal genes than Pol II CAST-ChIP. On the other hand, most of the Pol II peaks match neuronal transcripts. To demonstrate this point using a few examples, I took genome browser snapshots of neuronal genes that are enriched either in both Pol II CAST-ChIP and TRAP or enriched only in TRAP as well as carrying Pol II only (**Figure 5.15**). Neuronal genes such as *Ggamma30A* and *Similar to deadpan* (*Side*) carry RNA polymerase at their transcriptional start site and are enriched in the neuron-TRAP data (**Figure 5.15 first row**). Genes such as the neurosecretory insulin-like peptide encoding genes (*Dilp2, Dilp3* and *Dilp5*; see **section 2.1.2**) are highly enriched in the neuron-TRAP data but lack Pol II peaks (**Figure 5.12 and Figure 5.15**). The well-studied neuropeptide hormone (neuropeptide F [*npf*]; see **section 2.1.2**) was detected only by TRAP but not by Pol II CAST-ChIP (**Figure 5.15 second row**). In contrast, there are a few examples of neuronal Pol II binding but no enrichment in neuron TRAP, probably due to very low RNA levels (e.g. Dopamine receptor; *DopR*). There are also ambiguous situations where the neuronal Pol II peak is located near to a glia-TRAP-enriched gene but belongs to another TSS (*Syb-CG12913*).

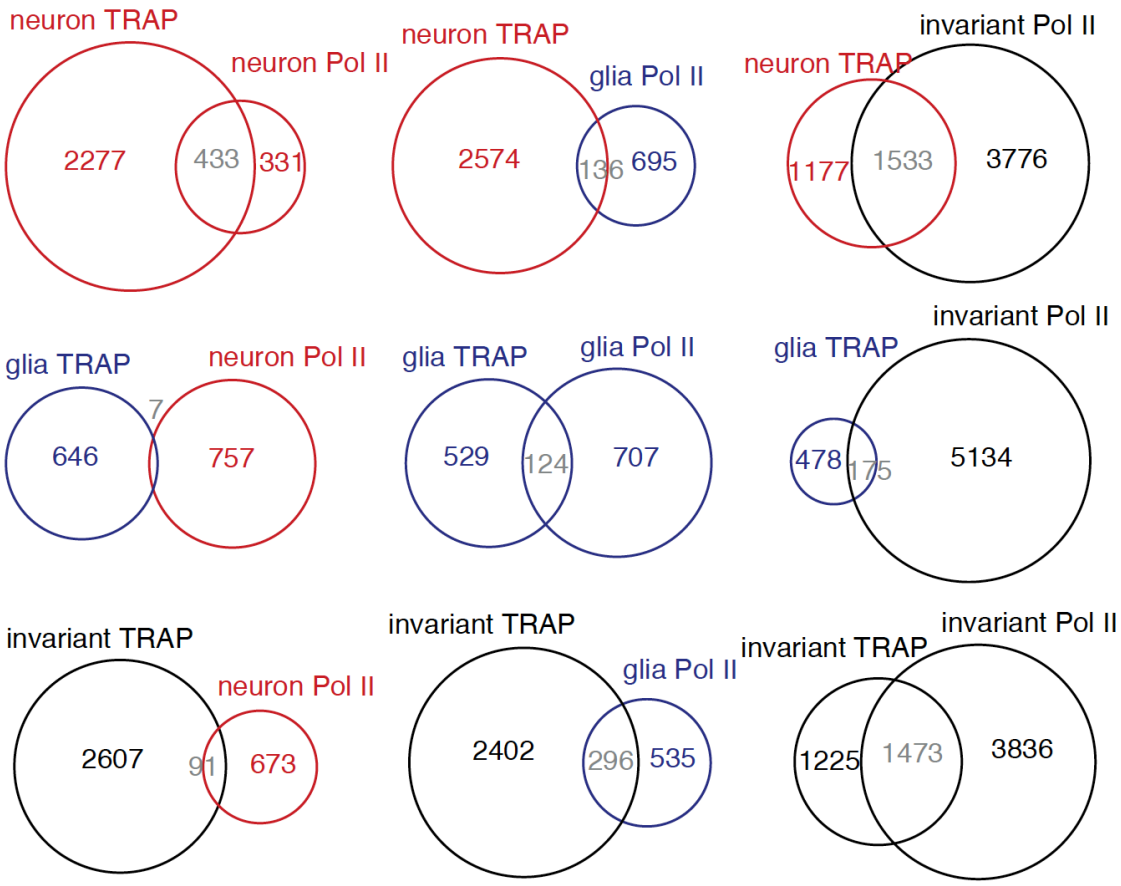


Figure 5.14 Overlap of cell-type-specific TRAP with Pol II genes.

Venn diagrams show the overlap of neuron TRAP genes (first row), glia TRAP genes (second row) and invariant TRAP genes (third row) with genes carrying neuronal Pol II (first column), glial Pol II (second column) and invariant Pol II (third column).

These results suggest that the dynamic range of cell-type-type specific TRAP is wider compared to chromatin-based approaches, such as ChIP against RNA polymerase II or H2A.Z (see **chapter 3** and **4**). To test this, I plotted all fold change values between neuron and glia obtained by TRAP, Pol II and H2A.Z ChIP (**Figure 5.16**). Indeed, TRAP FC values are clearly spread out, having a greater range compared to Pol II. As shown before, H2A.Z peaks do not differ among cell types (see **chapter 4**) and therefore, there is a small variation in FC between neurons and glia.

In summary, TRAP uncovers much more neuronal genes with relevant neuronal function compared to Pol II CAST-ChIP. These genes probably have very stable RNA molecules and their level is regulated primarily post-transcriptionally. On the other hand, I find in several cases that cell-type-specific Pol II binding does not match ribosome binding, suggesting that Pol II is in a stalled, transcriptionally inactive state.

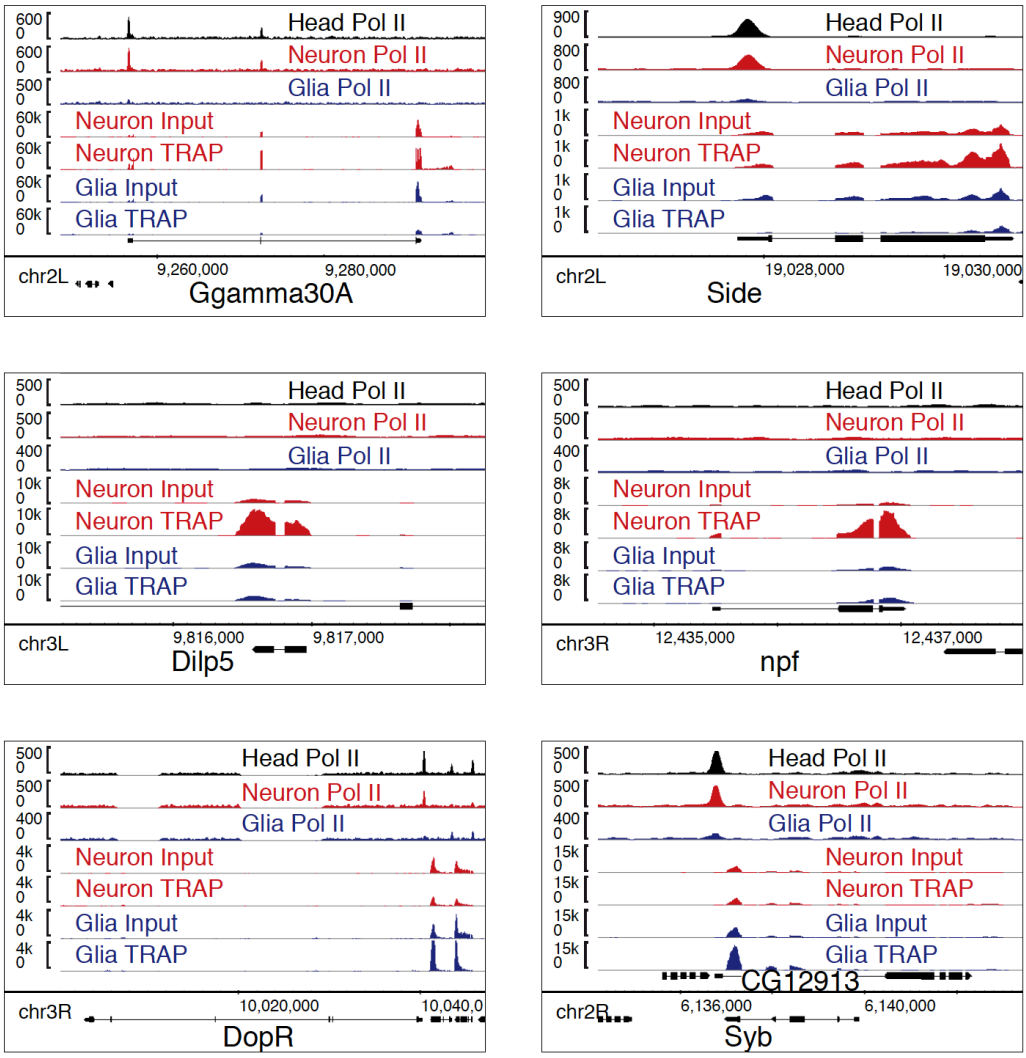


Figure 5.15 Genome browser view of neuronal Pol II peaks and TRAP.

Snapshots show Pol II CAST-ChIP peaks in head (black), neurons (red) and glia (blue) as well as TRAP from neurons (red) and glia (blue), Input and IP, respectively at genes such as *Ggamma30A*, *Side*, *Dilp5*, *npf*, *DopR* and *Syb*.

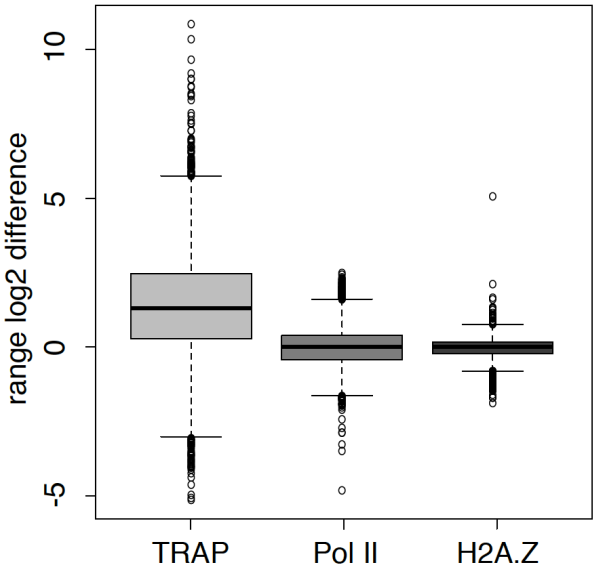


Figure 5.16 Dynamic range of TRAP, Pol II and H2A.Z.

Boxplots showing log2 fold change values between neuron and glia TRAP (left), neuron and glia Pol II CAST-ChIP (middle) as well as H2A.Z CAST-ChIP (right).

5.5 Discussion

In this chapter, I described an application of the translating ribosome affinity purification (TRAP; Heiman et al., 2008 and Doyle et al., 2008) method to dissect gene expression in distinct *Drosophila* cell types. Combining TRAP with the versatile UAS/Gal4 expression system ensures the translatomic profiling of specific cell populations (e.g. neurons) in *Drosophila* (Thomas et al., 2012). I established TRAP by cloning, expressing and immunoprecipitating a GFP-tagged ribosomal protein (GFP-L10A) in neurons, glia and fat body of the *Drosophila* head.

Affinity purification of GFP-L10A co-purified ribosomal and messenger RNAs, showing high enrichment over controls. I optimized cell-type-specific TRAP using different amounts of lysate in the comparison of neurons and glia cells. Quantitative PCR analysis of TRAPed mRNA confirmed the importance of both cell type to cell type and cell type to total comparisons as known from previous studies (Dougherty et al., 2010). I demonstrated this on transcripts highly enriched in neurons (i.e. *Dilp2*) or in glia (i.e. *repo*), which are depleted from the other cell type. Tripartite comparison revealed that there are transcripts (i.e. *Obp99b*) that are highly enriched in one cell type (in the fat body), present in another (i.e. neurons) and depleted from the third one (i.e. glia). The advantage of comparing at least three distinct cell populations is that transcripts invariant between two cell types might be different in a third one. Therefore, I suggest that the multiple-cell-type approach is the best way to identify most of the cell-type-specific vs. common transcripts.

