The EMBO Journal (2009) 28, 3799–3807 | © 2009 European Molecular Biology Organization | Some Rights Reserved 0261-4189/09
www.embojournal.org

THE EMBO JOURNAL

EMBO open

Neural stem cell transcriptional networks highlight genes essential for nervous system development

Tony D Southall and Andrea H Brand*

The Gurdon Institute and The Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK

Neural stem cells must strike a balance between self-renewal and multipotency, and differentiation. Identification of the transcriptional networks regulating stem cell division is an essential step in understanding how this balance is achieved. We have shown that the homeodomain transcription factor, Prospero, acts to repress self-renewal and promote differentiation. Among its targets are three neural stem cell transcription factors, Asense, Deadpan and Snail, of which Asense and Deadpan are repressed by Prospero. Here, we identify the targets of these three factors throughout the genome. We find a large overlap in their target genes, and indeed with the targets of Prospero, with 245 genomic loci bound by all factors. Many of the genes have been implicated in vertebrate stem cell self-renewal, suggesting that this core set of genes is crucial in the switch between self-renewal and differentiation. We also show that multiply bound loci are enriched for genes previously linked to nervous system phenotypes, thereby providing a shortcut to identifying genes important for nervous system development.

The EMBO Journal (2009) **28,** 3799–3807. doi:10.1038/emboj.2009.309; Published online 22 October 2009
Subject Categories: chromatin & transcription; development
Keywords: differentiation; multiple transcription factor binding; neural stem cell; self-renewal; transcriptional networks

Introduction

Recent work on *Drosophila* neural stem cells (or neuroblasts) has provided important insights into stem cell biology and tumour formation (Yu *et al*, 2006; Doe, 2008; Egger *et al*, 2008; Zhong and Chia, 2008). Neuroblasts divide in an asymmetric, self-renewing manner producing another neuroblast and a daughter cell that divides only once to give post-mitotic neurons or glial cells (Wodarz, 2005; Egger *et al*, 2008). During these asymmetric divisions the atypical homeodomain transcription factor, Prospero, is asymmetrically segregated to the smaller daughter cell, the ganglion mother cell (GMC), where it can enter the nucleus and regulate transcription (Figure 1B). Neuroblasts lacking Prospero form tumours in both the embryonic nervous system (Choksi *et al*, 2006) and the larval brain (Bello *et al*, 2006; Betschinger *et al*, 2006; Lee *et al*, 2006). Using the chromatin profiling technique DamID (van Steensel and Henikoff, 2000; van Steensel *et al*, 2001), together with expression profiling, we showed that Prospero represses neuroblast genes and is required to activate neuronal differentiation genes (Choksi *et al*, 2006). Therefore, Prospero acts as a binary switch to repress the genetic programs driving self-renewal (by directly repressing neuroblast transcription factors) and to promote differentiation. We find that Prospero represses the neuroblast transcription factors (Choksi *et al*, 2006), Asense, Deadpan and Snail, suggesting that these transcription factors may control genes involved in neural stem cell self-renewal and multipotency.

To identify the transcriptional networks promoting neural stem cell fate, we profiled, on a whole genome scale, the binding sites of Asense, Deadpan and Snail. These three proteins are members of a small group of transcription factors that are expressed in all embryonic neuroblasts (Figure 1A). The first, Asense, is a basic-helix–loop–helix protein, a member of the achaete–scute complex (Gonzalez *et al*, 1989; Jarman *et al*, 1993), and a homologue of the vertebrate neural stem cell factor, Ascl1 (Mash1). Unlike the other members of the achaete–scute complex, Asense is not expressed in proneural clusters in the embryo. Asense expression is initiated in the neuroblast and is maintained in at least a subset of GMC daughter cells (Brand *et al*, 1993). Asense is also expressed in most larval brain neuroblasts but is markedly absent from the DM/PAN neuroblast (Bello *et al*, 2008; Bowman *et al*, 2008). In these lineages, Asense expression is delayed and the daughter cells (secondary neuroblasts) of the Asense-negative DM/PAN neuroblasts undergo multiple cell divisions, expanding the stem cell pool before producing GMCs (Bello *et al*, 2008; Boone and Doe, 2008; Bowman *et al*, 2008). Ectopic expression of Asense limits the division potential of DM/PAN neuroblast progeny (Bowman *et al*, 2008). A study in the optic lobe showed that Asense expression coincides with the upregulation of *dacapo* and cell-cycle exit (Wallace *et al*, 2000). Perhaps in combination, these results suggest that Asense may also have a pro-differentiation role.

The second transcription factor, Deadpan, is a basic-helix–loop–helix protein (Bier *et al*, 1992) related to the vertebrate Hes family of transcription factors. Deadpan is expressed in all neuroblasts and has been shown to promote the proliferation of optic lobe neural stem cells (Wallace *et al*, 2000). Unlike Asense, Deadpan is also expressed in the DM/PAN neuroblasts of the larval brain (Boone and Doe, 2008).

*Corresponding author. Department of Physiology, Development and Neuroscience, University of Cambridge, Wellcome CRUK Gurdon Institute, Tennis Court Road, Cambridge CB2 1QN, UK.
Tel.: +44 1223 334 141; Fax: +44 1223 334 089;
E-mail: ahb@mole.bio.cam.ac.uk

Received: 15 June 2009; accepted: 14 September 2009; published online: 22 October 2009

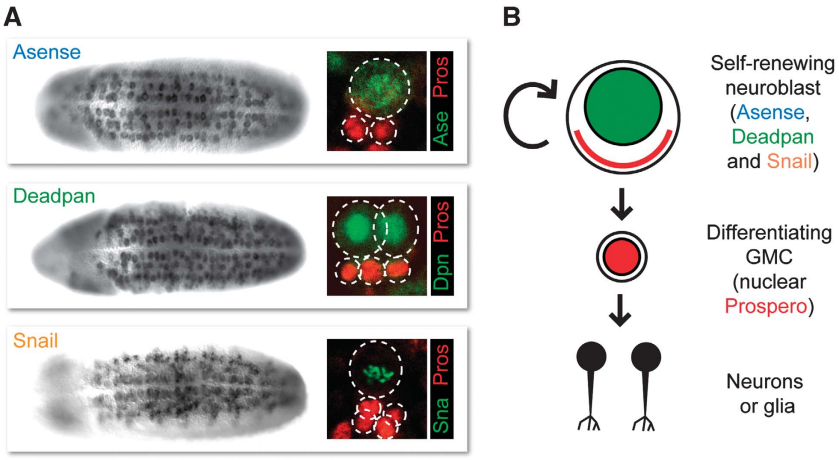

Figure 1 Expression of neural stem cell factors Asense, Deadpan and Snail in the *Drosophila* embryo. (**A**) Expression of *asense*, *deadpan* and *snail* in neuroblasts, left panel shows RNA *in situ* analysis and right panel shows expression of protein compared with Prospero. (**B**) Schematic showing the asymmetric division mode of neuroblasts, in which Asense, Deadpan and Snail are nuclear in neuroblasts and Prospero is segregated to the GMC daughter cell.

