Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development.

Marek Wojciechowski,[1,4] Robert Lowe,[2,4] Joanna Maleszka,[3] Danyal Conn,[1] Ryszard Maleszka,[3,5] and Paul J. Hurd[1,5]

[1] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK.

[2] The Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London, E1 2AT, UK.

[3] Research School of Biology, Australian National University, ACT, 0200, Australia.

[4] These authors contributed equally to this work.

[5] Corresponding authors

E-mail: p.j.hurd@qmul.ac.uk

E-mail: ryszard.maleszka@anu.edu.au

Running title: Caste-specific chromatin patterns in honey bees

Keywords:

Apis mellifera; phenotypic plasticity; caste differentiation; chromatin; histone modification; social insect; epigenetics.

Abstract

The capacity of the honey bee to produce three phenotypically distinct organisms or castes (queen, sterile female worker and a haploid male drone) from one genotype represents one of the most remarkable examples of developmental plasticity in any phylum. The queen-worker morphological and reproductive divide is environmentally controlled during post-embryonic development by differential feeding. Previous studies have implicated metabolic flux acting via epigenetic regulation, in particular DNA methylation and microRNAs, in establishing distinct patterns of gene expression underlying caste-specific developmental trajectories. We produce the first genome-wide maps of chromatin structure in the honey bee at a key larval stage where developmental canalization into queen or worker is virtually irreversible. We find extensive genome-wide differences in H3K4me3, H3K27ac and H3K36me3, many of which correlate with caste-specific transcription. Furthermore, we identify H3K27ac as a key chromatin modification, with caste-specific regions of intronic H3K27ac directing the worker caste. We suggest that these regions harbour the first examples of caste-specific enhancer elements in the honey bee. Our results demonstrate a key role for chromatin modifications in the establishment and maintenance of caste-specific transcriptional programmes in the honey bee. We show that at 96hrs of larval growth the queen-specific chromatin pattern is already established, whereas the worker determination is not, thus providing experimental support for the perceived timing of this critical point in developmental heterochrony in two types of honey bee females. In a broader context, our study provides novel data on environmentally-regulated organismal plasticity and the molecular foundation of the evolutionary origins of eusociality.

Introduction

Eusociality is an intriguing evolutionary invention found in many species of Hymenoptera (e.g. bees, wasps and ants), termites and even some mammals (Jarvis 1981; Gordon 2002; Nalepa 2015). In true eusocialism, large self-organising colonies are formed out of individuals partitioned into reproductive and non-reproductive types known as castes, each representing an organism with a distinct repertoire of morphological, physiological and behavioural characteristics. In some species this phenotypic divide is epigenetically rather than genetically determined consistent with the environmental impact by which these differences are implemented.

Insect pollinators such as the honey bee (*Apis mellifera*) play a crucial role in most ecosystems and strongly influence ecological relationships, for example by helping to maintain genetic variation in flowering plants. Furthermore, in farmed areas, the honey bee is used extensively for the commercial pollination of a variety of cultivated crops. Honey bees live in complex societies comprising tens of thousands of individuals, in which there is a division of labour that can be separated into two broad categories. The first is a reproductive distinction; each colony contains two diploid female castes comprising a single queen who is specialised for reproduction and thousands of sterile female worker bees (Winston 1991). The second distinction relates to the division of tasks performed by the worker caste which changes during the course of adult life from nurse through to forager, in a process termed behavioural maturation which results in two worker sub-castes (Winston 1991). A third main caste, which develops from unfertilised eggs, is a haploid male drone. The key feature in the establishment of these different female developmental trajectories and subsequent maintenance during adulthood, is nutrition. For the first 72h after hatching, both queen and worker larvae receive a certain amount of nutritious jelly, although the worker jelly contains lower concentration of sugars and a few other ingredients than the queen food known as royal jelly (Wang et al. 2016; Maleszka 2018). Drone larvae not only receive a distinct diet, but also in larger quantities compared to that of worker larvae. This

suggests that similarly to queens, nutrition provides important cues for their proper development (Hrassnigg and Crailsham 2007). After this, larvae developmentally destined to be workers or drones are then switched to a diet comprised of nectar and pollen, in contrast to larvae destined to become queens, which remain on a royal jelly diet. After 96h, larval chambers are capped and no further feeding occurs until after pupation. Differential feeding continues throughout adulthood, resulting in distinct but genetically indistinguishable organisms/castes. The honey bee genome therefore exemplifies environmentally driven phenotypic plasticity, where diet dictates the ability of different phenotypes to arise from a single genome and represents one of the most striking examples of developmental plasticity in any phylum.