After optimizing my TRAP protocol, I profiled cell-type-specific transcripts in a genome-wide manner. Gene expression profiling using RNA-Seq requires relatively large amounts of total RNA, because of the high proportion of non-informative ribosomal RNA compared to messenger RNA. Standard sequencing protocols (from Illumina) enrich polyadenylated mRNA using oligo-dT beads. Such approaches require at least four times more starting total RNA compared what is obtained with a single TRAP purification. Several other cell-type-specific transcriptomics methods overcome this problem using amplification kits (Steiner et al., 2012; Henry et al., 2012) to obtain the total RNA needed. To avoid artifacts from amplification prior to Illumina library preparation, (Tariq et al., 2011), I pooled several replicates to provide the minimal

amount required for the standard polyA protocol (see Methods). In this study, as a proof of principle, I showed a test sequencing run lacking biological replicates.

To identify differential ribosome-associated transcripts in neurons compared to glia, I applied a stringent fold change cutoff. I used this instead of DESeq due to the lack of replicates. For future analysis biological replicates with DESeq or the newer DEXSeq should be applied to ensure the statistical power and identify splice variants (Anders et al., 2010, 2012). Comparing TRAP results to FlyAtlas (Chintapalli et al., 2007) and gene ontology terms (FlyMine; Lyne et al., 2007), confirmed that I identified neuronal genes expressed in neuronal tissues (e.g. adult brain, larval CNS), having relevant neuronal function (e.g. receptor activity, neuropeptide hormone activity). In contrary, I do not find these genes in the glial dataset with specific function probably due to the diversity of glia cells (see **section 2.1.2**). Invariant genes, defined by TRAP, have general molecular functions, involved in metabolism or being part of the ribosome. Taken together, I am confident that the TRAP results provide a great collection of cell-type-specific and invariant genes.

Comparison of TRAP to other biochemical approaches, such as chromatin affinity purification from specific cell types (CAST-ChIP) against RNA polymerase II revealed the apparent advantage of TRAP for several reasons: 1) TRAP identified about a thousand more neuron-specific genes with a very stringent fold change cutoff; 2) the dynamic range of TRAP differences was much higher allowing the better distinction between specific and invariant genes and 3) TRAP overcomes the problem of stalled RNA polymerase with unknown elongation activity. On the other hand, stalled RNA polymerase might reflect those potential gene activation events that are hidden at the RNA level. Therefore, genes carrying RNA Polymerase II and having low mRNA levels are only detected by ChIP-based approaches. A combination of Pol II mapping and TRAP sequencing might thus better reveal gene activation events, where Pol II is paused under naive conditions and transcription is cell-type-specifically activated upon induction.

In summary, TRAP is a suitable tool to identify cell-type-specific transcripts by comparing distinct cell types from within an intact organism at the level of the translatome. TRAP is a versatile method, adaptable to several model organisms (mouse: Heiman et al., 2008, zebra fish: Tryon et al., 2012, fruit fly: Thomas et al., 2012 and this thesis). It has the advantage of an improved dynamic range compared to ChIP-based

approaches. Therefore, TRAP is likely to detect changes in gene expression upon environmental perturbation, which may be hidden at the level of Pol II binding

6 Discussion

The central dogma of molecular biology describes the process of how DNA is transcribed into RNA and RNA translated to proteins. Basically, this covers all the regulatory networks that determine how genes are expressed, manifesting in the organisms' phenotypes. Distinct cell types within a multicellular organism show various different phenotypes, but largely carry the same genome. Our understanding of how distinct cells obtain and maintain their phenotype, function and identity will profit from our ability to profile the distinct cell types of multicellular organisms at the level of epigenetics (or epigenomics) .

I sought to investigate several steps of gene expression in a comparative manner in differentiated cell types. Instead of comparing developing, dynamically changing cells, I chose terminally differentiated cell types, such as neurons of the fruit fly central nervous system. In the "epigenetic landscape" model, these are the end points of the slope where cell fate and identity have already been defined (see **chapter 2.1**). In order to identify genes specific for these cell populations and that likely play an important role in maintaining their function, I profiled: 1) transcription by chromatin immunoprecipitating (ChIPing) RNA polymerase II; 2) chromatin structure by mapping the histone variant H2A.Z and 3) cell-type-specific mRNA transcripts by purifying ribosome-associated mRNAs.

In this thesis, I presented novel genomic approaches to study gene regulation in a cell-type-specific manner. I combined the versatile UAS/Gal4 targeting system of *Drosophila melanogaster* with biochemical techniques. My methods rely on the specific expression of a tagged reporter in distinct cell populations followed by tag-mediated affinity purification and genome-wide profiling by next generation sequencing. I developed Chromatin Affinity purification from Specific cell Types (CAST-ChIP) for chromatin mapping and applied Translating Ribosome Affinity Purification (TRAP) for mRNA profiling of distinct cell types within the *Drosophila* head.

6.1 CAST-ChIP, a tool for cell-type-specific chromatin mapping

CAST-ChIP is a cell-type-specific chromatin immunoprecipitation method that uses a tagged chromatin-associated reporter expressed in the cell type of interest. Targeting of the reporter is mediated by the *Drosophila* UAS/Gal4 system that ensures its expression in a large variety of cell populations. I chose well-characterized cell types of the *Drosophila* head including neurons, glia and fat body. These cells have completely different function (neurons: information transfer, glia: buffering, fat body: energy storage), morphology (neurons: projections, glia and fat body: round-shaped) and developmental origin (neurons and glia: ectodermal, fat body: mesodermal [see **section 2.1.1** and **2.1.2**]). CAST-ChIP identified neuron-specific genes having specific functions such as neurotransmitter, neuropeptide or receptor (see **chapter 3**). The comparison of these three distinct populations allowed us to distinguish between cell-type-specific and -invariant genes.

As a proof of principle, Carla Margulies and myself applied CAST-ChIP to profile a subunit of RNA Polymerase II, as well as the histone variant H2A.Z. However, the CAST-ChIP reporter could be any abundant chromatin-bound factor, whose expression does not lead to mis-incorporation events. This includes other histones including variants (e.g H3.3) and general chromatin modifying enzymes (e.g. SAGA; Weake et al., 2011) involved in transcription (see **section 2.2.2**). Profiling H3.3 in cell types might reveal new insight of nucleosomes incorporating both H2A.Z and H3.3 or only one of them. On the other hand, transcription factors with low expression levels should be avoided in CAST-ChIP analysis, since their expression under a Gal4 promoter may induce overexpression artefacts. CAST-ChIP is a suitable tool to profile chromatin-associated proteins, however, it cannot be used to specifically profile post-translationally modified target proteins. One extension of CAST-ChIP, which my lab has started working on is the tagging of a canonical histone (H4 and H2B) coupled with a protease-site in the linker between the GFP and the histone. These flies should allow the purification of all chromatin from a specific cell type, followed by protease cleavage, further chromatin fragmentation to the level of mononucleosomes and then coupled with a conventional ChIP assay using antibodies that may recognize post-translationally modified histones

and non-histone proteins. Complementary to CAST-ChIP, fluorescent activated cell or nuclei sorting ChIP (FACS or FANS) could be used for highly abundant histone modifications.

CAST-ChIP is a rapid, efficient and sensitive technique (**Table 6.1**). A critical point of any genome-wide profiling experiment (ChIP- or RNA-Seq) is to preserve the gene expression state of cells during the experimental procedure. First, CAST-ChIP is an X-ChIP method using formaldehyde cross-linking, which stops ongoing transcription and new initiation. This is an advantage compared to RNA-based gene expression profiling methods, where fixing should be avoided (see **section 2.3.3**). Second, CAST-ChIP is carried out within the regular time frame of any other ChIP experiment without additional steps, such as for example treatment of the sample at temperatures higher than 4 °C. Other methods such as FACS sorting require several hours to separate labeled nuclei needed for ChIP-seq (see **section 2.3.2** and Bonn et al., 2012). Third, CAST-ChIP does not need any special, expensive equipment (i.e. FACS sorter), only a cooled bench environment, similar to other commonly available biochemical purification methods. Fourth, CAST-ChIP is easily parallelized to obtain the required amount of ChIPed DNA for sequencing, which is a usual problem of cell-type-specific methods. Methods such as INTACT use amplification of ChIPed DNA upon purifying whole nuclei (see **section 2.3.2** and Deal et al., 2010). Although such kits might offer a linear amplification, I wanted to avoid any potential bias coming from such a step. Fifth, I demonstrate that CAST-ChIP is a sensitive approach to identify thousands of cell-type-specific sites between neurons and glia. RNA polymerase II profiles obtained by CAST-ChIP provide a compendium of cell-type-specific transcription.

Table 6.1	CAST-ChIP	Other methods
preservation	crosslinking	RNA-based: no crosslinking
time	regular ChIP	FACS: long sorting time
equipment:	bench	FACS sorter
required DNA	pooling replicates	INTACT: amplification
sensitivity	thousands of diff. genes	thousands of diff. genes
targets	chromatin bound proteins	post-translational modifications

Table 6.1 CAST-ChIP is a rapid, efficient and sensitive method.

Comparison of CAST-ChIP to different methods by technical features such as how the sample is preserved, the time needed, the required amount of DNA for sequencing, sensitivity (the number of differential genes identified) and the potential targets for profiling.

6.2 Ubiquitous genes, are they special?

Genes in multicellular organism can be grouped according to their cell-type-specific or ubiquitous expression. Ubiquitous genes include housekeeping genes, which are not only expressed in most of the cell types but they maintain essential cellular functions. In contrast, cell-type-specific genes are involved in specific processes during development; therefore, their expression differs in space and time. Genes within these two groups are not only distinguishable by their expression pattern, but also by several genomic and epigenomic features. Housekeeping genes are shorter, more compact, having less exons compared to longer, cell-type-specific genes with long introns (De Ferrari et al., 2006). Promoter architecture of housekeeping and cell-type-specific genes also differs, ensuring their ubiquitous or specific expression. Genes with ubiquitous expression in developmental space and time have broad promoters, whereas genes with restricted expression have peaked promoters (Hoskins et al., 2011 and Rach et al., 2011). Co-regulated genes, such as housekeeping genes, tend to cluster together, whereas specific ones are often in gene poor regions (Weber et al., 2011; Filion et al., 2010).

Cell-type-invariant and specific genes not only differ in their genomic but also in their transcriptional and epigenetic features. RNA polymerase II is uniformly bound to ubiquitously expressed genes and preferentially stalled at the promoter of developmental genes (Zeitlinger et al., 2007). Using CAST-ChIP, I identified thousands of cell-type-specific RNA polymerase II sites, indicating that ubiquitous and specific genes are distinguishable at the level of Pol II binding (see **chapter 3.4**). On the other hand, by profiling RPB3 (or GFP-RPB3) in *Drosophila* head, I could not detect elongating polymerase to distinguish between elongating and stalled sites. Usually studies investigating Pol II stalling (Zeitlinger et al., 2007; Gilchrist et al., 2010) use embryos or embryonic S2 cells, which might differ in chromatin structure, Pol II accessibility or "ChIPibility" compared to adult tissues. To find active elongating polymerase in a cell-type-specific manner, ChIP against Pol II CTD Ser2 phosphorylation (reviewed by Egloff

et al., 2012 and see **section 2.1.2.2**) should be carried out, probably on FACS sorted nuclei.

Housekeeping genes share epigenetic marks including chromatin modifications and binding of specific protein complexes. Genes with ubiquitous expression associate with the constitutively active chromatin (YELLOW) in the 5-state model (Filion et al., 2010). This type of chromatin carries the histone mark H3K36me3 and associated factor MRG15. Housekeeping genes overlap with the NSL (non-specific-lethal) complex, which contains the histone acetyl-transferase MOF responsible for H4K16ac (Feller et al., 2011; Lam et al., 2012; *Schauer et al. unpublished*). NSL regulates housekeeping genes with particular promoter architecture (with DRE motifs) by modulating Pol II initiation.