The third factor, Snail, is a zinc-finger transcription factor whose vertebrate homologues have roles in the epithelial to mesenchymal transition and in cancer metastasis (Hemavathy *et al*, 2000). The Snail family members (Snail, Worniu and Escargot) are known to regulate neuroblast spindle orientation and cell-cycle progression (Ashraf and Ip, 2001; Cai *et al*, 2001).

To further understand the role of these pan neural stem cell transcription factors, we have mapped their targets throughout the genome. This, combined with expression profiling, allows us to begin to build the gene regulatory networks governing neural stem cell self-renewal, and to enhance our knowledge of the function and mode of action of these transcription factors in neural stem cells.

Results

Asense, Deadpan, Snail and Prospero bind to many common targets

To identify the genes regulated by Asense, Deadpan and Snail in the embryo, we mapped their binding sites *in vivo* by DamID (van Steensel and Henikoff, 2000; van Steensel *et al*, 2001), as we have previously done for Prospero (Choksi *et al*, 2006). In brief, DamID involves tagging a DNA or chromatin-associated protein with a *Escherichia coli* DNA adenine methyltransferase (Dam). Wherever the fusion protein binds, surrounding DNA sequences are methylated. Methylated DNA fragments can then be isolated, labelled and hybridised on a microarray. Here, we express Dam fusion proteins *in vivo*, in transgenic *Drosophila* embryos. Methylated DNA fragments from transgenic embryos expressing Dam alone serve as a reference. Target sites identified by DamID have been shown to match targets identified by chromatin immunoprecipitation (Sun *et al*, 2003; Song *et al*, 2004; Tolhuis *et al*, 2006), by mapping to polytene chromosomes (Bianchi-Frias *et al*, 2004) and by 3D microscopy data (Pickersgill *et al*, 2006; Guelen *et al*, 2008).

In comparing our results for Asense, Deadpan, Snail and Prospero, we observe a high degree of overlap between their targets (Figure 2A and C; Supplementary Table S1). The average overlap for the four factors in pairwise comparisons is 40%, with the highest overlap between Deadpan and Snail (66%). The similarity in binding is illustrated by the binding of all four factors to the intronic regions of the cell-cycle regulation gene *CycE* (Figure 2B). 245 genes are bound by all four proteins, including genes involved in neuroblast cell fate determination, cell-cycle control and differentiation. These loci are unlikely to represent regions of chromatin accessible to all transcription factors as we find only 17/245 (7%) also bound by another neural transcription factor, Pdm1 (Wu and Brand, unpublished). The large overlap in the targets of Asense, Deadpan, Snail and Prospero implies that these may be a core set of genes involved in neuroblast self-renewal and differentiation.

Properties of loci bound by Asense, Deadpan and Snail

Genome-wide analysis of Asense DamID peaks shows that Asense binding is associated with increased levels of DNA conservation (determined by the alignment of eight insect species (Remm *et al*, 2001) (Figure 3B). A representation of Asense binding around a generic gene (Figure 3D) shows an enrichment of ~2 kb upstream of the transcriptional start site, binding within intronic regions (32%) and also downstream of the gene (20%). This distribution is consistent with transcription factor-binding analysis and regulatory sequence studies in mice and humans (Birney *et al*, 2007; Chen *et al*, 2008).

The resolution of DamID is ~1 kb (Vogel *et al*, 2007) and there are currently no motif discovery tools available that can analyse the large amount of sequence data generated by full genome DamID. Therefore, we developed a motif discovery tool, called MICRA (**M**otif **I**dentification using **C**onservation and **R**elative **A**bundance) to identify overrepresented motifs in low-resolution data. In brief, 1 kb of sequence from each binding site is extracted and filtered for conserved sequences. The relative frequency of each 6–10 mer is then calculated and compared with background frequency (see Supplementary data for more details). Using MICRA we identified the E-box, CAGCTG, as the most overrepresented 6 mer in the regions of Asense binding (131% overrepresented using a conservation threshold of 0.6; see Supplementary Figure S6; Figure 3C). In support of our *in vivo* binding data, *in vitro* studies (Jarman *et al*, 1993) had previously shown that

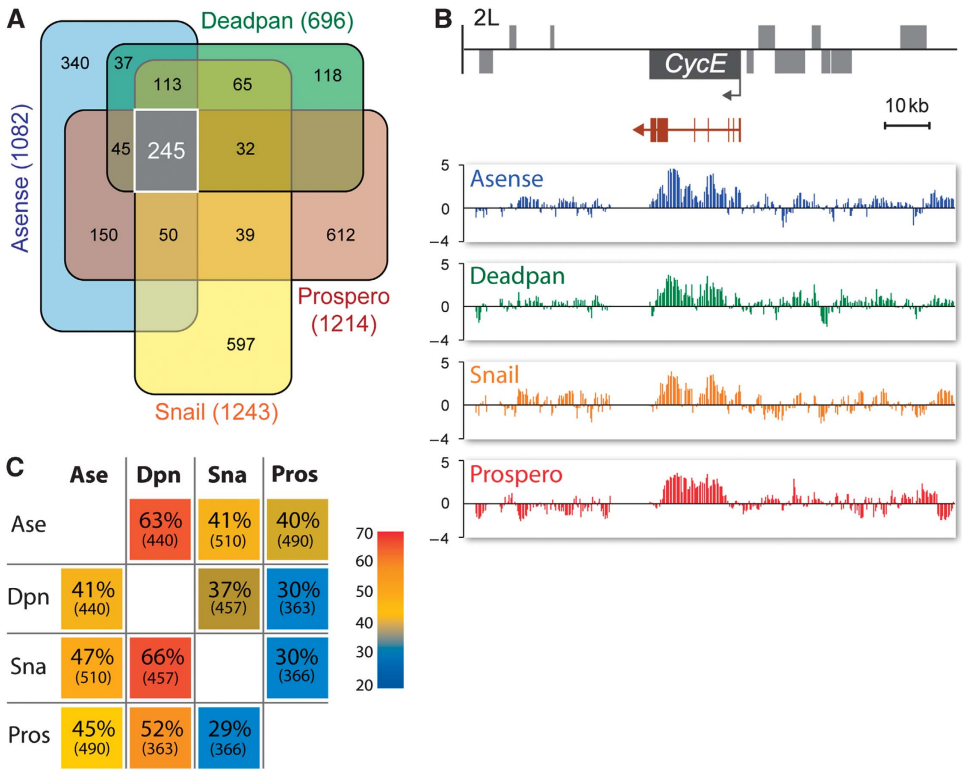

Figure 2 Asense, Deadpan, Snail and Prospero target genes overlap. (**A**) Venn diagram showing shared targets. Numbers in brackets show total number of genes bound. (**B**) Binding of Asense, Deadpan, Snail and Prospero at the *CycE* loci. Grey and black boxes represent genes. Brown bars represent exons. Bar heights are proportional to the average of normalised log2-transformed ratio of intensities from DamID *in vivo* binding site experiments. (**C**) Heat diagram with individual overlap information.