The establishment, maintenance and modulation of transcriptional programmes such as those during development are reliant on the inherent plasticity of chromatin and recent evidence indicates that chromatin-based epigenetic mechanisms direct nutrition-mediated caste differentiation in the honey bee. RNAi knock-down of the putative *de novo* DNA methyltransferase DNMT3 in newly hatched larvae has been shown to lead to royal jelly-like effects on developmental trajectory, resulting in a significantly high proportion of queens with fully developed ovaries (Kucharski et al. 2008). The potential role of differential DNA methylation in influencing alternate developmental outcomes of queens and workers has been confirmed by genome-wide mapping of methylated CpGs in both castes at 96hrs of larval growth (Foret et al. 2012). While the exact function of this common epigenomic modification in the honey bee remains poorly understood several studies have shown that differential DNA methylation correlates with alternative splicing and modulation of gene expression in a context-dependent manner (Lyko et al. 2010; Foret et al. 2012; Kucharski et al. 2016; Wedd et al. 2016). More recently, using proteomic approaches, we have demonstrated that honey bee histone proteins are extensively post-translationally modified and show caste-specific signatures (Dickman et al. 2013). Given the conservation of both histone sequences and epigenetic machinery in the honey bee, we hypothesise that histone post-translational modifications (PTM) are also pivotal in determining developmental

trajectory in response to nutrition in this organism. Furthermore, a direct link between a component of royal jelly and potential caste-specific histone PTM changes has been provided by a biochemical study of a fatty acid, (*E*)-10-hydroxy-2-decenoic acid (10-HDA), which comprises up to 5% of royal jelly. 10-HDA has been shown to be a histone deacetylase inhibitor and can reactivate the expression of epigenetically silenced genes in mammalian cells (Spannhoff et al. 2011).

Given that hundreds of genes have been implicated in queen-worker differentiation (Barchuk et al. 2007; Foret et al. 2012) we reasoned that their coordinated differential expression has to be regulated at the level of chromatin. Chromatin structure has not previously been studied in the honey bee and for the first time, we have determined the genome-wide distribution of three histone H3 modifications (H3K4me3, H3K27ac and H3K36me3) in both queen and worker female castes, at a crucial development time point that has been shown to be critical for caste determination with both types of females essentially committed to a specific trajectory (Weaver 1966; Maleszka 2018). We sequenced chromatin immunoprecipitation (ChIP) and RNA samples from 96h larval heads and have identified thousands of genomic regions that show caste-specific chromatin states, many of which are linked to caste-specific gene transcription.

Results

Histone post-translational modifications in the honey bee associate with transcribed regions

We have previously identified over 20 different histone PTM states in queen and worker honey bee castes using mass spectrometry (Dickman et al. 2013; P Hurd, unpublished). In order to study chromatin structure in honey bees, for the first time we determined the genome-wide distribution of three histone PTMs that are associated with transcription and active *cis*-elements in other organisms; H3K4me3, H3K27ac and H3K36me3 (Pokholok et al. 2005; Heintzman et al. 2007; Guenther et al. 2007; Négre et al. 2011; Simola et al. 2013).