Housekeeping genes not only have broad promoter architecture, but also display a special chromatin structure carrying well-positioned nucleosomes downstream of the TSS (Rach et al., 2011). Genes having broad promoters associate with H2A.Z and class I insulator-binding proteins in a conserved way (human: CTCF; fly: CTCF, BEAF-32 and CP190; Rach et al., 2011). We confirm previous findings in our adult fly head data: 1) H2A.Z is incorporated to positioned nucleosomes close to the TSS and together with Pol II associates with 2) broad promoters, 3) constitutive chromatin (YELLOW in Filion et al., 2010) and 4) class I insulator binding proteins (**chapter 4.4**). Using CAST-ChIP, I show that H2A.Z marks ubiquitous genes and is absent from cell-type-specific ones (defined by Pol II). H2A.Z is invariant among cell types of the fly head and also compared to embryo. H2A.Z bound genes fulfill all the features of housekeeping genes; they do not differ among dissected tissues of the FlyAtlas (Chintapalli et al., 2007); they cluster such as low-specificity Tau cluster genes (Weber et al., 2011); and they are in general short in length (data not shown). Therefore, our cell-type-specific approach, CAST-ChIP, identifies a novel features of H2A.Z as a specific feature of cell-type-invariant genes.

Further classifying genes, there are genes that are constantly active (or inactive), whereas some are inducible upon developmental or environmental changes. During development genes are rapidly activated and inactivated in a spatio-temporal manner. Developmentally induced genes are often cell-type-specific, e.g. the developmental hormone ecdysone induces a differential response in different cell lineages (Davis et al., 2011). In contrary, a group of inducible genes might be ubiquitous in drastic

environmental changes, where the whole organism is exposed to a shock. The *Drosophila* heat shock response, for example, is such a process, showing rapid activation of RNA polymerase II at the *Hsp* genes in all cell types. The heat shock gene *Hsp70* carries H2A.Z and a paused RNA polymerase in an un-induced state. We and others show that Pol II occupies the entire *Hsp70* gene and nucleosomes are lost upon heat shock in adult head (see **section 4.4.1**), as well as in embryonic cells (Petesch et al., 2008). This indicates that *Hsp70* is indeed a ubiquitous but also inducible gene, and therefore it is marked by H2A.Z, allowing for its rapid induction.

Taken together, ubiquitously expressed genes are co-regulated in a specially coordinated manner. They share genomic and epigenomic features that define the local chromatin domains/clusters of such genes. Using CAST-ChIP, I could distinguish between genes ubiquitously expressed (marked by H2A.Z and Pol II), ubiquitous genes with low expression (H2A.Z only) and specifically regulated genes (marked by cell-type-specific Pol II without H2A.Z (see **chapter 4.4** and **Figure 6.1**). A future direction would be to define the relationship of insulator and other *cis*-regulatory elements acting on ubiquitous *vs*. specific genes and separating these domains in a three-dimensional nuclear space.

Figure 6.1

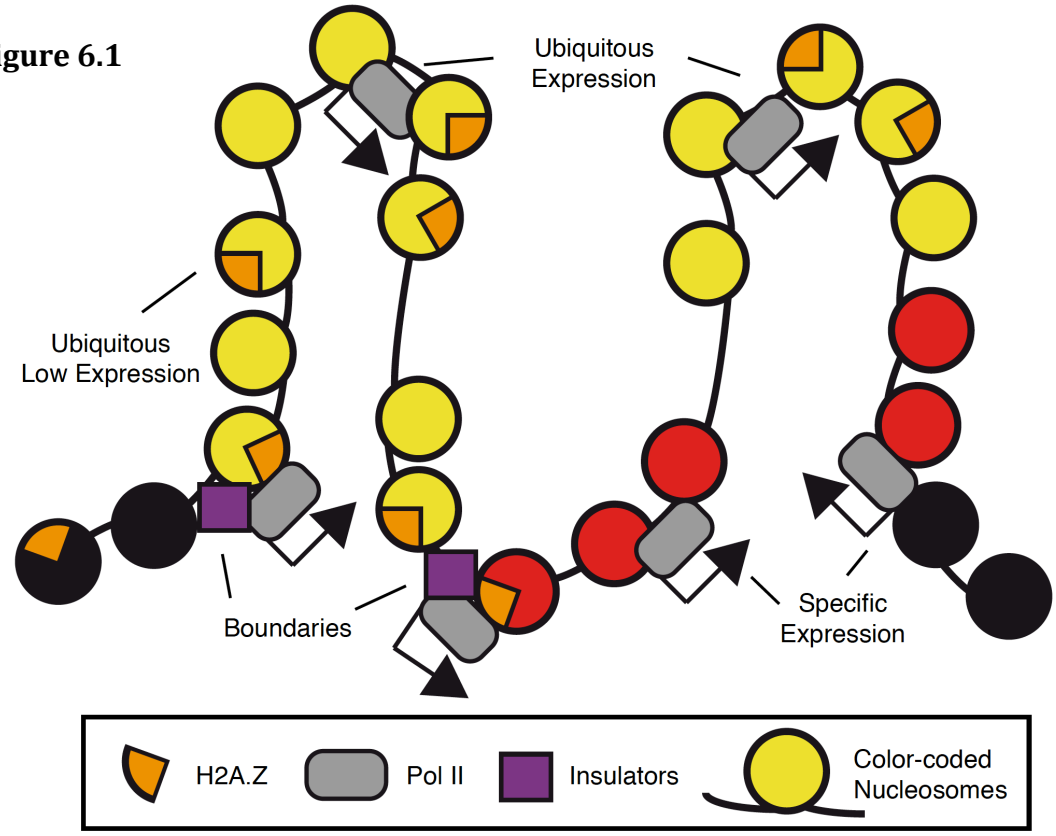


Figure 6.1 Summary of ubiquitously and specifically expressed genes.
Ubiquitously expressed genes are marked by H2A.Z and Pol II whereas specific genes by Pol II only. Insulator binding proteins present at the domain (color coded domains described in Filion et al., 2010) boundaries associate with H2A.Z enriched regions.

6.3 TRAP, profiling mRNA from cell types

Chromatin profiling gives insights about the transcriptional state of cell-type-specific and ubiquitous genes. However, ChIP-based approaches lack the information about steady-state mRNA levels. Each step of post-transcriptional regulation gives additional complexity to the system, influencing which transcripts are translated. To better understand cell identity and function, mRNA levels have to be profiled in a cell-type-specific manner.

Several approaches tried to address the difficulty of isolating mRNA from distinct cell types. 1) RNA molecules are less stable especially compared to chromatin-protected DNA. 2) Cross-linking (especially with formaldehyde) should be avoided to keep the RNA intact for high-throughput profiling (Karsten et al., 2002; VanDeerlin et al., 2002). 3) The amount of purified RNA needed for sequencing is about 1 µg compared to few nano-grams of ChIPed DNA. 4) Several previous approaches (e.g. laser-capture micro-dissection) do not give the required specificity, showing contamination from other tissues (see **section 2.3.3** and Okaty et al., 2011). Fulfilling these criteria when isolating mRNAs from different cell population has been a challenge. Applying and optimizing TRAP in *Drosophila*, I developed the conditions where I obtained ribosome-bound mRNAs that are highly cell-type-specific (**Figure 5.6-5.7**) and the amount of RNA is enough for high-throughput sequencing (**section 5.4.2**). To avoid potential degradation or sample loss, TRAP is carried out in a protecting buffer system (cycloheximide blocks new translation, RNasIn protects from RNA degradation). Therefore, TRAP is a reliable and suitable method for profiling mRNAs obtained from specific cell types.

TRAP is an ideal approach to investigate the cell-type-specific translatome. The pool of ribosome-associated mRNAs is determined by three factors: 1) transcription of newly synthesized mRNA, 2) new binding of existing transcripts to the ribosome and 3) regulated degradation of mRNA molecules. TRAP measures the steady-state level of ribosome-bound mRNA, which is the net effect of these three factors. Changes in either

ribosome binding or nascent transcription would influence the result of TRAP. TRAP was not yet performed in a changing environment, where perturbations or mutations induce regulatory changes to gene expression. Recent studies report the uncoupling of the transcriptome and the translatome in mammals (Tebaldi et al., 2012) and show preferentially translatomic changes under mild stress conditions in yeast (Halbeisen et al., 2009). Therefore, TRAP is an ideal tool to capture changes in the cell-type-specific translatome upon environmental induction.

TRAP in combination with other cell-type-specific methods that detect Pol II binding (CAST-ChIP; **chapter 3**), Pol II elongation (Ser2-P ChIP; see **section 2.2.1.2**), nascent transcription (NET-Seq) or short nuclear RNA (see **section 2.2.3**) would uncover all the main regulatory steps of cell-type-specific gene expression. Taken together, cell-type-specific approaches are complementary to each other; each of them gives insight of gene regulation in a different perspective. In order to understand how cell-type-specific phenotypes are manifested, a combination of methods (ChIP *vs*. RNA) and multiple cell populations will have to be investigated.

6.4 Perspective

Multicellular organism contain many distinct cell types that carry the same genome but show different phenotypes. To study the relationship between genotype and cell-type-specific phenotype, each regulatory step in between has to be investigated on a cellular basis. Thus far, genome-wide profiling methods allowed us to investigate the epigenomes, transcriptomes or translatomes on the organismal and tissue/whole organ level, as well as in culture (including, for example, stem cells). The rapidly developing high-throughput sequencing technologies make it possible to profile smaller and smaller amounts of biological starting material; therefore, we can reach more specific cell populations.

In this thesis, I presented cell-type-specific approaches to investigate transcription, by RNA polymerase II binding and chromatin, by histone variant incorporation using CAST-ChIP, as well as mRNA, by ribosome purification using TRAP. All of these techniques can be further developed to dissect differences among more specific groups of cells. Recently, ChIP-seq was optimized starting from a few thousand

of cells using high fidelity, linear amplification (Shankaranarayanan et al., 2011). Furthermore, Thomas et al. profiled the translatome of about 200 neurosecretory cells using TRAP without amplification. The next step has been to profile the transcriptome on the single cell level (Tang et al., 2011). Such techniques have been used to study individual neurons or even sub-compartments of neurons (Qiu et al., 2012; Batish et al., 2012), as well as individual tumor cells (Ramskold et al., 2012). Single cell transcriptomics has its limitations, such as the bias towards the 3´ end of transcripts and the lack of simultaneous sequencing of both genome and transcriptome (Tang et al., 2011). Although such data are still noisy, differential transcripts could be identified providing useful information at the single cell level.

Using novel cell-type-specific approaches allows us to systematically dissect cell populations within complex tissues, such as the central nervous system (CNS). Here, I presented a comparison of the two main cell types (neurons vs. glia), which are present in close proximity within the CNS, making manual dissection very tedious and time consuming. CAST-ChIP and TRAP enabled the biochemical dissection of these cell types and identified thousands of neuron- or glia-specific genes (see **chapter 3** and **5**). However, neurons and also glia cells are quite a diverse population of cells (see **section 2.1.2**). There are specific subgroups of neurons that are characterized by different neurotransmitters (e.g. dopamine) and their receptors (e.g. dopamine receptor). Neurons that are located to different parts of the brain, but express the same neurotransmitter, could be labeled by specific Gal4 lines (e.g. TH-Gal4) and systematically profiled by CAST-ChIP or TRAP (see **section 2.1.2.1**). The diverse populations of glia cells (e.g. ensheathing or astrocyte-like; see **section 2.12**) could also be categorized to better understand their function. In addition, there are important sub-anatomical structures within the fly brain involved in different neuronal and behavioral processes (see **section 2.1.2**). Identifying genes that play a specific function in such structures, including the mushroom body or ellipsoid body, would provide a great resource for neurobiological and behavioural research.