Asense binds to CAGCTG, which is also the binding site of the vertebrate Asense homologue Ascl1 (Mash1) (Castro *et al*, 2006).

A GO annotation analysis (Martin *et al*, 2004) of the genes bound by Asense shows a highly significant overrepresentation of genes involved in nervous system development and cell fate determination (Supplementary Figure S1). Similar analyses were performed for Deadpan and Snail and for both transcription factors; DNA conservation was enriched surrounding their binding sites (Supplementary Figure S5). Deadpan and Snail targets fall broadly into the same gene ontology classes as Asense and Prospero (Choksi *et al*, 2006) (Supplementary Figures S2 and S3**)** and the binding peaks show a similar distribution relative to gene structure as for Asense (Supplementary Figure S4). Motif discovery using MICRA identifies sites consistent with previously published *in vitro* studies for Deadpan (CACGCG and CACGTG) (Winston *et al*, 1999, 2000) and Snail (CAGGTA) (Mauhin *et al*, 1993) (Supplementary Figure S6). These analyses provide unbiased support for the Deadpan and Snail DamID experiments.

Multiple transcription factor binding is associated with increased conservation of regulatory sequences and with genes critical for the development of the nervous system

When comparing our data sets for Asense, Deadpan, Snail and Prospero we find genomic loci in which multiple transcription factors bind. This phenomenon has been described previously in a *Drosophila* cell line (Moorman *et al*, 2006) and, more recently, in mouse embryonic stem (ES) cells (Chen *et al*, 2008) in which these loci are termed 'multiple transcription factor-binding loci' (MTL). The ES cell MTLs are associated with ES-cell-specific gene expression and are thought to identify genes important for stem cell self-renewal. Our data provide an independent and direct, *in vivo* demonstration of the phenomenon described in these two earlier studies. Analysis of neural MTLs (as determined by binding of Asense, Deadpan, Prospero and Snail within a 2 kb window) shows increased sequence constraint, correlating with the number of transcription factors bound (Figure 4A and B). The increase in conservation is higher than expected solely based on the combined binding sites of the factors studied (Figure 4B, orange hashed line, presuming a single binding event that would fully constrain 8 bp of sequence). This suggests that further factors may bind to these loci. The loci associated with MTLs are enriched for genes required for proper neural development (Figure 4C) and for viability (lethal alleles, see Figure 4D).

To investigate further the relationship between the number of transcription factors bound at a locus and the importance of the associated target gene in neural development, we assembled a database (www.neuroBLAST.org) comprising DamID data, expression profiling of neural transcription factors, and data on *Drosophila* nervous system development collated from genetic screens, expression screens, gene homology and text mining screens (see Supplementary data for details of database construction). Using a random permutation algorithm and training sets of known nervous system development genes we assigned weighted scores to each screen (Supplementary Figure S7). A total score is

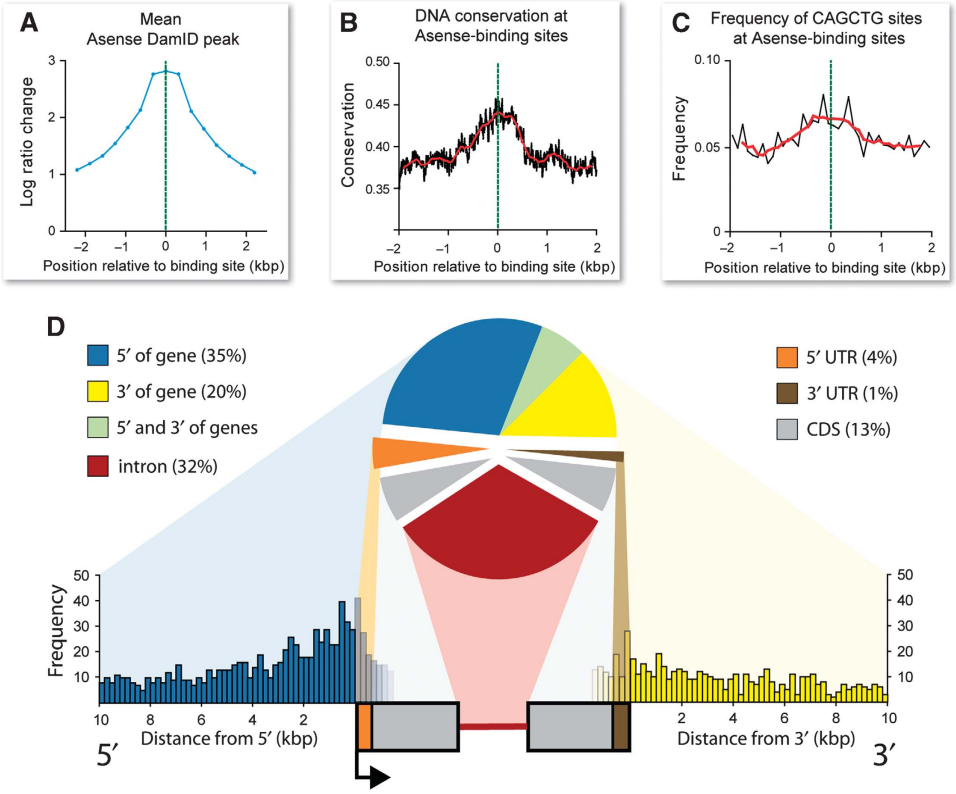

Figure 3 Binding profile of Asense. (**A**) Mean DamID peak as determined from all identified binding sites. The green dashed line represents the peak of enrichment. (**B**) Plot of mean non-exonic conservation (across eight insect species) at sites of Asense binding. Red line is a moving average plot using a window of 100 bp. (**C**) Frequency of predicted Asense-binding motif (CAGCTG) at sites of Asense binding (as determined by DamID). Red line is a moving average plot using a window of 500 bp. (**D**) Distribution of Asense binding. UTR, untranslated region; CDS, coding sequence.

calculated for each gene, providing an indication of the gene's involvement in nervous system development. Multiple gene lists can be searched in the database, which is a useful method to pinpoint key genes in user generated gene lists (e.g. expression array results).