We profiled replicate pools of worker (W) and queen (Q) larval heads at 96h post-hatching by ChIP-seq (n = 50 per caste). Replicates show a very strong and significant correlation across all histone PTMs and castes ($\rho > 0.96$; p-value < 2.2×10^{-16}; Supplemental Fig. S1). We find enrichment of H3K4me3 and H3K27ac and depletion of H3K36me3 around the transcriptional start sites (TSS) of genes in both 96hW and 96hQ castes (Fig. 1A and Supplemental Fig. S2), similar to what has previously been reported in other organisms including vertebrates, invertebrates and plants (Pokholok et al. 2005; Heintzman et al. 2007; Guenther et al. 2007; Négre et al. 2011; Simola et al. 2013). H3K36me3 is mostly found downstream of TSSs (Fig. 1A) suggesting that it demarcates gene bodies. Overall, the majority of the 10,746 protein coding genes in the honey bee show high levels of enrichment (> 3-fold over input) for at least one of the three histone PTMs profiled (57.6% in 96hW and 61.3% in 96hQ). Since each of these histone PTMs are mostly localised close to or within genes, we next investigated their correlation with transcription. We profiled gene expression in 96hW and 96hQ larval heads (n = 4 per caste) by RNA-seq. In both 96hW and 96hQ castes, genes marked uniquely by H3K4me3 or H3K36me3 show a significant (p-value < 3.1×10^{-4}) increase in expression compared to background, while genes marked uniquely by H3K27ac do not show any significant association with expression (Fig. 1B). Furthermore, in both castes, genes marked by two or more histone PTMs also show significant increases in expression compared to background (p-value < 0.047). Thus, in both worker and queen honey bee castes; H3K4me3, H3K36me3 and H3K27ac when occurring in combination, associate with actively transcribed regions and therefore have the potential to regulate caste-specific gene expression.

Caste-specific chromatin patterns correlate with differential gene expression

Having established that within each caste, H3K4me3, H3K27ac and H3K36me3 are significantly enriched at transcribed regions, we next wanted to determine whether these three histone PTMs show caste-specific distributions. To investigate this, we called differences between 96hW and 96hQ castes for H3K4me3, H3K27ac and H3K36me3 (Fig.

2A). For H3K4me3, we identify 1,834 unique genomic regions that are significantly more enriched in 96hW and 3,333 in 96hQ (adjusted p-value < 0.01 and $|\Delta Enrichment| > 3$). For H3K27ac, we identify 2,027 unique genomic regions that are significantly more enriched in 96hW and 489 in 96hQ (adjusted p-value < 0.01 and $|\Delta Enrichment| > 3$). Finally, for H3K36me3 we identifiy 1,196 unique regions that are significantly more enriched in 96hW and 3,285 in 96hQ (adjusted p-value < 0.01 and $|\Delta Enrichment| > 3$). Gene ontology analysis of those genes that show any significant caste-specific chromatin marks reveal a distinct developmental separation of castes at 96h. In 96hQ, overrepresented GO terms significantly associate with physio-metabolic functions and processes, such as the structural constituents of ribosome (GO:0003735) and biological processes including cellular amide metabolic processes (GO:0043603), cytoplasmic translation (GO:0002181) and peptide metabolic processes (GO:0006518), suggesting that at 96h, queen developmental trajectory is established (Fig. 2B and Supplemental Table S1). In stark contrast, 96hW GO terms associate with development and transcriptional programming, including the molecular functions of transcription factor activity and binding (GO:0003700) along with biological processes of anatomical structure morphogenesis (GO:0009653), system development (GO:0048731) and developmental processes (GO:0032502). Importantly, this suggests that relative to the 96hQ, worker caste developmental trajectory is not yet established at 96h (Fig. 2B, Supplemental Fig. S3 and Supplemental Table S2). Since caste-specific DNA methylation patterns have previously been reported (Lyko et al. 2010; Foret et al. 2012), we asked whether the observed caste-specific differences in histone PTM enrichment correlated with differentially DNA methylated positions (DMPs). We reanalysed the 96h larval DNA methylation data of Foret et al using Fisher's exact test and detected a total of 24,663 DMPs (adjusted BH p-value < 0.05) and then analysed the enrichment of these DMPs with each of our differential histone PTMs. We find only a strong enrichment in H3K36me3 (permutation test; p-value < 0.001; Fold Enrichment = 15). However, this is driven by an overlap in genome distribution rather than a strong functional association (see Supplemental Figs. S4 and S5). This suggests that DNA methylation and histone PTMs provide different signals in