In the last decades several chromatin and transcription studies used heat shock as model system for transcription activation in *Drosophila* (Rasmussen et al., 1993). This is a drastic response with a high increase of transcript levels within few minutes upon induction (Adelman et al., 2006). Heat shock induces a general response in all cell types, since proteins have to be protected everywhere independently of the cell type. In

support of this hypothesis, the heat shock response in the head fly head was very similar to S2 cells or embryos (Adelman et al., 2006 and **Figure 4.3**). In contrast, changes occurring in a "natural environment" are probably regulated in a cell-type-specific manner. Such environmental changes may be altered by the supply of nutrients, different light-dark cycles, mild temperature changes or differences in social environment, such as altered crowdedness or sleep deprivation. These fluctuations influence fly behavior, such as circadian rhythm, feeding, mating or egg-laying.

One of the basic biological processes that follows environmental perturbation in a cell-type-specific manner is metabolism. Starvation in flies induces a complex response including metabolic, transcriptomic and translatomic changes, which are tissue- and cell-type-specific (reviewed by Teleman et al., 2010). Signaling molecules (e.g. insulin-like peptides), transcription factors (e.g. FOXO and CREB; Alic et al., 2011), translation regulators (e.g. 4E-BP) regulate metabolism in cell types, such as neurosecretory cells, serotoninergic neurons or fat body cells (Teleman et al., 2010 and see **section 2.1.2**). Our knowledge about these metabolic gene regulatory networks is limited to the organismal or to the manually dissectible level (Fujikawa et al., 2009; Farhadian et al., 2012). Now, using the CAST-ChIP and *Drosophila* TRAP procedures that I have developed in this thesis, cell-type-specific gene expression profiling approaches will make it possible to investigate these metabolic pathways in distinct cell populations. Understanding the spatio-temporal gene expression dynamics upon nutritional stress may allow us to dissect and distinguish general metabolic pathways (e.g. lipid metabolism Gronke et al., 2005) and specific mediators, for example, involved in odorant recognition (e.g. odorant binding proteins: Fujikawa et al., 2009; Farhadian et al., 2012 and Swarup et al., 2011). CAST-ChIP and TRAP will provide a more precise picture of how gene expression is regulated upon environmental and behavioral alteration.

7 List of Figures and Tables

8 Abbreviations

5-HT1A - serotonin (5-HydroxyTryptamine) receptor 1A
AchR - Acetylcholine Receptor
BEAF-32 - Boundary Element-Associated Factor of 32kD
CAST-ChIP - Chromatin Affinity purification from Specific cell Types - ChIP
ChIP - Chromatin ImmunoPrecipitation
CP190 - Centrosomal Protein 190kD
CRM - Cis-Regulatory Module
CTCF - CCCTC-binding Factor
CTD - C-Terminal Domain
DBD - DNA-Binding Domain
DopR - Dopamine Receptor
elav - embryonic lethal abnormal vision
FACS - Fluorescence Activated Cell Sorting
GO - Gene Ontology
GFP - Green Fluorescent Protein
H2A.Z - Histone-2A variant Z
H3 - Histone-3
H3.3 - Histone variant 3.3
H3K4me3 - Histone-3-tri-methyl-Lysine(K)-4
H3K9me3 - Histone-3-tri-methyl-Lysine(K)-9
H3K27me3 - Histone-3-tri-methyl-Lysine(K)-27
H3K36me3 - Histone-3-tri-methyl-Lysine(K)-36
H4K16ac - Histone-4-acetyl-Lysine(K)-16
HAT - Histone AcetylTransferase
HDAC - Histone DeAcetylase Complex
HP1 - Heterochromatin Protein 1
Hsp70 - Heat shock protein 70
igl - *igloo*
ILP - Insulin-Like Peptide
nAcRbeta-96A - nicotinic Acetylcholine Receptor beta 96A
Nmdar1 - NMDA receptor 1
Obp99b - odorant binding protein 99b
PcG - Polycomb Group protein
Pol II - RNA polymerase II
repo - reversed polarity
Rp49/RpL32 - Ribosomal protein L32
RpL10A - Ribosomal protein L10A
sNPF - short NeuroPeptide F
SWI/SNF - SWItch/Sucrose NonFermentable
TF - Transcription Factor
to - takeout
TRAP - Translating Ribosome Affinity Purification
TSS - Transcription Start Site
UAS - Upstream Activating Sequence

9 Appendix

9.1 CAST-ChIP protocol

9.1.1 Chromatin preparation

9.1.1.1 Fly collection

1. Collect GFP-RPB3 flies in polypropylene tubes, freeze in liquid nitrogen and store at -80°C (Small prep: 5 ml flies in 15ml tube)

9.1.1.2 Head removal

1. Vortex or hit tubes of flies on the bench top to knock off the heads (keep it cold in liquid nitrogen; 3 rounds of hitting, 5x on each side of the tube)
2. Use a sieve sized 630 µm to separate heads from the bodies and a sieve sized 400 µm to separate the heads from the wings and legs (keep it cold on dry ice)

9.1.1.3 Chromatin preparation procedure

Pre-preparation

1. Use clean gloves and try to work as clean as possible
2. Use filter tips and low-binding tubes (Costar)
3. Buffers should be filtered (0.2 µm filter) prior storage
4. Don´t use old buffers (discard older than 3 month)
5. Freshly add PIC tablets and PMSF
6. Cool down the sonicator and centrifuges to 4°C (1-2 hours before usage)
7. Work on ice and use cooled centrifuges

Buffers

NE Buffer

15 mM HEPES pH 7.6
10 mM KCl
5 mM MgCl2
0.1 mM EDTA
0.5 mM EGTA
350 mM sucrose
0.1% Tween 20
freshly add: PIC (2 tablets in 100ml), PMSF (200x), DTT (final: 1 mM)

RIPA Buffer

25 mM HEPES pH 7.6
150 mM NaCl
1 mM EDTA
1% Triton X-100

0.1% sodium deoxycholate
0.1% SDS
freshly add: PIC (2 tablets in 100ml), PMSF (200x stock)

Homogenization
1. Grind frozen heads with mortar pestle to a fine powder (chilled on dry ice)
2. Collect the heads in a 100 ml mechanical Dounce homogenizer (chilled on ice)
3. Immediately add 25 ml ice-cold NE buffer (DTT+PIC+PMSF), incubate 5 min on ice
   100 ml NE Buffer
   freshly add :
   100 µl 1M DTT
   500 µl 200x PMSF stock
   2 tablets of PIC (Protease Inhibitor Cocktail; Roche)
4. Dounce in the mechanical homogenizer (rotation: 2000rpm, 20 up-downs, slowly, no bubbles)
5. Transfer homogenate to a 50 ml Falcon tube
6. Add 675 µl formaldehyde (stock: 37%; final concentration: 1%) to fix
7. Rotate 10 min at room temperature
8. Add 1.25 ml glycine (stock: 2.5M) to stop fixing
9. Rotate 5 min at room temperature
10. Let stand on ice until the other samples are done
11. Filter through 60 µm Steriflip filter (Millipore)

Washing nuclei
1. Collect nuclei: spin 2000G 5 min at 4°C (pre-cooled centrifuge)
2. Resuspend the pellet in 1 ml RIPA (ice cold; PIC+PMSF freshly added)
3. Transfer to low-binding 1.5 ml tubes
4. Wash nuclei 2x in RIPA (spin 3500G 1 min)
5. Take up the pellet in 400 µl RIPA

Sonication
1. Pre-sonication (opening the nuclei)
   Branson250 sonifier
   400 µl in low-binding tubes
   settings:
   intensity: 5; pulse 60
   15x bursts; 45s break
   7 cycles (after 2 cycles longer break)
(keep it cold in water-ice, avoid bubbles, the tip should be close to the bottom of the tube, spin down if it´s foaming)
2. Sonication (shearing the chromatin)
   Covaris sonicator (cool down and degas in advance)
   small tubes: 130 µl (split the 400 µl to 3x 130µl)
   settings:
   200bp program; DC 20%; PIP 175; 4 min
   Collect the 3 samples (of the same) into 1 tube
3. Spin down 10 min; full speed; at 4°C

Measure DNA concentration with Nanodrop
1. Blank with RIPA (ice cold like your sample)
2. Avoid bubbles
3. Measure concentration (200-500 ng/µl)
4. Print the profile
5. Aliquot 50 µg (enough for 4 IP)

6. Store chromatin at -80°C

Crosslink reversal, DNA purification and size check (5 µg chromatin required)
1. Dilute 5 µg chromatin to 100 µl with TE
2. Incubate overnight at 65°C (shaking 1000 rpm)
3. Add 1 µl RNaseA (10 mg/ml), incubate 30 min at 37°C
4. Add 5µl 10% SDS and 1µl proteaseK (10mg/ml), incubate 1.5 h at 55°C
5. Purify DNA using Minelute columns (Qiagen)
   5.1 Add 500 µl PB (binding buffer) and mix
   5.2 Apply sample (600µl) to the column and spin for 1 min at room temperature
   5.3 Discard the flow-through
   5.4 Wash the column **2x** with 500 µl PE (wash buffer)
   5.5 Discard the flow-through
   5.6 Centrifuge for an additional 2 min
   5.7 Place the column in a clean low-binding 1.5ml tube
   5.8 Let it stand on the bench with open lid for 1 min (let the ethanol evaporate)
   5.9 Elute with 15µl EB (elution buffer), let it stand on the bench for 5 min, spin at 10000rpm (not max speed) for 2 minutes (caps can break)
6. Add 6 µl orange loading buffer (total volume: ~20 µl)
7. Run 10 µl on a 1.5% agarose gel (80V; 45 min; DNA ladder: "1kb plus" 7 µl)

9.1.2 Chromatin Immunoprecipitation

9.1.2.1 ChIP procedure

Pre-preparation
1. Use clean gloves and try to work as clean as possible
2. Use filter tips and low-binding tubes (Costar)
3. Buffer stocks should be filtered (0.2 µm filter) prior storage
4. Don´t use old buffers (discard older than 3 month)
5. Freshly add PIC tablets and PMSF
6. Work on ice and use cooled centrifuges

Buffers

RIPA Buffer
- 25 mM HEPES pH 7.6
- 150 mM NaCl
- 1 mM EDTA
- 1% Triton X-100
- 0.1% sodium deoxycholate
- 0.1% SDS
- freshly add: PIC (2 tablets in 100ml), PMSF (200x stock)

LiCl wash buffer
- 250 mM LiCl
- 10 mM Tris-HCl PH 8.0
- 1 mM EDTA
- 0.5% NP-40
- 0.5% sodium deoxycholate
- freshly add: PIC (2 tablets in 100ml), PMSF (200x stock)

TE buffer
- 10 mM Tris-HCl pH 8.0

1 mM EDTA

Beads equilibration
1. Always spin beads 500G 1 min and wash with at least 750 µl buffer
2. Re-suspend Sepharose protein G beads (GE Healthcare, Product code: 17-0618-01) in RIPA (25 µl/IP)
3. Wash 2x with RIPA
4. Incubate for at least 1h in RIPA + 1µg/µl salmon sperm DNA + 1 µg/µl BSA (freshly added)
5. Wash 3x 5 min with RIPA

IP
1. Thaw chromatin on ice (10-15 µg / IP)
2. Dilute chromatin to 650 µl with RIPA
3. Pre-absorb chromatin with equilibrated beads for at least 2h
4. Spin down the beads, use the supernatant
5. Add 2 µl anti-GFP antibody (goat, lab stock) to the pre-absorbed chromatin and rotate overnight at 4°C
6. Take 50 µl as input, add 500 µl chromatin to freshly equilibrated beads (25 µl slurry/ IP)
7. Rotate 2-3 h at 4 °C
8. Wash at least 3x 5 min with RIPA, 1x 10 min LiCl wash buffer and rinse 1x with TE buffer
9. Re-suspend beads and input in TE (final volume 100 µl)