Using the data collected for the database, we consistently find a correlation between gene sets bound by increasing numbers of transcription factors and genes in *Drosophila* genetic screens for defects in nervous system development, eye development and cell-cycle progression ($r = 0.83$, Supplementary Figure S8A) or in text mining screens (occurrence of the gene or its homologue with neural or stem cell terms; $r = 0.98$, Supplementary Figure S8B).

Neural stem cell gene regulatory networks highlight genes crucial for neural development

We have shown that Asense, Deadpan, Prospero and Snail bind to genes essential for neural development. This finding enables us to highlight novel genes that may be involved in neural development. The neuroBLAST database ranks genes based on the number of transcription factors bound, together with their appearance in external screens. In this way it identifies known key players in neural development (Figure 5) such as *prospero*, *brain tumour* (Bello *et al*, 2006; Betschinger *et al*, 2006; Lee *et al*, 2006), *miranda* (Shen *et al*, 1997), *seven up* (Kanai *et al*, 2005) and *glial cells missing* (Jones *et al*, 1995). The majority of these genes are identified by multiple binding information (DamID data), independent of external screens and weighted scores.

Interestingly, there are many high scoring genes that have not previously been characterised for a role in *Drosophila* neural development (a selection are highlighted in Figure 5). These include CG32158, an adenylate cyclase known to be expressed in the CNS, two putative transcription factors (CG2052 and CG33291), an NADH dehydrogenase (CG2014) and an F-box protein (CG9772). There is also cenG1A, an ARF GTPase activator, is bound by all four transcription factors and is expressed in neuroblasts (BDGP expression screen (Tomancak *et al*, 2007)). CG9650 is bound by Prospero and Deadpan, and is a homologue of the BCL11b oncogene, which is essential for proper corticospinal neuron development in vertebrates (Arlotta *et al*, 2005). Another high scoring gene identified by this method is *canoe* (bound by all four transcription factors, neuroBLAST score of 33.7), which has recently been shown to regulate neuroblast asymmetric divisions (Speicher *et al*, 2008).

High-resolution expression profiling suggests a dual role for Asense

Using the binding data for these four transcription factors as a foundation, we sought to construct the transcriptional networks governing neural stem cell self-renewal and differentiation. Although DamID reports protein-binding sites, it cannot show how individual target genes are regulated in response to binding. Expression profiling of neuroblasts and

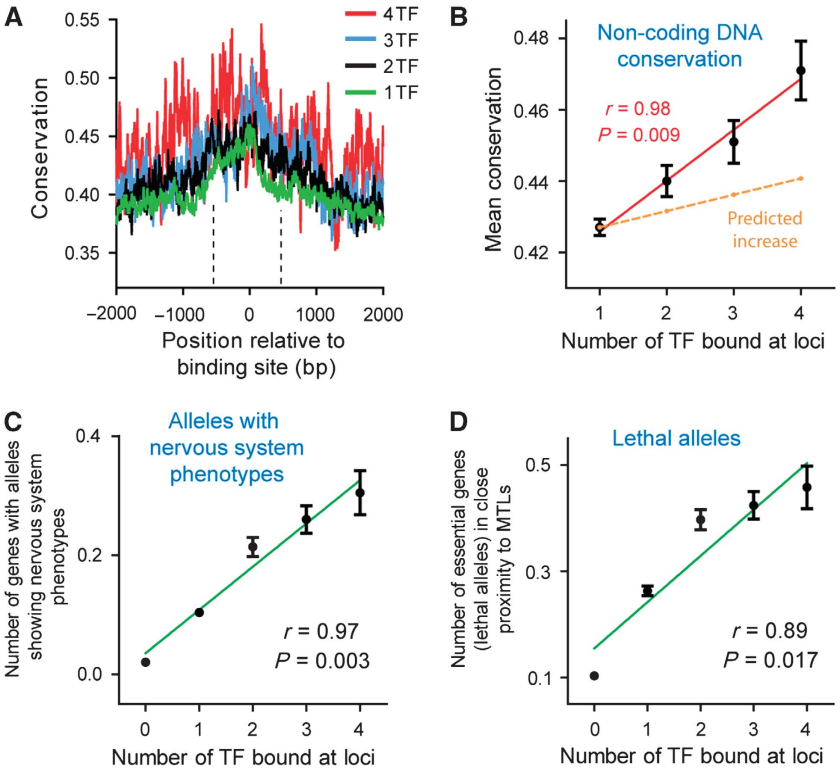

Figure 4 Properties of identified multiple transcription-binding loci (MTLs). (**A**) Mean plot of DNA conservation at loci with one and multiple instances of transcription factor binding (determined by DamID). (**B**) Correlation between number of transcription factors bound and DNA conservation. Hashed orange line represents predicted increase if increments are only because of the binding of 1, 2 or 3 additional factors. (**C**) Frequency of genes proximal to MTLs that have alleles with nervous system phenotypes. (**D**) Frequency of genes proximal to MTLs that have lethal alleles. Bars represent standard error (s.e.m.).

GMCs from wild type and mutant embryos can provide this information, and provide greater insight into the biological function of each of the transcriptional regulators (see, for example, Choksi *et al*, 2006).

Expression profiling of *asense* mutants was performed on 50–100 neuroblasts and GMCs microdissected from the ventral nerve cord of stage 11 wild type and mutant embryos (see Material and methods). Figure 6A shows genes that are bound by Asense and exhibit a significant change in expression level in *asense* mutant neuroblasts and GMCs. In many cases, neuronal differentiation and Notch pathway genes (enhancer of split complex (E(spl)-C) and bearded complex) are upregulated in the mutant, suggesting that Asense normally represses them, whereas neuroblast genes are downregulated, suggesting they require Asense for expression (Figure 6A; full data in Supplementary Table S2). This contrasts with our data for Prospero, which represses neuroblast genes and is required for the activation of differentiation genes. Combined with the fact that Prospero represses expression of Asense, these data support an antagonistic relationship between Prospero and Asense. For example, the neuroblast genes *miranda* (Shen *et al*, 1997) and *grainy head* (Brody and Odenwald, 2000) are activated by Asense and repressed by Prospero, whereas transcription of the differentiation gene *Fasciclin I* (Elkins *et al*, 1990) is promoted by Prospero but inhibited by Asense (Figure 6B). Interestingly, however, there are also examples of differentiation and cell-cycle exit genes activated by Asense, such as *commissureless* (Tear *et al*, 1996), *hikaru genki* (Hoshino *et al*, 1996) and *dacapo* (de Nooij *et al*, 1996). Furthermore, when the full expression array data from *prospero* mutants (Choksi *et al*, 2006) and *asense* mutants are compared by cluster analysis (Supplementary Figure S9) we find two clusters in which genes are regulated antagonistically, but also two clusters in which genes are similarly regulated. These data suggest a dual role for Asense: activating the expression of neuroblast genes and repressing differentiation genes in the neuroblast, whereas promoting differentiation when present in the GMC.