197 the process of caste determination. We next wanted to determine whether the observed
198 caste-specific differences in histone PTM enrichment, correlated with any differential gene
199 expression between castes. Unsupervised multi-dimensional scaling analysis of our RNA-
200 seq data reveal a strong separation of the two castes (see Supplemental Fig. S6). We
201 identify a total of 1,060 significant differences (genome wide adjusted p-value < 0.01) in
202 transcript levels between 96hW and 96hQ (Fig. 2C), with 386 transcripts showing an
203 increase in expression in the 96hW and 674 showing an increase in the 96hQ (see
204 Supplemental Fig. S7 and Supplemental Tables S3 and S4). To confirm that we were
205 robustly measuring differences between castes, we compared these differentially expressed
206 transcripts to previously published RNA-seq data taken from whole 96hW and 96hQ larvae
207 ((Ashby et al. 2016); see Supplemental Fig. S8). We find a strong correlation ($\rho = 0.51$; p-
208 value < 5.3×10^{-72}) between the transcriptional differences detected in our experiment and
209 that of Ashby et al. We find that genes differentially enriched with H3K4me3 ($\rho = 0.28$; p-
210 value = 2.7×10^{-17}) and H3K36me3 ($\rho = 0.15$; p-value = 4.6×10^{-5}) show significant
211 correlation with transcriptional differences, suggesting that caste-specific H3K4me3 and
212 H3K36me3 patterns associate with caste-specific transcriptional profiles (Fig. 2D). In
213 contrast, genes differentially enriched with H3K27ac show a non-significant correlation with
214 caste-specific transcriptional differences ($\rho = 0.04$; p-value = 0.074). To investigate the
215 function of genes which show consistent caste-specific changes in both gene expression
216 and histone PTM enrichment, we performed gene ontology analysis. We observe that these
217 genes reveal a distinct developmental separation of castes at 96h. In 96hQ, we again find
218 significant GO terms for biological processes and functions that associate with physio-
219 metabolic processes such as translation (GO:0006412 and GO:0002181) and peptide
220 biosynthetic processes (GO:0043043) (Fig. 2E and Supplemental Table S5). In stark
221 contrast, GO terms for system development (GO:0048731), nervous system development
222 (GO:0007399), generation of neurons (GO:0048699) and neuron differentiation
223 (GO:0030182) are enriched in 96hW (Fig. 2E, Supplemental Fig. S9 and Supplemental

Table S6). In agreement with our earlier analyses, this again suggests that at 96h and relative to worker, queen development is set while the worker-specific developmental programme (which will determine a distinct phenotype) is yet to be established. Representative examples of two genes that show some of the most significant caste-specific changes in both gene expression and histone PTM enrichment are shown in Fig. 3. Pyruvate kinase (*PYK*; 552007) is shown as an example of a physio-metabolic 96hQ-specific gene, where H3K27ac and H3K4me3 enrichment differences associate with the TSS and H3K36me3 over the gene body (Fig. 3A and Supplemental Fig. S10). Similarly, the 96hW-specific gene Ten-eleven translocation (*TET*; 412878), also shows TSS and gene body differences in H3K4me3, H3K27ac and H3K36me3. However, *TET* also has significant 96hW-specific differences in H3K27ac in the intronic region between two alternative TSSs (Fig. 3B). Notably, the longer transcript containing caste-specific intronic H3K27ac (XM_006561197) shows not only the highest level of expression but also the most significant caste-specific difference in expression (LogFC = 0.38; p-value = 0.0075). Taken together, these data indicate that aged-matched worker and queen honey bee castes exhibit distinct patterns of histone PTMs. Furthermore, patterns of H3K4me3 and H3K36me3 correlate with caste-specific gene expression and reveal that at this crucial time point and relative to the queen, the worker-specific developmental pathway is not yet established.