Crosslink reversal, DNA purification
1. Incubate overnight at 65°C (shaking 1000 rpm)
2. Add 1 µl RNaseA (10 mg/ml), incubate 30 min at 37°C
3. Add 5µl 10% SDS and 1µl proteaseK (10mg/ml), incubate 1.5 h at 55°C
4. Spin down the beads, use the supernatant
5. Purify DNA using Minelute columns (Qiagen; room temperature)
   5.1 Add 500 µl PB (binding buffer) and mix
   5.2 Apply sample (600µl) to the column and spin for 1 min at room temperature
   5.3 Discard the flow-through
   5.4 Wash the column **2x** with 500 µl PE (wash buffer)
   5.5 Discard the flow-through
   5.6 Centrifuge for an additional 2 min
   5.7 Place the column in a clean low-binding 1.5ml tube
   5.8 Let it stand on the bench with open lid for 1 min (let the ethanol evaporate)
   5.9 Elute with 15µl EB (elution buffer), let it stand on the bench for 5 min, spin at 10000rpm (not max speed) for 2 minutes (caps can break)
6. Store purified DNA at -20 °C (avoid freezing thawing)
7. Check ChIP enrichment with qPCR (purified DNA can be diluted up to 20x)

9.2 TRAP protocol

9.2.1 TRAP procedure

9.2.1.1 Fly collection

1. Collect GFP-L10A flies in polypropylene tubes, freeze in liquid nitrogen and store at -80°C

(1000 flies per lysate: 4-5 ml flies in 15ml tube)

9.2.1.2 Head removal

1. Vortex or hit tubes of flies on the bench top to knock off the heads (keep it cold in liquid nitrogen; 3 rounds of hitting, 5x on each side of the tube)
2. Use a sieve sized 630 µm to separate heads from the bodies and a sieve sized 400 µm to separate the heads from the wings and legs (keep it cold on dry ice)

9.2.1.3 Ribosome extraction

Pre-preparation
1. Use clean gloves and try to work as clean as possible
2. Clean the bench environment with RNaseZAP (Sigma)
3. Use filter tips and low-binding tubes (Costar)
4. Buffer stocks should be filtered (0.2 µm filter) prior storage
5. Don´t use old buffers (discard older than 3 month)
6. Prepare Polysome buffers freshly
7. Work on ice and use cooled centrifuges

Buffers (freshly prepared)
REB (Ribosome Extract Buffer)
10 mM HEPES pH 7.6
150 mM KCl
5 mM MgCl2
protease inhibitor cocktail (2 tablets in 100 ml; PIC, Roche)
0.5 mM DTT
100 µg/ml cycloheximide (CHX)
100 U/ml RNasIn (Promega)

Beads equilibration
1. Always spin beads 500G 1 min and wash with at least 500 µl buffer
2. Re-suspend Sepharose protein G beads (GE Healthcare, Product code: 17-0618-01) in REB + 1% NP-40 (25 µl/IP)

Lysate preparation
1. Dounce 1000 fly heads (1-200 µl in 1.5 ml tube) in 1 ml ice cold REB with a manual mechanical homogenizer (slowly, 20x with loose pestle and 20x with tight pestle)
2. Spin down at 2000G for 10 min and use the supernatant (discard nuclear pellet)
3. Add 100 µl REB + 10% NP-40 (final: 1% NP-40) and 30 µl DHPC (1M; final: 30 mM)
4. Re-suspend detergents by gentle pipetting and incubate for 5 min on ice
5. Spin down at 13000G for 10 min and use the supernatant (discard the debris)

9.2.1.4 Affinity Purification

Buffers (freshly prepared)
RWB (Ribosome Wash Buffer)
10 mM HEPES pH 7.6
350 mM KCl
5 mM MgCl2

1% NP-40
protease inhibitor cocktail (2 tablets in 100 ml; PIC, Roche)
0.5 mM DTT
100 µg/ml cycloheximide (CHX)
100 U/ml RNasIn (Promega)

Immunoprecipitation

1. Immediately use the ~1ml lysate (from 9.2.1.3) for IP
2. Take out 25 µl input (keep on ice) and split the lysate into 4 parts (225 µl; technical replicates)
3. Dilute each part with 250 µl REB + 1% NP-40
4. Add 2 µl anti-GFP antibody to each (goat, lab stock), mix gently and incubate for 15 min on ice
5. Add lysate+antibody to 25 µl equilibrated beads and incubate for 45 min on ice (mix gently every 10 min)
6. Wash beads 3x quickly with RWB
7. Use beads and input for RNA purification

RNA purification

1. Purify beads-bound and input total RNA using RNeasy MinElute kit (Qiagen)
2. Re-suspend beads in 350 µl RLT buffer and incubate for 5 min at room temperature
3. Spin down the beads and use the supernatant
4. Add 250 µl 100% ethanol, mix by pipetting and immediately transfer to the column
5. Spin down for 15 s at 8000G, discard the flow-through and place the column to a new collection tube
6. Wash the column with 500 µl RPE (spin for 15 s at 8000G) and with 80% ethanol (spin for 2 min at 8000G)
7. Place the column to a new collection tube and spin at full speed for 5 min with open lid
8. Place the column to an elution tube (1.5 ml) and elute RNA with 14 µl RNase-free water (spin at 10000G for 2 min)
9. Measure RNA concentration using Qubit RNA Assay Kit
10. Store purified RNA at -80°C (avoid freezing thawing)
11. Use 1-2 µl RNA for cDNA synthesis and check TRAP enrichment with qPCR (cDNA can diluted up to 5x)

10 Bibliography

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of Drosophila melanogaster. Science 287, 2185–2195.

Adelman, K., Marr, M.T., Werner, J., Saunders, A., Ni, Z., Andrulis, E.D., and Lis, J.T. (2005). Efficient Release from Promoter-Proximal Stall Sites Requires Transcript Cleavage Factor TFIIS. Molecular Cell 17, 103–112.

Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., and Lis, J.T. (2006). Drosophila Paf1 modulates chromatin structure at actively transcribed genes. Molecular and Cellular Biology 26, 250–260.

Adryan, B., Woerfel, G., Birch-Machin, I., Gao, S., Quick, M., Meadows, L., Russell, S., and White, R. (2007). Genomic mapping of Suppressor of Hairy-wing binding sites in Drosophila. Genome Biology 8, R167.

Akalal, D.-B.G., Wilson, C.F., Zong, L., Tanaka, N.K., Ito, K., and Davis, R.L. (2006). Roles for Drosophila mushroom body neurons in olfactory learning and memory. Learn. Mem. 13, 659–668.

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Molecular Cell 5, 367–375.

Akhtar, A., and Gasser, S.M. (2007). The nuclear envelope and transcriptional control. Nature Reviews Genetics 8, 507–517.

Alic, N., Andrews, T.D., Giannakou, M.E., Papatheodorou, I., Slack, C., Hoddinott, M.P., Cochemé, H.M., Schuster, E.F., Thornton, J.M., and Partridge, L. (2011). Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. Molecular Systems Biology 7, 502.

ALLFREY, V.G., FAULKNER, R., and MIRSKY, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proc. Natl. Acad. Sci. U.S.a. 51, 786–794.

Anders, S., Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Nature Precedings.

Andretic, R., van Swinderen, B., and Greenspan, R.J. (2005). Dopaminergic modulation of arousal in Drosophila. Curr. Biol. 15, 1165–1175.

Arrese, E.L., and Soulages, J.L. (2010). Insect fat body: energy, metabolism, and regulation. Annu. Rev. Entomol. 55, 207–225.

Avvakumov, N., Nourani, A., and Côté, J. (2011). Histone chaperones: modulators of chromatin marks. Molecular Cell 41, 502–514.

Awasaki, T., Lai, S.L., Ito, K., and Lee, T. (2008). Organization and Postembryonic Development of Glial Cells in the Adult Central Brain of Drosophila. Journal of Neuroscience 28, 13742–13753.

Baek, M., and Mann, R.S. (2009). Lineage and birth date specify motor neuron targeting and dendritic architecture in adult Drosophila. Journal of Neuroscience 29, 6904–6916.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res 21, 381–395.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120–124.

Bargaje, R., Alam, M.P., Patowary, A., Sarkar, M., Ali, T., Gupta, S., Garg, M., Singh, M., Purkanti, R., Scaria, V., et al. (2012). Proximity of H2A.Z containing nucleosome to the transcription start site influences gene expression levels in the mammalian liver and brain. Nucleic Acids Research.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell 129, 823–837.

Batish, M., van den Bogaard, P., Kramer, F.R., and Tyagi, S. (2012). Neuronal mRNAs travel singly into dendrites. Proceedings of the National Academy of Sciences 109, 4645–4650.

Becker, P.B., and Hörz, W. (2002). ATP-dependent nucleosome remodeling. Annu. Rev. Biochem. 71, 247–273.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science 301, 1090–1093.

Bonn, S., and Furlong, E.E.M. (2008). cis-Regulatory networks during development: a view of Drosophila. Current Opinion in Genetics & Development 18, 513–520.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44, 148–156.

Bonn, S., Zinzen, R.P., Perez-Gonzalez, A., Riddell, A., Gavin, A.-C., and Furlong, E.E.M. (2012). Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. Nature Protocols 7, 978–994.

Boynton, S., and Tully, T. (1992). latheo, a new gene involved in associative learning and memory in Drosophila melanogaster, identified from P element mutagenesis. Genetics 131, 655–672.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118, 401–415.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.-C., Widom, J., and Brickner, J.H. (2007). H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. PLoS Biol 5, e81.

Brickner, J.H. (2009). Transcriptional memory at the nuclear periphery. Current Opinion in Cell Biology 21, 127–133.

Brody, T., and Odenwald, W.F. (2002). Cellular diversity in the developing nervous system: a temporal view from Drosophila. Development 129, 3763–3770.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D.J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. The Journal of Cell Biology 141, 469–481.

Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. Molecular Cell 36, 541–546.

Busch, S., Selcho, M., Ito, K., and Tanimoto, H. (2009). A map of octopaminergic neurons in the Drosophilabrain. J. Comp. Neurol. 513, 643–667.

Bushey, D., Tononi, G., and Cirelli, C. (2009). The Drosophila fragile X mental retardation gene regulates sleep need. Journal of Neuroscience 29, 1948–1961.

Campos-Ortega, J.A. (1995). Genetic mechanisms of early neurogenesis in Drosophila melanogaster. Mol Neurobiol 10, 75–89.

Carey, M., Li, B., and Workman, J.L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Molecular Cell 24, 481–487.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P., et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123, 581–592.

Chandy, M., Gutiérrez, J.L., Prochasson, P., and Workman, J.L. (2006). SWI/SNF displaces SAGA-acetylated nucleosomes. Eukaryotic Cell 5, 1738–1747.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. Science 318, 1780–1782.

Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science 299, 721–725.

Chintapalli, V.R., Wang, J., and Dow, J.A.T. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39, 715–720.

Chung, C.Y., Seo, H., Sonntag, K.C., Brooks, A., Lin, L., and Isacson, O. (2005). Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. Hum. Mol. Genet. 14, 1709–1725.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368–373.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. 78, 273–304.

Clarkson, M.J., Wells, J.R., Gibson, F., Saint, R., and Tremethick, D.J. (1999). Regions of variant histone His2AvD required for Drosophila development. Nature 399, 694–697.

Coleman-Derr, D., and Zilberman, D. (2012). Deposition of Histone Variant H2A.Z within Gene Bodies Regulates Responsive Genes. PLoS Genet 8, e1002988.

Coleman, C.M., and Neckameyer, W.S. (2005). Serotonin synthesis by two distinct enzymes in Drosophila melanogaster. Arch. Insect Biochem. Physiol. 59, 12–31.

Comas, D., Petit, F., and Preat, T. (2004). Drosophila long-term memory formation involves regulation of cathepsin activity. Nature 430, 460–463.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science 322, 1845–1848.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the Status of RNA Polymerase at Promoters. Cell Rep.

Costa-Mattioli, M., Sossin, W.S., Klann, E., and Sonenberg, N. (2009). Translational control of long-lasting synaptic plasticity and memory. Neuron 61, 10–26.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863–1876.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288, 640–649.

Crick, F. (1970). Central dogma of molecular biology. Nature 227, 561–563.

Dauwalder, B. (2002). The Drosophila takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. Genes & Development 16, 2879–2892.

Davis, M.B., SanGil, I., Berry, G., Olayokun, R., and Neves, L.H. (2011). Identification of common and cell type specific LXXLL motif EcR cofactors using a bioinformatics refined

candidate RNAi screen in Drosophila melanogaster cell lines. BMC Developmental Biology 11, 66.