Discussion

We combined *in vivo* chromatin profiling and cell-specific expression profiling to identify the gene regulatory networks directing neural stem cell fate and promoting differentiation in the *Drosophila* embryo. We find that the transcription factors Asense, Deadpan, Snail and Prospero bind to many of the same target genes. The targets of Asense, Deadpan and Snail include neuroblast genes but also many differentiation genes. The binding of these neural stem cell factors to differentiation genes is not entirely unexpected. In vertebrates, stem cell transcription factors bind to and repress differentiation genes to maintain the stem cell state (Boyer *et al*, 2005; Loh *et al*, 2006). Additionally, it is becoming apparent that transcription factors can have roles in both activation and repression, in *Drosophila* (Choksi *et al*, 2006) and in vertebrate stem cell transcriptional networks (Boyer *et al*, 2005; Loh *et al*, 2006). The ability to either repress or activate is likely to be due to interaction with co-factors, and

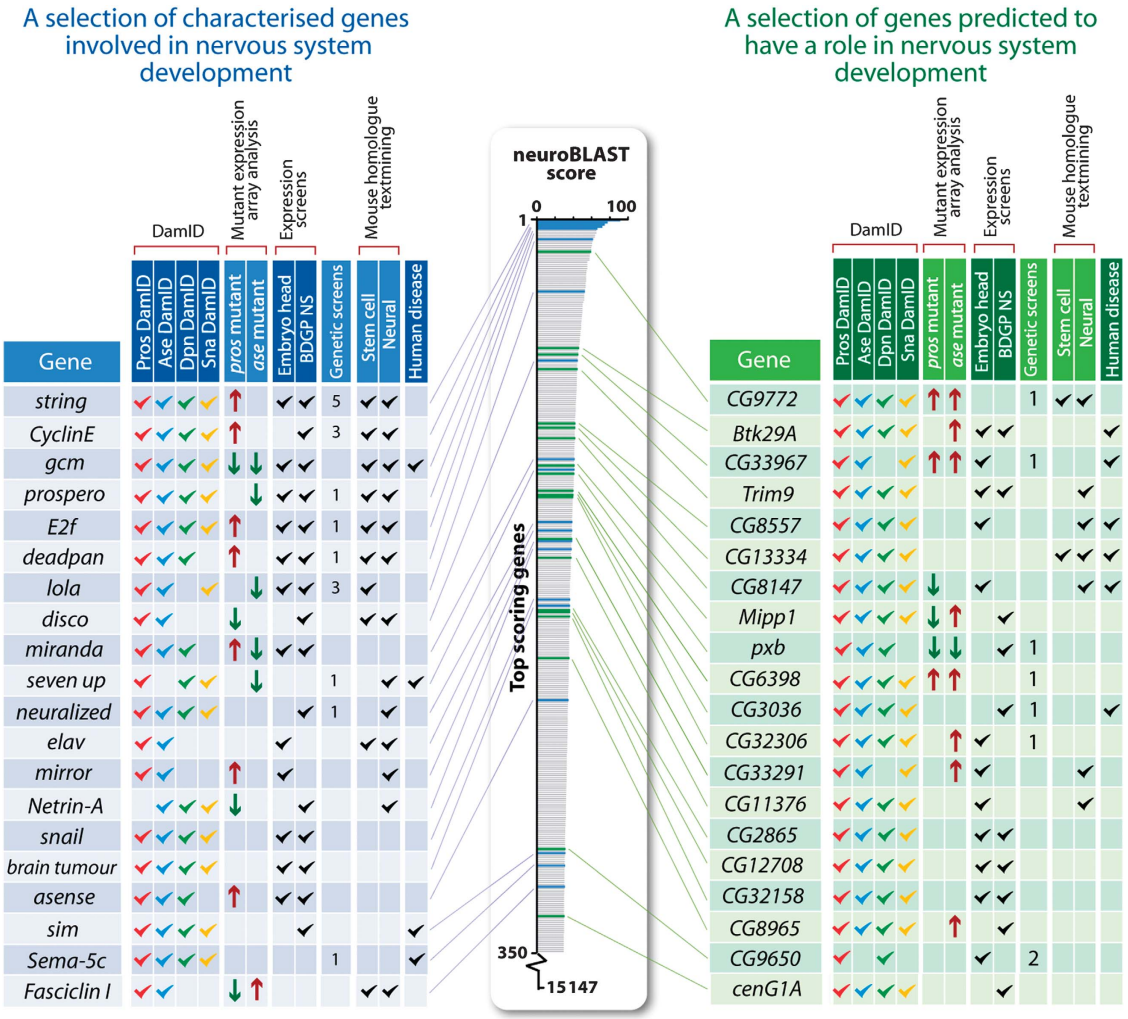

Figure 5 Prediction of key neural developmental genes using the neuroBLAST database. Identification of key neurogenesis genes using the neuroBLAST database. Red arrows indicate higher expression levels and green arrows lower expression levels in the mutant tissue. Weighted scores for each screen were generated by a random permutation algorithm, using a training set containing known nervous system development genes.

the ability to recruit chromatin remodelling complexes to specific loci.

We showed previously that Prospero represses the expression of Asense and Deadpan in GMCs, supporting a model whereby a core set of genes involved in neuroblast self-renewal and multipotency is activated by the neuroblast transcription factors and repressed by Prospero. Here, we show that, in part, Asense acts oppositely to Prospero, promoting the expression of neuroblast genes and repressing certain differentiation genes. However, our data also indicate that Asense can promote the expression of some genes required for differentiation, including the cell-cycle inhibitor *dacapo*, which is a member of the p21/p27 family of cdk inhibitors (de Nooij *et al*, 1996). *dacapo* expression initiates in the GMC (Liu *et al*, 2002) and we observe a reduction in levels of *dacapo* mRNA in the *asense* mutant neuroblasts and GMCs, similar to what has been reported in the developing optic lobe (Wallace *et al*, 2000). *asense* mRNA is known to be expressed in at least a subset of GMCs (Brand *et al*, 1993) and Asense protein is present in larval GMCs (Bowman *et al*, 2008). This suggests that Asense has a secondary role, to promote GMC cell-cycle exit and differentiation. Asense is absent in larval PAN neuroblasts whose progeny, unlike GMCs, divide in a stem cell-like manner. (Bello *et al*, 2008; Bowman *et al*, 2008). Ectopic expression of Asense prevents formation of these daughter cells, which can undergo extra divisions (Bowman *et al*, 2008), possibly by the upregulation of *dacapo*, and other differentiation genes.