Intronic H3K27ac regions most readily define the worker caste and are enriched for transcription factor binding sites

Previously, we observed a strong correlation between caste-specific H3K4me3 and H3K36me3 patterns and transcriptional profiles but not for H3K27ac. However, at the *TET* locus, an alternative longer transcript containing 96hW-specific intronic H3K27ac did correlate strongly with caste-specific transcription, whereas a shorter transcript with caste-specific H3K27ac around the TSS did not. This led us to examine more closely the distribution of H3K27ac caste differences and to this end, we plotted unique ChIP-seq regions relative to TSSs. In both castes, the majority of unique H3K4me3 regions are

similarly located around the TSS. For H3K36me3, the majority of unique regions are in gene bodies with the distribution in 96hW more downstream than 96hQ. In contrast, differences in H3K27ac have a much more pronounced caste-specific distribution (Fig 4A). An increase in enrichment of H3K27ac in 96hQ is almost exclusively located within 0-1kbp of TSSs whereas in 96hW, enrichment is mostly located outside these regions (Fig. 4A). Furthermore, in comparison to H3K4me3 and H3K36me3, there is significantly more caste-specific intergenic H3K27ac (16% in 96hW and 15% in 96hQ; see Fig. 4B). In order to further define the caste-specific distribution of intragenic H3K27ac, we mapped differences to either exons or introns. We observe a stark difference in the location of 96hW and 96hQ -specific intragenic H3K27ac, with 83% of all 96hQ-specific enrichments occurring within exons while conversely, over 53% of 96hW-specific differences occur within introns (Fig. 4B). We then asked whether there was any correlation between these distinct caste-specific intronic H3K27ac regions and expression of the associated gene. In agreement with our previous observations at the *TET* locus, genes which are significantly more expressed in 96hW show highly significant enrichment for 96hW intronic H3K27ac (1.75-fold enrichment; p-value < 0.001) while those genes significantly more expressed in 96hQ are actually depleted for 96hQ intronic H3K27ac (2.06-fold depletion; p-value < 0.001; see Fig. 4C). However, in both castes, intronic H3K27ac is associated with gene expression (Supplemental Fig. S11). It is also likely that H3K27ac acts at distal regulatory elements and therefore we determined the average expression of genes at various distances from H3K27ac enriched regions, revealing a specific peak of gene expression at a distance of 130-140kbp (Supplemental Figs. S12 and S13). In order to try and gain a better understanding of the functional significance of caste-specific intronic H3K27ac, we performed motif enrichment analysis on these regions using CentriMo from the MEME suite software package (Bailey and MacHanick 2012). Using transcription factor motifs annotated in *Drosophila melanogaster*, we identify highly significant enrichment for Trithorax-like (Trl; p-value < 1.3×10^{-2}) and Mothers against dpp (Mad; p-value < 3.2×10^{-2}), which accounts for 70% of all 96hW-specific regions of intronic H3K27ac (Fig. 4D). In contrast,

66% of all 96hQ-specific intronic H3K27ac regions are enriched in motifs for Brinker (Brk; p-value < 2.3×10^{-5}), Hairy (H; p-value < 1.3×10^{-4}) and Mothers against dpp (Mad; p-value < 5.9×10^{-4}) (Fig. 4D). Analysis of Trl and Mad binding sites in relation to peaks of intronic 96hW-specific H3K27ac, reveals highly significant motif enrichment flanking peaks of H3K27ac, most likely a reflection of nucleosome displacement associated with transcription factor binding (Fig. 4E). Conversely, binding sites for Brk, H and Mad are centred on peaks of 96hQ-specific H3K27ac (Fig. 4E). Given the enrichment for Trl and Mad transcription factor binding sites, the presence of the enhancer associated histone modification H3K27ac, the intronic genomic locations and increased gene expression, these results suggest that 96hW-specific H3K27ac enriched regions are marking active enhancers and play an important role in worker and queen honey bee caste determination.