Davis, R.L., and Giurfa, M. (2012). Mushroom-body memories: an obituary prematurely written? Curr. Biol. 22, R272–R275.

De Ferrari, L., and Aitken, S. (2006). Mining housekeeping genes with a Naive Bayes classifier. BMC Genomics 7, 277.

Deal, R.B., and Henikoff, S. (2010). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. Developmental Cell 18, 1030–1040.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science 295, 1306–1311.

Deng, W., and Roberts, S.G.E. (2007). TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma 116, 417–429.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature 399, 491–496.

Dierick, H.A., and Greenspan, R.J. (2007). Serotonin and neuropeptide F have opposite modulatory effects on fly aggression. Nat Genet 39, 678–682.

Dinant, C., Houtsmuller, A.B., and Vermeulen, W. (2008). Chromatin structure and DNA damage repair. Epigenetics & Chromatin 1, 9.

Dion, M.F., Altschuler, S.J., Wu, L.F., and Rando, O.J. (2005). Genomic characterization reveals a simple histone H4 acetylation code. Proc. Natl. Acad. Sci. U.S.a. 102, 5501–5506.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature.

Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Research 38, 4218–4230.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 135, 749–762.

Dubnau, J., Chiang, A.-S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., et al. (2003). The staufen/pumilio pathway is involved in Drosophila long-term memory. Curr. Biol. 13, 286–296.

Ebert, A., Lein, S., Schotta, G., and Reuter, G. (2006). Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res 14, 377–392.

Edwards, T.N., and Meinertzhagen, I.A. (2010). The functional organisation of glia in the adult brain of Drosophila and other insects. Progress in Neurobiology 90, 471–497.

Egger, B., Leemans, R., Loop, T., Kammermeier, L., Fan, Y., Radimerski, T., Strahm, M.C., Certa, U., and Reichert, H. (2002). Gliogenesis in Drosophila: genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. Development 129, 3295–3309.

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends Genet. 24, 280–288.

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet. 28, 333–341.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science 318, 1777–1779.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 1–9.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Molecular Cell 11, 1549–1561.

Farhadian, S.F., Suárez-Fariñas, M., Cho, C.E., Pellegrino, M., and Vosshall, L.B. (2012). Post-fasting olfactory, transcriptional, and feeding responses in Drosophila. Physiol. Behav. 105, 544–553.

Feller, C., Prestel, M., Hartmann, H., Straub, T., Söding, J., and Becker, P.B. (2011). The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. Nucleic Acids Research.

Fichelson, P., Audibert, A., Simon, F., and Gho, M. (2005). Cell cycle and cell-fate determination in Drosophila neural cell lineages. Trends in Genetics 21, 413–420.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell 143, 212–224.

Finley, K.D., Edeen, P.T., Foss, M., Gross, E., Ghbeish, N., Palmer, R.H., Taylor, B.J., and McKeown, M. (1998). Dissatisfaction encodes a tailless-like nuclear receptor expressed in a subset of CNS neurons controlling Drosophila sexual behavior. Neuron 21, 1363–1374.

Fischle, W., Wang, Y., and Allis, C.D. (2003). Binary switches and modification cassettes in histone biology and beyond. Nature 425, 475–479.

Fox, R.M., and Miller, D. (2007). The embryonic muscle transcriptome of Caenorhabditis elegans. Genome Biology 1–20.

Fox, R.M., Stetina, Von, S.E., Barlow, S.J., Shaffer, C., Olszewski, K.L., Moore, J.H., Dupuy, D., Vidal, M., and Miller, D.M. (2005). A gene expression fingerprint of C. elegans embryonic motor neurons. BMC Genomics 6, 42.

Friggi-Grelin, F., Coulom, H., Meller, M., Gomez, D., Hirsh, J., and Birman, S. (2003). Targeted gene expression in Drosophila dopaminergic cells using regulatory sequences from tyrosine hydroxylase. J. Neurobiol. 54, 618–627.

Fu, S., Fan, J., Blanco, J., Gimenez-Cassina, A., Danial, N.N., Watkins, S.M., and Hotamisligil, G.S. (2012). Polysome profiling in liver identifies dynamic regulation of endoplasmic reticulum translatome by obesity and fasting. PLoS Genet 8, e1002902.

Fujikawa, K., Takahashi, A., Nishimura, A., Itoh, M., Takano-Shimizu, T., and Ozaki, M. (2009). Characteristics of genes up-regulated and down-regulated after 24 h starvation in the head of Drosophila. Gene 446, 11–17.

Furlong, E.E., Andersen, E.C., Null, B., White, K.P., and Scott, M.P. (2001). Patterns of gene expression during Drosophila mesoderm development. Science 293, 1629–1633.

Geminard, C., Rulifson, E.J., and LEopold, P. (2009). Remote Control of Insulin Secretion by Fat Cells in Drosophila. Cell Metabolism 10, 199–207.

Gershman, B., Puig, O., Hang, L., Peitzsch, R.M., Tatar, M., and Garofalo, R.S. (2007). High-resolution dynamics of the transcriptional response to nutrition in Drosophila: a key role for dFOXO. Physiol. Genomics 29, 24–34.

Gilbert, S.F. (2000). Developmental Biology.

Gilchrist, D.A., Santos, Dos, G., Fargo, D.C., Bin Xie, Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. Cell 143, 540–551.

Glaus, P., Honkela, A., and Rattray, M. (2012). Identifying differentially expressed transcripts from RNA-seq data with biological variation

Granderath, S., Bunse, I., and Klämbt, C. (2000). gcm and pointed synergistically control glial transcription of the Drosophila gene loco. Mechanisms of Development 91, 197–208.

Green, M.R. (2005). Eukaryotic transcription activation: right on target. Molecular Cell 18, 399–402.

Grewal, S.I.S., and Jia, S. (2007). Heterochromatin revisited. Nature Reviews Genetics 8, 35–46.

Grönke, S., Mildner, A., Fellert, S., Tennagels, N., Petry, S., Müller, G., Jäckle, H., and Kühnlein, R.P. (2005). Brummer lipase is an evolutionary conserved fat storage regulator in Drosophila. Cell Metabolism 1, 323–330.

Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007). Chromatin challenges during DNA replication and repair. Cell 128, 721–733.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88.

Haberland, M., Montgomery, R.L., and Olson, E.N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nature Reviews Genetics 10, 32–42.

Haberman, A., Williamson, W.R., Epstein, D., Wang, D., Rina, S., Meinertzhagen, I.A., and Hiesinger, P.R. (2012). The synaptic vesicle SNARE neuronal Synaptobrevin promotes endolysosomal degradation and prevents neurodegeneration. The Journal of Cell Biology 196, 261–276.

Haenni, S., Ji, Z., Hoque, M., Rust, N., Sharpe, H., Eberhard, R., Browne, C., Hengartner, M.O., Mellor, J., Tian, B., et al. (2012). Analysis of C. elegans intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3'-end-seq. Nucleic Acids Research 40, 6304–6318.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". Proc. Natl. Acad. Sci. U.S.a. 103, 6428–6435.

Halbeisen, R.E., and Gerber, A.P. (2009). Stress-dependent coordination of transcriptome and translatome in yeast. PLoS Biol 7, e1000105.

Halbeisen, R.E., Galgano, A., Scherrer, T., and Gerber, A.P. (2008). Post-transcriptional gene regulation: from genome-wide studies to principles. Cell. Mol. Life Sci. 65, 798–813.

Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. Cell 113, 429–432.

Han, K.A., Millar, N.S., and Davis, R.L. (1998). A novel octopamine receptor with preferential expression in Drosophila mushroom bodies. J. Neurosci. 18, 3650–3658.

Harrison, J.B., Chen, H.H., Sattelle, E., Barker, P.J., Huskisson, N.S., Rauh, J.J., Bai, D., and Sattelle, D.B. (1996). Immunocytochemical mapping of a C-terminus anti-peptide antibody to the GABA receptor subunit, RDL in the nervous system in Drosophila melanogaster. Cell Tissue Res. 284, 269–278.

Hayashi, S., Ito, K., Sado, Y., Taniguchi, M., Akimoto, A., Takeuchi, H., Aigaki, T., Matsuzaki, F., Nakagoshi, H., Tanimura, T., et al. (2002). GETDB, a database compiling expression patterns and molecular locations of a collection of gal4 enhancer traps. Genesis 34, 58–61.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types. Cell 135, 738–748.

Heitz, E. (1928). Das Heterochromatin der Moose.

Hempel, C.M., Sugino, K., and Nelson, S.B. (2007). A manual method for the purification of fluorescently labeled neurons from the mammalian brain. Nature Protocols 2, 2924–2929.

Henikoff, S. (2009). Labile H3.3+H2A.Z nucleosomes mark “nucleosome-free regions.” Nat Genet 41, 865–866.

Henry, G.L., Davis, F.P., Picard, S., and Eddy, S.R. (2012). Cell type-specific genomics of Drosophila neurons. Nucleic Acids Research.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal Domain of RNA Polymerase II Plays Central Roles in the Integrated Events of Eucaryotic Gene Expression. Journal of Biochemistry 141, 601–608.

Hochedlinger, K., and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. Development 136, 509–523.

Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S., et al. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. Nature.

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H., et al. (2011). Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Research 21, 182–192.

Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science 334, 683–686.

Iijima, K., Zhao, L., Shenton, C., and Iijima-Ando, K. (2009). Regulation of Energy Stores and Feeding by Neuronal and Peripheral CREB Activity in Drosophila. PLoS ONE 4, e8498.

Ito, K., Awano, W., Suzuki, K., Hiromi, Y., and Yamamoto, D. (1997). The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. Development 124, 761–771.

Jackson, D.A., Iborra, F.J., Manders, E.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. Mol. Biol. Cell 9, 1523–1536.

Jackson, F.R., and Haydon, P.G. (2008). Glial cell regulation of neurotransmission and behavior in Drosophila. Ngb 4, 11.

Jacobs, J.R. (2000). The midline glia of Drosophila: a molecular genetic model for the developmental functions of glia. Progress in Neurobiology 62, 475–508.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science 295, 2080–2083.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science 288, 1422–1425.

James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A., and Elgin, S.C. (1989). Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. Eur. J. Cell Biol. 50, 170–180.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science 293, 1074–1080.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature Reviews Genetics 10, 161–172.

Jiang, Y., Matevossian, A., Huang, H.-S., Straubhaar, J., and Akbarian, S. (2008). Isolation of neuronal chromatin from brain tissue. BMC Neurosci 9, 1–9.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant–containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. Nat Genet 41, 941–945.

Jones, B.W. (2005). Transcriptional control of glial cell development in Drosophila. Developmental Biology 278, 265–273.

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Molecular Cell 20, 971–978.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental Biology 339, 225–229.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science 301, 1096–1099.

Karsten, S.L., Van Deerlin, V.M.D., Sabatti, C., Gill, L.H., and Geschwind, D.H. (2002). An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. Nucleic Acids Research 30, E4.

Kasai, Y., Nambu, J.R., Lieberman, P.M., and Crews, S.T. (1992). Dorsal-ventral patterning in Drosophila: DNA binding of snail protein to the single-minded gene. Proc. Natl. Acad. Sci. U.S.a. 89, 3414–3418.

Kasten, M., Szerlong, H., Erdjument-Bromage, H., Tempst, P., Werner, M., and Cairns, B.R. (2004). Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. The EMBO Journal 23, 1348–1359.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2010). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471, 480–485.

Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science 288, 1418–1422.

Klaes, A., Menne, T., Stollewerk, A., Scholz, H., and Klämbt, C. (1994). The Ets transcription factors encoded by the Drosophila gene pointed direct glial cell differentiation in the embryonic CNS. Cell 78, 149–160.

Knapek, S., Sigrist, S., and Tanimoto, H. (2011). Bruchpilot, a synaptic active zone protein for anesthesia-resistant memory. Journal of Neuroscience 31, 3453–3458.

Kolodziej, P.A., and Young, R.A. (1991). Mutations in the three largest subunits of yeast RNA polymerase II that affect enzyme assembly. Molecular and Cellular Biology 11, 4669–4678.

Kolodziejczyk, A., Sun, X., Meinertzhagen, I.A., and Nässel, D.R. (2008). Glutamate, GABA and acetylcholine signaling components in the lamina of the Drosophila visual system. PLoS ONE 3, e2110.