The expression pattern, function and binding site specificity of Asense all correlate strongly with its vertebrate counterpart, Ascl1 (Mash1). Mash1 is expressed in neural precursors in vertebrates (Parras *et al*, 2004), is known to regulate genes involved in Notch signalling (*Delta, Jag2, Lfng* and *Magi1*), cell-cycle control (*Cdc25b*) and neuronal differentiation (*Insm1*) (Castro *et al*, 2006) and recognises the E-box sequence, CAGCTG (Hu *et al*, 2004; Castro *et al*, 2006). Furthermore, Mash1 is consistently found to promote neuronal differentiation (Sommer *et al*, 1995; Tomita *et al*, 1996), which is consistent with a pro-differentiation role for Asense. Conversely, we show that Asense activates the expression of certain neuroblast genes, such as *miranda*, which is expressed in all neuroblasts and repressed by Prospero. Deadpan and Snail bind to many neuroblast genes. Given that the expression of *deadpan* and *snail* is restricted to pan-neural neuroblasts, it is likely that they can

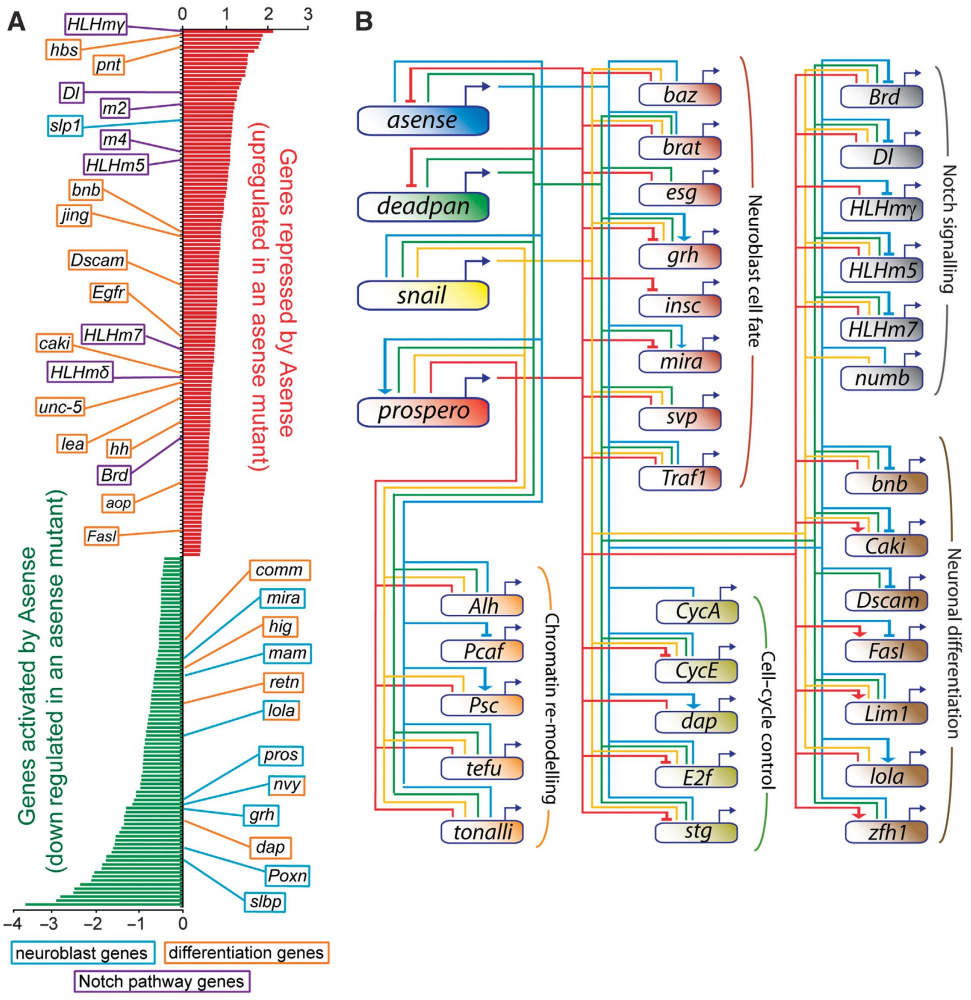

Figure 6 Neural stem cell transcriptional network integrating data from DamID chromatin profiling and high-resolution transcriptional profiling of neuroblasts and GMCs. (**A**) Graph represents genes bound by Asense that show a significant change in expression level in *asense* mutant neuroblasts and GMCs. *x* axis represents log2-transformed ratio of mRNA levels between *wild type* and *asense* mutant cells. (**B**) An abbreviated neural stem cell transcriptional network integrating data from DamID chromatin profiling and high-resolution transcriptional profiling of neuroblasts and GMCs. Key groupings of genes are emphasised. Arrows represent activation, bars represent repression.

also activate the expression of neuroblast genes. However, confirmation of this awaits expression profiling of *deadpan* and *snail* mutant neuroblasts and GMCs.

Finally, we show that multiple transcription factor binding is associated with genes that have critical functions in neural development. We show that this relationship can be used to identify novel genes involved in neural development, including those with vertebrate counterparts. A similar gene network and data mining study, using two pair-rule genes in *Drosophila*, has recently been used to identify a new marker for kidney cancer (Liu *et al*, 2009). Therefore, large-scale analysis of gene regulatory networks, as used here, provides a powerful approach to identifying key genes involved in development and disease.

Materials and methods

Fly lines

asense[1] is a deficiency removing the entire *asense* gene (Gonzalez *et al*, 1989). Control flies used for expression profiling were w^{1118}. UAS-Dam has been described previously (Choksi *et al*, 2006).