Discussion

We have used ChIP-seq to provide the first description of genome-wide caste-specific chromatin patterns in the honey bee and furthermore, the first between *Hymenoptera* castes that show a reproductive division of labour. Combined with RNA-seq analysis and at a crucial developmental stage when developmental trajectory has been shown to be irreversible (Weaver 1966), we identify numerous queen and worker -specific chromatin differences many of which correlate with caste-specific transcription. Importantly, regions of the genome which show the most robust caste-specific differences are suggestive of previously unidentified enhancer regions that are important in specifying the worker caste development from that of the queen.

We hypothesized that in order for the honey bee genome to specify two different female castes, different chromatin patterns and transcriptional programmes have to be established during development. Previous work has mainly focused on the role of caste-specific DNA methylation patterns in adult honey bees (Lyko et al. 2010), where differentially methylated regions are mainly localized to exons and are thought to mediate alternative splicing (Foret

et al. 2012). More recently, proteomic analysis of histone modifications (Dickman et al. 2013) and miRNAs (Ashby et al. 2016) have also suggested potential roles in caste-determination. Since chromatin structure has not previously been studied in the honey bee, we first established that within each caste H3K4me3, H3K27ac and H3K36me3 were associated with transcribed regions in a manner consistent in other organisms. Importantly, this also suggested that these highly conserved histone modifications could have the potential to regulate the widely reported caste-specific transcription of the honey bee genome (Evans et al. 1999; Evans and Wheeler 2000; Barchuk et al. 2007; Foret et al. 2012; Chen et al. 2012; Cameron et al. 2013; Ashby et al. 2016). We found that in both worker and queen castes, H3K4me3 and H3K36me3 strongly correlated with transcription whereas H3K27ac alone, did not. Importantly, worker and queen castes had contrasting chromatin patterns for all three histone PTMs at 96h post-hatching and these genomic regions were strongly suggestive of fundamentally different caste developmental states. In queen, enrichment for genes involved in body growth suggested developmental trajectory is established whereas in worker, there was a strong enrichment for processes concerned with continued development and specialisation. This was further supported by transcriptome analysis which also showed strong developmental separation of worker and queen caste at 96h in agreement with other studies (Barchuk et al. 2007; Ashby et al. 2016). Moreover, we found that for H3K4me3 and H3K36me3, caste-specific chromatin signatures correlated with caste-specific transcription suggesting that histone PTMs play a role in determining alternate developmental trajectories. Furthermore, analysis of these regions again highlighted contrasting caste-specific developmental stages at 96h. Biological processes associated with body growth in the queen is in sharp contrast to worker development where neurogenesis is strongly evident. This is consistent with previous observations that queen and worker development is distinct for neurogenesis (Ashby et al. 2016), possibly due to the fact that workers show remarkable behavioural complexity in adult life and would therefore be expected to require a more complex nervous system.

In contrast to H3K4me3 and H3K36me3, we found that caste-specific H3K27ac did not correlate with caste-specific transcriptional differences. Whilst the distribution of caste-specific differences for H3K4me3 and H3K36me3 occurred mainly over similar genomic locations, H3K27ac showed a much more pronounced caste bias. Queen-specific H3K27ac was localized mainly within exons and close to transcriptional start sites whereas conversely, worker-specific H3K27ac was more pervasive and most frequently located within introns. Furthermore, genes with caste-specific regions of intronic H3K27ac correlated with higher levels of caste-specific expression suggesting that these regions may play important *cis*-regulatory roles. Enhancers are *cis*-acting elements that are frequently found in noncoding regions of genomes and are characterized by nucleosome-free regions enriched in transcription factor binding sites (Calo and Wysocka 2013; Li et al. 2016). Activation of enhancers most commonly requires the repositioning of nucleosomes through the activity of ATP-dependent chromatin remodellers followed by transcription factor binding and recruitment of coactivators which modify adjacent nucleosomes most often at H3K27ac (Creyghton et al. 2010; Rada-Iglesias et al. 2011). The transcriptional co-activator CREB binding protein (CBP) is the main histone acetyltransferase that catalyses H3K27ac (Tie et al. 2009; Jin et al. 2011) and honey bees have a single CBP enzyme (726332) that is differentially DNA methylated in *A. mellifera* larvae (Foret et al. 2012). H3K27ac, transcription factor binding motifs and CBP occupancy have been widely used to successfully map enhancers in numerous cell types, tissues and organisms (Visel et al. 2009; Rada-Iglesias et al. 2011; Simola et al. 2013; Koenecke et al. 2016; Négre et al. 2011).