Kotova, E., Lodhi, N., Jarnik, M., Pinnola, A.D., Ji, Y., and Tulin, A.V. (2011). Drosophila histone H2A variant (H2Av) controls poly(ADP-ribose) polymerase 1 (PARP1) activation in chromatin. Proceedings of the National Academy of Sciences 108, 6205–6210.

Koushika, S.P., Soller, M., and White, K. (2000). The neuron-enriched splicing pattern of Drosophila erect wing is dependent on the presence of ELAV protein. Molecular and Cellular Biology 20, 1836–1845.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell 128, 693–705.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., et al. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. Molecular and Cellular Biology 23, 4207–4218.

Ku, M., Jaffe, J.D., Koche, R.P., Rheinbay, E., Endoh, M., Koseki, H., Carr, S.A., and Bernstein, B.E. (2012). H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. Genome Biology 13, R85.

Kusch, T. (2004). Acetylation by Tip60 Is Required for Selective Histone Variant Exchange at DNA Lesions. Science 306, 2084–2087.

Ladurner, A.G., Inouye, C., Jain, R., and Tjian, R. (2003). Bromodomains mediate an acetyl-histone encoded antisilencing function at heterochromatin boundaries. Molecular Cell 11, 365–376.

Lam, K.C., Mühlpfordt, F., Vaquerizas, J.M., Raja, S.J., Holz, H., Luscombe, N.M., Manke, T., and Akhtar, A. (2012). The NSL complex regulates housekeeping genes in Drosophila. PLoS Genet 8, e1002736.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Larson, J.L., and Yuan, G.-C. (2010). Epigenetic domains found in mouse embryonic stem cells via a hidden Markov model. BMC Bioinformatics 11, 557.

Latchman, D.S. (1997). Transcription factors: an overview. Int. J. Biochem. Cell Biol. 29, 1305–1312.

Lazareva, A.A., Roman, G., Mattox, W., Hardin, P.E., and Dauwalder, B. (2007). A role for the adult fat body in Drosophila male courtship behavior. PLoS Genet 3, e16.

Leach, T.J. (2000). Histone H2A.Z Is Widely but Nonrandomly Distributed in Chromosomes of Drosophila melanogaster. Journal of Biological Chemistry 275, 23267–23272.

Lee, K.-S., Hong, S.-H., Kim, A.-K., Ju, S.-K., Kwon, O.-Y., and Yu, K. (2009). Processed short neuropeptide F peptides regulate growth through the ERK-insulin pathway in Drosophila melanogaster. FEBS Letters 583, 2573–2577.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat Rev Mol Cell Biol 8, 284–295.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298, 799–804.

Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Reviews Genetics 13, 233–245.

Lercher, M.J., Urrutia, A.O., and Hurst, L.D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 31, 180–183.

Lewis, B.A., Sims, R.J., Lane, W.S., and Reinberg, D. (2005). Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. Molecular Cell 18, 471–481.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell 128, 707–719.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Current Opinion in Genetics & Development 21, 175–186.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of

long-range interactions reveals folding principles of the human genome. Science 326, 289–293.

Lisbin, M.J., Gordon, M., Yannoni, Y.M., and White, K. (2000). Function of RRM domains of Drosophila melanogaster ELAV: Rnp1 mutations and rrm domain replacements with ELAV family proteins and SXL. Genetics 155, 1789–1798.

Liu, T., Dartevelle, L., Yuan, C., Wei, H., Wang, Y., Ferveur, J.-F., and Guo, A. (2008). Increased dopamine level enhances male-male courtship in Drosophila. Journal of Neuroscience 28, 5539–5546.

Lobo, M.K., Karsten, S.L., Gray, M., Geschwind, D.H., and Yang, X.W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. Nat Neurosci 9, 443–452.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260.

Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., Mclaren, P., North, P., et al. (2007). FlyMine: an integrated database for Drosophila and Anopheles genomics. Genome Biology 8, R129.

Maeda, R.K., and Karch, F. (2007). Making connections: boundaries and insulators in Drosophila. Current Opinion in Genetics & Development 17, 394–399.

Marsh, E.D., Minarcik, J., Campbell, K., Brooks-Kayal, A.R., and Golden, J.A. (2008). FACS-array gene expression analysis during early development of mouse telencephalic interneurons. Dev Neurobiol 68, 434–445.

Martín-Bermudo, M.D., Carmena, A., and Jiménez, F. (1995). Neurogenic genes control gene expression at the transcriptional level in early neurogenesis and in mesectoderm specification. Development 121, 219–224.

Martin, J.R., Raabe, T., and Heisenberg, M. (1999). Central complex substructures are required for the maintenance of locomotor activity in Drosophila melanogaster. J. Comp. Physiol. A 185, 277–288.

Matsuno, M., Horiuchi, J., Tully, T., and Saitoe, M. (2009). The Drosophila cell adhesion molecule Klingon is required for long-term memory formation and is regulated by Notch. Proceedings of the National Academy of Sciences 106, 310–315.

Maurange, C., Cheng, L., and Gould, A.P. (2008). Temporal transcription factors and their targets schedule the end of neural proliferation in Drosophila. Cell 133, 891–902.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature 453, 358–362.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science 336, 1723–1725.

McGuire, S.E., Le, P.T., and Davis, R.L. (2001). The role of Drosophila mushroom body signaling in olfactory memory. Science 293, 1330–1333.

Meunier, N., Belgacem, Y.H., and Martin, J.R. (2007). Regulation of feeding behaviour and locomotor activity by takeout in Drosophila. Journal of Experimental Biology 210, 1424–1434.

Millar, C.B., Xu, F., Zhang, K., and Grunstein, M. (2006). Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. Genes & Development 20, 711–722.

Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet 39, 1507–1511.

Nagoshi, E., Sugino, K., Kula, E., Okazaki, E., Tachibana, T., Nelson, S., and Rosbash, M. (2009). Dissecting differential gene expression within the circadian neuronal circuit of Drosophila. Nat Neurosci 13, 60–68.

Nässel, D.R., Enell, L.E., Santos, J.G., Wegener, C., and Johard, H.A.D. (2008). A large population of diverse neurons in the Drosophila central nervous system expresses short neuropeptide F, suggesting multiple distributed peptide functions. BMC Neurosci 9, 90.

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. BBA - Gene Regulatory Mechanisms 1809, 34–45.

Nechaev, S., Fargo, D.C., Santos, dos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. Science 327, 335–338.

Neel, V.A., and Young, M.W. (1994). Igloo, a GAP-43-related gene expressed in the developing nervous system of Drosophila. Development 120, 2235–2243.

Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.H., et al. (2010). A Comprehensive Map of Insulator Elements for the Drosophila Genome. PLoS Genet 6, e1000814.

Ng, H.-H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular Cell 11, 709–719.

Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., Ohler, U., and Zhu, J. (2010). A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nature Methods 1–10.

Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25, 2730–2731.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385.

Nozaki, T., Yachie, N., Ogawa, R., Kratz, A., Saito, R., and Tomita, M. (2011). Tight associations between transcription promoter type and epigenetic variation in histone positioning and modification. BMC Genomics 12, 416.

Nüsslein-Volhard, C., Frohnhöfer, H.G., and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. Science 238, 1675–1681.

Okaty, B.W., Sugino, K., and Nelson, S.B. (2011). A Quantitative Comparison of Cell-Type-Specific Microarray Gene Expression Profiling Methods in the Mouse Brain. PLoS ONE 6, e16493.

Orlicky, S.M., Tran, P.T., Sayre, M.H., and Edwards, A.M. (2001). Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. J. Biol. Chem. 276, 10097–10102.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36, 1065–1071.

Pandey, R., and Dou, Y. (2013). H2A.Z Sets the Stage in ESCs. Cell Stem Cell 12, 143–144.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes & Development 20, 2041–2054.

Paredes, J.A., Carreto, L., Simões, J., Bezerra, A.R., Gomes, A.C., Santamaria, R., Kapushesky, M., Moura, G.R., and Santos, M.A.S. (2012). Low level genome mistranslations deregulate the transcriptome and translatome and generate proteotoxic stress in yeast. BMC Biol 10, 55.

Pearson, B.J., and Doe, C.Q. (2003). Regulation of neuroblast competence in Drosophila. Nature 425, 624–628.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Molecular Cell 23, 297–305.

Petesch, S.J., and Lis, J.T. (2008). Rapid, Transcription-Independent Loss of Nucleosomes over a Large Chromatin Domain at Hsp70 Loci. Cell 134, 74–84.

Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.T.B., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R., et al. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. Proceedings of the National Academy of Sciences 105, 9715–9720.

Pichaud, F., and Desplan, C. (2001). A new visualization approach for identifying mutations that affect differentiation and organization of the Drosophila ommatidia. Development 128, 815–826.

Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M., and van Steensel, B. (2006). Characterization of the Drosophila melanogaster genome at the nuclear lamina. Nat Genet 38, 1005–1014.

Pile, L.A., and Wassarman, D.A. (2002). Localizing transcription factors on chromatin by immunofluorescence. 1–7.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517–527.

Pospisilik, J.A., Schramek, D., Schnidar, H., Cronin, S.J.F., Nehme, N.T., Zhang, X., Knauf, C., Cani, P.D., Aumayr, K., Todoric, J., et al. (2010). Drosophila Genome-wide Obesity Screen Reveals Hedgehog as a Determinant of Brown versus White Adipose Cell Fate. Cell 140, 148–160.

Puig, O., Marr, M.T., Ruhf, M.L., and Tjian, R. (2003). Control of cell number by Drosophila FOXO: downstream and feedback regulation of the insulin receptor pathway. Genes & Development 17, 2006–2020.

Qiu, S., Luo, S., Evgrafov, O., Li, R., Schroth, G.P., Levitt, P., Knowles, J.A., and Wang, K. (2012). Single-neuron RNA-Seq: technical feasibility and reproducibility. Front Genet 3, 124.

Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011a). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. PLoS Genet 7, e1001274.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell 123, 233–248.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature Biotechnology 30, 777–782.

Rao, B., Shibata, Y., Strahl, B.D., and Lieb, J.D. (2005). Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. Molecular and Cellular Biology 25, 9447–9459.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. U.S.a. 90, 7923–7927.

Riechmann, V., Rehorn, K.P., Reuter, R., and Leptin, M. (1998). The genetic control of the distinction between fat body and gonadal mesoderm in Drosophila. Development 125, 713–723.

Rival, T., Soustelle, L., Strambi, C., Besson, M.-T., Iché, M., and Birman, S. (2004). Decreasing Glutamate Buffering Capacity Triggers Oxidative Stress and Neuropil Degeneration in the Drosophila Brain. Current Biology 14, 599–605.

Robinow, S., and White, K. (1988). The locus elav of Drosophila melanogaster is expressed in neurons at all developmental stages. Developmental Biology 126, 294–303.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology 29, 24–26.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rossner, M.J., Hirrlinger, J., Wichert, S.P., Boehm, C., Newrzella, D., Hiemisch, H., Eisenhardt, G., Stuenkel, C., Ahsen, von, O., and Nave, K.-A. (2006). Global transcriptome analysis of genetically identified neurons in the adult cortex. Journal of Neuroscience 26, 9956–9966.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature 418, 975–979.

Rulifson, E.J., Kim, S.K., and Nusse, R. (2002). Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. Science 296, 1118–1120.

Sajan, S.A., and Hawkins, R.D. (2012). Methods for identifying higher-order chromatin structure. Annu Rev Genomics Hum Genet 13, 59–82.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246–3248.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E.M. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes & Development 21, 436–449.

Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., and Amieux, P.S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proceedings of the National Academy of Sciences 106, 13939–13944.

Sarcinella, E., Zuzarte, P.C., Lau, P.N.I., Draker, R., and Cheung, P. (2007). Monoubiquitylation of H2A.Z Distinguishes Its Association with Euchromatin or Facultative Heterochromatin. Molecular and Cellular Biology 27, 6457–6468.

Sarov-Blat, L., So, W.V., Liu, L., and Rosbash, M. (2000). The Drosophila takeout gene is a novel molecular link between circadian rhythms and feeding behavior. Cell 101, 647–656.