DamID

Full-length coding sequences from *asense*, *deadpan* and *snail* were PCR amplified from an embryonic cDNA library and cloned into pUASTNDam (Choksi *et al*, 2006) using *Bgl*II and *Not*I sites. Transgenic lines were generated as described previously (Choksi *et al*, 2006). Stage 10–11 embryos (4–7 h AEL) were collected from *UAS-Dam*, *UAS-Dam-Asense*, *UAS-Dam-Deadpan* and *UAS-Dam-Snail*. DNA isolation, processing and amplification were performed as described previously (Choksi *et al*, 2006). Samples were labelled and hybridised to a custom whole genome 375 000 feature tiling array, with 60-mer oligonucleotides spaced at ~300 bp intervals (Choksi *et al*, 2006). Arrays were scanned and intensities extracted (Nimblegen Systems). Two biological replicates for each TF were performed. Log2 mean ratios of each spot were median normalised.

DamID analysis

A peak finding algorithm with false discovery rate (FDR) analysis (PERL script available on request) was developed to identify significant binding sites. All peaks spanning four or more consecutive probes (> ~1200 bp) over a two-fold increase were analysed and assigned an FDR value (using 100 iterations). This analysis was performed for each chromosome arm. Nimblegen genome coordinates were converted to Release 5.0 of the *Drosophila* genome and genes were defined as targets in which a binding event (with an FDR of 0%) occurred within up to 5 kbp of the gene structure (depending on the proximity of adjacent genes). MTLs were defined as ~1800 bp genomic regions (six consecutive

probes) containing binding events from 2, 3 or 4 transcription factors.

Expression profiling

Using a drawn out and bevelled capillary, samples of ~100 cells were extracted *in vivo* from the ventral and intermedial columns of the ventral nerve cords of late stage 11 homozygous *ase*^{1} embryos (*ase*^{1}*/ FM7c, Kruppel-GAL4; UAS-GFP*) and w^{1118} embryos. cDNA was amplified from the cells, labelled and hybridised to microarrays (FL002; Flychip Cambridge Microarray Facility) as described previously (Choksi *et al*, 2006).

Statistics

FDR analysis was performed for DamID determined binding sites (see above). Linear correlation regression and significance were calculated in Graphpad Prism. Significance analysis of microarrays (Tusher *et al*, 2001) was used to analyse the expression microarray data. Genes with less than 1% FDR were identified as significant.

MICRA, conservation analysis and database construction

See Supplementary data.

Immunohistochemistry and in situ hybridisation

See Supplementary data.

Public database access of microarray data

The raw and normalised data for the DamID- binding experiments and the expression profiling experiments described here are available on the GEO public database (http://www.ncbi.nlm.nih.gov/projects/geo/). The accession numbers are as follows: Four DamID data sets (Asense, Deadpan, Snail and Prospero)—GSE18270. *asense* mutant expression profiling—GSE18214. *prospero* mutant expression profiling—GSE18213.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank A Jarman for providing fly stocks; D Griffiths, M Goodwin, S Sansom and members of the Brand lab for critical reading of the paper; K Edoff and FlyChip for assisting with expression profiling experiments and A Carr for advice on bioinformatics. This work was funded by a Wellcome Trust Programme Grant and an MRC Project Grant to AHB. TDS is funded by a Herchel Smith Fellowship.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD (2005) Neuronal subtype-specific genes that control corticospinal motor neuron development *in vivo*. *Neuron* **45:** 207–221

Ashraf SI, Ip YT (2001) The Snail protein family regulates neuroblast expression of inscuteable and string, genes involved in asymmetry and cell division in Drosophila. *Development* **128:** 4757–4767

Bello B, Reichert H, Hirth F (2006) The brain tumor gene negatively regulates neural progenitor cell proliferation in the larval central brain of Drosophila. *Development* **133:** 2639–2648

Bello BC, Izergina N, Caussinus E, Reichert H (2008) Amplification of neural stem cell proliferation by intermediate progenitor cells in Drosophila brain development. *Neural Dev* **3:** 5

Betschinger J, Mechtler K, Knoblich JA (2006) Asymmetric segregation of the tumor suppressor brat regulates self-renewal in Drosophila neural stem cells. *Cell* **124:** 1241–1253

Bianchi-Frias D, Orian A, Delrow JJ, Vazquez J, Rosales-Nieves AE, Parkhurst SM (2004) Hairy transcriptional repression targets and cofactor recruitment in Drosophila. *PLoS Biol* **2:** E178

Bier E, Vaessin H, Younger-Shepherd S, Jan LY, Jan YN (1992) deadpan, an essential pan-neural gene in Drosophila, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev* **6:** 2137–2151

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816

Boone JQ, Doe CQ (2008) Identification of Drosophila type II neuroblast lineages containing transit amplifying ganglion mother cells. *Dev Neurobiol* **68:** 1185–1195

Bowman SK, Rolland V, Betschinger J, Kinsey KA, Emery G, Knoblich JA (2008) The tumor suppressors Brat and Numb regulate transit-amplifying neuroblast lineages in Drosophila. *Dev Cell* **14:** 535–546

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956

Brand M, Jarman AP, Jan LY, Jan YN (1993) asense is a Drosophila neural precursor gene and is capable of initiating sense organ formation. *Development* **119:** 1–17

Brody T, Odenwald WF (2000) Programmed transformations in neuroblast gene expression during Drosophila CNS lineage development. *Dev Biol* **226:** 34–44

Cai Y, Chia W, Yang X (2001) A family of snail-related zinc finger proteins regulates two distinct and parallel mechanisms that mediate Drosophila neuroblast asymmetric divisions. *EMBO J* **20:** 1704–1714

Castro DS, Skowronska-Krawczyk D, Armant O, Donaldson IJ, Parras C, Hunt C, Critchley JA, Nguyen L, Gossler A, Gottgens B, Matter JM, Guillemot F (2006) Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Dev Cell* **11:** 831–844

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106–1117

Choksi SP, Southall TD, Bossing T, Edoff K, de Wit E, Fischer BE, van Steensel B, Micklem G, Brand AH (2006) Prospero acts as a binary switch between self-renewal and differentiation in Drosophila neural stem cells. *Dev Cell* **11:** 775–789

de Nooij JC, Letendre MA, Hariharan IK (1996) A cyclin-dependent kinase inhibitor, Dacapo, is necessary for timely exit from the cell cycle during Drosophila embryogenesis. *Cell* **87:** 1237–1247

Doe CQ (2008) Neural stem cells: balancing self-renewal with differentiation. *Development* **135:** 1575–1587

Egger B, Chell JM, Brand AH (2008) Insights into neural stem cell biology from flies. *Philos Trans R Soc Lond B Biol Sci* **363:** 39–56