Saurin, A.J., Shiels, C., Williamson, J., Satijn, D.P., Otte, A.P., Sheer, D., and Freemont, P.S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. The Journal of Cell Biology 142, 887–898.

Schmidt-Ott, U., and Technau, G.M. (1992). Expression of en and wg in the embryonic head and brain of Drosophila indicates a refolded band of seven segment remnants. Development 116, 111–125.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell 132, 887–898.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. The EMBO Journal 21, 1121–1131.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol 7, e13.

Schwabe, T., Bainton, R.J., Fetter, R.D., Heberlein, U., and Gaul, U. (2005). GPCR signaling is required for blood-brain barrier formation in drosophila. Cell 123, 133–144.

Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.-B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A., et al. (2012). Nature and function of insulator protein binding sites in the Drosophila genome. Genome Research.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annu. Rev. Biochem. 79, 271–293.

Sepp, K. (2001). Peripheral Glia Direct Axon Guidance across the CNS/PNS Transition Zone. Developmental Biology 238, 47–63.

Sexton, T., Schober, H., Fraser, P., and Gasser, S.M. (2007). Gene regulation through nuclear organization. Nature Structural & Molecular Biology 14, 1049–1055.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell 1–15.

Shandilya, J., and Roberts, S.G.E. (2012). The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. BBA - Gene Regulatory Mechanisms 1819, 391–400.

Shankaranarayanan, P., Mendoza-Parra, M.-A., Walia, M., Wang, L., Li, N., Trindade, L.M., and Gronemeyer, H. (2011). Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. Nature Methods 8, 565–567.

Shinomiya, K., Matsuda, K., Oishi, T., Otsuna, H., and Ito, K. (2011). Flybrain neuron database: a comprehensive database system of the Drosophila brain neurons. J. Comp. Neurol. 519, 807–833.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10, 697–708.

Skeath, J.B., Panganiban, G.F., and Carroll, S.B. (1994). The ventral nervous system defective gene controls proneural gene expression at two distinct steps during neuroblast formation in Drosophila. Development 120, 1517–1524.

Soller, M., and White, K. (2003). ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA. Genes & Development 17, 2526–2538.

Sousa-Nunes, R., Yee, L.L., and Gould, A.P. (2011). Fat cells reactivate quiescent neuroblasts via TOR and glial insulin relays in Drosophila. Nature 471, 508–512.

Southall, T.D., and Brand, A.H. (2009). Neural stem cell transcriptional networks highlight genes essential for nervous system development. The EMBO Journal 28, 3799–3807.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer 6, 846–856.

Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J., et al. (2011). A spatial and temporal map of C. elegans gene expression. Genome Research 21, 325–341.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics 13, 613–626.

Srinivasan, A., and Mishra, R.K. (2012). Chromatin domain boundary element search tool for Drosophila. Nucleic Acids Research.

Stegmaier, P., Kel, A.E., and Wingender, E. (2004). Systematic DNA-binding domain classification of transcription factors. Genome Inform 15, 276–286.

Steiner, F.A., Talbert, P.B., Kasinathan, S., Deal, R.B., and Henikoff, S. (2012). Cell type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Research.

Stetina, Von, S.E., Watson, J.D., Fox, R.M., Olszewski, K.L., Spencer, W.C., Roy, P.J., and Miller, D.M. (2007). Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C. elegans nervous system. Genome Biology 8, R135.

Stork, T., Engelen, D., Krudewig, A., Silies, M., Bainton, R.J., and Klämbt, C. (2008). Organization and function of the blood-brain barrier in Drosophila. Journal of Neuroscience 28, 587–597.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature 403, 41–45.

Straub, T., and Becker, P.B. (2011). Transcription modulation chromosome-wide: universal features and principles of dosage compensation in worms and flies. Current Opinion in Genetics & Development 21, 147–153.

Suh, J., and Jackson, F.R. (2007). Drosophila ebony activity is required in glia for the circadian regulation of locomotor activity. Neuron 55, 435–447.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321, 956–960.

Swaminathan, J. (2005). The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin. Genes & Development 19, 65–76.

Swarup, S., Williams, T.I., and Anholt, R.R.H. (2011). Functional dissection of Odorant binding protein genes in Drosophila melanogaster. Genes, Brain and Behavior 10, 648–657.

Sweeney, S.T., Broadie, K., Keane, J., Niemann, H., and O'Kane, C.J. (1995). Targeted expression of tetanus toxin light chain in Drosophila specifically eliminates synaptic transmission and causes behavioral defects. Neuron 14, 341–351.

Swinton (2009). Venn diagrams in R with the Vennerable package. 1–33.

Taddei, A., Hediger, F., Neumann, F.R., and Gasser, S.M. (2004). The function of nuclear architecture: a genetic approach. Annu. Rev. Genet. 38, 305–345.

Talbert, P.B., and Henikoff, S. (2010). Histone variants — ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11, 264–275.

Tanabe, M., Kouzmenko, A.P., Ito, S., Sawatsubashi, S., Suzuki, E., Fujiyama, S., Yamagata, K., Zhao, Y., Kimura, S., Ueda, T., et al. (2008). Activation of facultatively silenced Drosophilaloci associates with increased acetylation of histone H2AvD. Genes to Cells 13, 1279–1288.

Tanaka, N.K., Tanimoto, H., and Ito, K. (2008). Neuronal assemblies of theDrosophila mushroom body. J. Comp. Neurol. 508, 711–755.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods 6, 377–382.

Tang, F., Lao, K., and Surani, M.A. (2011). Development and applications of single-cell transcriptome analysis. Nature Methods 8, S6–S11.

Tariq, M.A., Kim, H.J., Jejelowo, O., and Pourmand, N. (2011). Whole-transcriptome RNAseq analysis from minute amount of total RNA. Nucleic Acids Research 39, e120–e120.

Tebaldi, T., Re, A., Viero, G., Pegoretti, I., Passerini, A., Blanzieri, E., and Quattrone, A. (2012). Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. BMC Genomics 13, 220.

Teleman, A.A. (2010). Molecular mechanisms of metabolic regulation by insulin in Drosophila. Biochem. J. 425, 13–26.

Terzi, N., Churchman, L.S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011). H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway. Molecular and Cellular Biology 31, 3569–3583.

The modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., et al. (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science 330, 1787–1797.

Thomas, A., Lee, P.-J., Dalton, J.E., Nomie, K.J., Stoica, L., Costa-Mattioli, M., Chang, P., Nuzhdin, S., Arbeitman, M.N., and Dierick, H.A. (2012). A Versatile Method for Cell-Specific Profiling of Translated mRNAs in Drosophila. PLoS ONE 7, e40276.

Thomas, M.C., and Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. Critical Reviews in Biochemistry and Molecular Biology 41, 105–178.

Tran, E.J., and Wente, S.R. (2006). Dynamic nuclear pore complexes: life on the edge. Cell 125, 1041–1053.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7, 562–578.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 1–8.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 1–8.

Truman, J.W. (1990). Metamorphosis of the central nervous system of Drosophila. J. Neurobiol. 21, 1072–1084.

Tryon, R.C., Pisat, N., Johnson, S.L., and Dougherty, J.D. (2012). Development of translating ribosome affinity purification for zebrafish. Genesis.

Valdés-Mora, F., Song, J.Z., Statham, A.L., Strbenac, D., Robinson, M.D., Nair, S.S., Patterson, K.I., Tremethick, D.J., Stirzaker, C., and Clark, S.J. (2012). Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome Research 22, 307–321.

Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J., and Corces, V. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Research 1–40.

Van Deerlin, V.M.D., Gill, L.H., and Nelson, P.T. (2002). Optimizing gene expression analysis in archival brain tissue. Neurochem. Res. 27, 993–1003.

Varga-Weisz, P.D., and Becker, P.B. (2006). Regulation of higher-order chromatin structures by nucleosome-remodelling factors. Current Opinion in Genetics & Development 16, 151–156.

Wagh, D.A., Rasse, T.M., Asan, E., Hofbauer, A., Schwenkert, I., Dürrbeck, H., Buchner, S., Dabauvalle, M.-C., Schmidt, M., and Qin, G. (2006). Bruchpilot, a Protein with Homology to ELKS/CAST, Is Required for Structural Integrity and Function of Synaptic Active Zones in Drosophila. Neuron 49, 833–844.

Wang, G.G., Allis, C.D., and Chi, P. (2007). Chromatin remodeling and cancer, Part I: Covalent histone modifications. Trends Mol Med 13, 363–372.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40, 897–903.

Weake, V.M., and Workman, J.L. (2012). SAGA function in tissue-specific gene expression. Trends Cell Biol. 22, 177–184.

Weake, V.M., Dyer, J.O., Seidel, C., Box, A., Swanson, S.K., Peak, A., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2011). Post-transcription initiation function of the ubiquitous SAGA complex in tissue-specific gene activation. Genes & Development 25, 1499–1509.

Weber, C.C., and Hurst, L.D. (2011). Support for multiple classes of local expression clusters in Drosophila melanogaster, but no evidence for gene order conservation. Genome Biology 12, R23.

Weber, C.M., Henikoff, J.G., and Henikoff, S. (2010). H2A.Z nucleosomes enriched over active genes are homotypic. Nature Structural & Molecular Biology 17, 1500–1507.

Whittle, C.M., McClinic, K.N., Ercan, S., Zhang, X., Green, R.D., Kelly, W.G., and Lieb, J.D. (2008). The Genomic Distribution and Function of Histone Variant HTZ-1 during C. elegans Embryogenesis. PLoS Genet 4, e1000187.

Wilczyński, B., and Furlong, E.E.M. (2010). Dynamic CRM occupancy reflects a temporal map of developmental progression. Molecular Systems Biology 6, 383.

Wu, C.-L., Xia, S., Fu, T.-F., Wang, H., Chen, Y.-H., Leong, D., Chiang, A.-S., and Tully, T. (2007). Specific requirement of NMDA receptors for long-term memory consolidation in Drosophila ellipsoid body. Nat Neurosci 10, 1578–1586.

Wu, J.S., and Luo, L. (2006). A protocol for dissecting Drosophila melanogaster brains for live imaging or immunostaining. Nature Protocols 1, 2110–2115.

Xu, H., Handoko, L., Wei, X., Ye, C., Sheng, J., Wei, C.-L., Lin, F., and Sung, W.-K. (2010). A signal-noise model for significance analysis of ChIP-seq with negative control. Bioinformatics 26, 1199–1204.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 97, 41–51.

Yang, Z. (2005). Isolation of mRNA from specific tissues of Drosophila by mRNA tagging. Nucleic Acids Research 33, e148–e148.

Yao, J., Munson, K.M., Webb, W.W., and Lis, J.T. (2006). Dynamics of heat shock factor association with native gene loci in living cells. Nature 442, 1050–1053.

Younossi-Hartenstein, A., Nassif, C., Green, P., and Hartenstein, V. (1996). Early neurogenesis of the Drosophila brain. J. Comp. Neurol. 370, 313–329.

Yuasa, Y., Okabe, M., Yoshikawa, S., Tabuchi, K., Xiong, W.-C., Hiromi, Y., and Okano, H. (2003). Drosophila homeodomain protein REPO controls glial differentiation by cooperating with ETS and BTB transcription factors. Development 130, 2419–2428.

Yusuf, D., Butland, S.L., Swanson, M.I., Bolotin, E., Ticoll, A., Cheung, W.A., Zhang, X.Y., Dickman, C.T., Fulton, D.L., Lim, J.S., et al. (2012). The Transcription Factor Encyclopedia. Genome Biology 13, R24.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet 39, 1512–1516.

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Research 34, 6621–6628.

Zhang, Z., and Pugh, B.F. (2011). High-resolution genome-wide mapping of the primary structure of chromatin. Cell 144, 175–186.

Zhao, J., Herrera-Diaz, J., and Gross, D.S. (2005). Domain-Wide Displacement of Histones by Activated Heat Shock Factor Occurs Independently of Swi/Snf and Is Not Correlated with RNA Polymerase II Density. Molecular and Cellular Biology 25, 8985–8999.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456, 125–1