Elkins T, Zinn K, McAllister L, Hoffmann FM, Goodman CS (1990) Genetic analysis of a Drosophila neural cell adhesion molecule: interaction of fasciclin I and Abelson tyrosine kinase mutations. *Cell* **60:** 565–575

Gonzalez F, Romani S, Cubas P, Modolell J, Campuzano S (1989) Molecular analysis of the asense gene, a member of the achaete-scute complex of Drosophila melanogaster, and its novel role in optic lobe development. *EMBO J* **8:** 3553–3562

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951

Hemavathy K, Ashraf SI, Ip YT (2000) Snail/slug family of repressors: slowly going into the fast lane of development and cancer. *Gene* **257:** 1–12

Hoshino M, Suzuki E, Nabeshima Y, Hama C (1996) Hikaru genki protein is secreted into synaptic clefts from an early stage of synapse formation in Drosophila. *Development* **122:** 589–597

Hu Y, Wang T, Stormo GD, Gordon JI (2004) RNA interference of achaete-scute homolog 1 in mouse prostate neuroendocrine cells reveals its gene targets and DNA binding sites. *Proc Natl Acad Sci USA* **101:** 5559–5564

Jarman AP, Brand M, Jan LY, Jan YN (1993) The regulation and function of the helix-loop-helix gene, asense, in Drosophila neural precursors. *Development* **119:** 19–29

Jones BW, Fetter RD, Tear G, Goodman CS (1995) glial cells missing: a genetic switch that controls glial versus neuronal fate. *Cell* **82:** 1013–1023

Kanai MI, Okabe M, Hiromi Y (2005) seven-up Controls switching of transcription factors that specify temporal identities of Drosophila neuroblasts. *Dev Cell* **8:** 203–213

Lee CY, Wilkinson BD, Siegrist SE, Wharton RP, Doe CQ (2006) Brat is a Miranda cargo protein that promotes neuronal differentiation and inhibits neuroblast self-renewal. *Dev Cell* **10:** 441–449

Liu J, Ghanim M, Xue L, Brown CD, Iossifov I, Angeletti C, Hua S, Negre N, Ludwig M, Stricker T, Al-Ahmadie HA, Tretiakova M, Camp RL, Perera-Alberto M, Rimm DL, Xu T, Rzhetsky A, White KP (2009) Analysis of Drosophila segmentation network identifies a JNK pathway factor overexpressed in kidney cancer. *Science*

Liu TH, Li L, Vaessin H (2002) Transcription of the Drosophila CKI gene dacapo is regulated by a modular array of cis-regulatory sequences. *Mech Dev* **112:** 25–36

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL *et al.* (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38:** 431–440

Martin DM, Berriman M, Barton GJ (2004) GOtcha: a new method for prediction of protein function assessed by the annotation of seven genomes. *BMC Bioinformatics* **5:** 178

Mauhin V, Lutz Y, Dennefeld C, Alberga A (1993) Definition of the DNA-binding site repertoire for the Drosophila transcription factor SNAIL. *Nucleic Acids Res* **21:** 3951–3957

Moorman C, Sun LV, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu XJ, White KP, Bussemaker HJ, van Steensel B (2006) Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proc Natl Acad Sci USA* **103:** 12027–12032

Parras CM, Galli R, Britz O, Soares S, Galichet C, Battiste J, Johnson JE, Nakafuku M, Vescovi A, Guillemot F (2004) Mash1 specifies neurons and oligodendrocytes in the postnatal brain. *EMBO J* **23:** 4495–4505

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

Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* **314:** 1041–1052

Shen CP, Jan LY, Jan YN (1997) Miranda is required for the asymmetric localization of Prospero during mitosis in Drosophila. *Cell* **90:** 449–458

Sommer L, Shah N, Rao M, Anderson DJ (1995) The cellular function of MASH1 in autonomic neurogenesis. *Neuron* **15:** 1245–1258

Song S, Cooperman J, Letting DL, Blobel GA, Choi JK (2004) Identification of cyclin D3 as a direct target of E2A using DamID. *Mol Cell Biol* **24:** 8790–8802

Speicher S, Fischer A, Knoblich J, Carmena A (2008) The PDZ protein Canoe regulates the asymmetric division of Drosophila neuroblasts and muscle progenitors. *Curr Biol* **18:** 831–837

Sun LV, Chen L, Greil F, Negre N, Li TR, Cavalli G, Zhao H, Van Steensel B, White KP (2003) Protein-DNA interaction mapping using genomic tiling path microarrays in Drosophila. *Proc Natl Acad Sci USA* **100:** 9428–9433

Tear G, Harris R, Sutaria S, Kilomanski K, Goodman CS, Seeger MA (1996) commissureless controls growth cone guidance across the CNS midline in Drosophila and encodes a novel membrane protein. *Neuron* **16:** 501–514

Tolhuis B, de Wit E, Muijrers I, Teunissen H, Talhout W, van Steensel B, van Lohuizen M (2006) Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nat Genet* **38:** 694–699

Tomancak P, Berman BP, Beaton A, Weiszmann R, Kwan E, Hartenstein V, Celniker SE, Rubin GM (2007) Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol* **8:** R145

Tomita K, Nakanishi S, Guillemot F, Kageyama R (1996) Mash1 promotes neuronal differentiation in the retina. *Genes Cells* **1:** 765–774

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98:** 5116–5121

van Steensel B, Henikoff S (2000) Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18:** 424–428

van Steensel B, Delrow J, Henikoff S (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* **27:** 304–308

Vogel MJ, Peric-Hupkes D, van Steensel B (2007) Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* **2:** 1467–1478

Wallace K, Liu TH, Vaessin H (2000) The pan-neural bHLH proteins DEADPAN and ASENSE regulate mitotic activity and cdk inhibitor dacapo expression in the Drosophila larval optic lobes. *Genesis* **26:** 77–85

Winston RL, Ehley JA, Baird EE, Dervan PB, Gottesfeld JM (2000) Asymmetric DNA binding by a homodimeric bHLH protein. *Biochemistry* **39:** 9092–9098

Winston RL, Millar DP, Gottesfeld JM, Kent SB (1999) Characterization of the DNA binding properties of the bHLH domain of Deadpan to single and tandem sites. *Biochemistry* **38:** 5138–5146

Wodarz A (2005) Molecular control of cell polarity and asymmetric cell division in Drosophila neuroblasts. *Curr Opin Cell Biol* **17:** 475–481

Yu F, Kuo CT, Jan YN (2006) Drosophila neuroblast asymmetric cell division: recent advances and implications for stem cell biology. *Neuron* **51:** 13–20

Zhong W, Chia W (2008) Neurogenesis and asymmetric cell division. *Curr Opin Neurobiol* **18:** 4–11