Therefore, in the absence of any previous enhancer annotation in the honey bee genome, we analysed regions of caste-specific intronic H3K27ac for conserved transcription factor binding motifs. Trl (552090; GAGA factor in mammals) and Mad (409301; SMAD1 in mammals) motifs accounted for 70% of all worker-specific intronic H3K27ac. Trl/GAGA factor is a multifunctional transcriptional regulator that is differentially DNA methylated in *A. mellifera* larvae (Foret et al. 2012). Trl/GAGA factor activates transcription by promoting

chromatin remodelling at enhancers, primarily by recruiting the nucleosome remodelling factor (NURF) (Okada and Hirose 1998; Kwon et al. 2016) before associating with, or allowing, other proteins to modulate transcription including CBP (Philip et al. 2015; Boija et al. 2017). NURF can also directly interact with ecdysone receptor (EcR; 406084) in order to potentially target different enhancer regions (Badenhorst et al. 2005). In our present study, EcR shows highly significant worker-specific chromatin and expression patterns (LogFC relative to queen = 0.24), suggesting that NURF-mediated transcriptional activation of enhancers could also be directed via this crucial developmental steroid hormone signalling pathway in honey bees. Mad/SMAD1 is also an enhancer-associated transcription factor that mediates the bone morphogenetic protein (BMP) signalling cascade, acting downstream of decapentaplegic (dpp) (Deignan et al. 2016). Mad/SMAD1 has been demonstrated to interact directly with CBP (Pearson et al. 1999), co-localise with regions of H3K27ac at enhancers (Koenecke et al. 2016) and affect gene expression in a CBP-dependent manner (Waltzer and Bienz 1999). In our study, Magu (411502; LogFC 0.12), BMP receptor 1B (408442; LogFC 0.12) and Mad/SMAD1 (LogFC 0.19) show a significant worker-caste bias in gene expression compared to queen and are all differentially DNA methylated in larvae (Foret et al. 2012), suggesting that BMP signalling via Mad-mediated enhancer activation may also play an important role in worker-caste determination.

In contrast to worker, motifs for the transcriptional repressors Brk and Hairy (410468; HES1 in mammals) showed the most significant enrichment in queen and accounted for 66% of 96hQ-specific H3K27ac regions. Brk and Hairy have been demonstrated to bind enhancers and mediate transcriptional repression (Campbell and Tomlinson 1999; Jaźwińska et al. 1999). Although Hairy can directly recruit the Sirt1 histone deacetylases to repress transcription (Rosenberg and Parkhurst 2002; Takata and Ishikawa 2003), both Hairy and Brk associate with the co-repressors Groucho (Gro; TLE in mammals) and C-terminal binding protein (CtBP) (Paroush et al. 1994; Poortinga et al. 1998; Hasson et al. 2001; Zhang et al. 2001; Morel et al. 2001; Barolo et al. 2002; Nagel et al. 2005). Groucho functions downstream of key signalling pathways such as Wg/Wnt and Dpp/TGF-beta and

391 mediates deacetylation of histones through recruitment of HDAC1. In *A. mellifera* it is also a
392 differentially DNA methylated gene (Foret et al. 2012). CtBP is an NAD(H)-regulated
393 transcription factor that functions through the recruitment of a variety of histone modification
394 enzymes or via the inhibition of CBP in order to repression transcription (reviewed in
395 (Chinnadurai 2007)), although there is evidence that CtBP can also activate transcription in
396 certain contexts (Phippen et al. 2000; Fang et al. 2006). Significantly and in stark contrast to
397 worker, in 96hQ-specific H3K27ac regions transcription factor motifs were centred on peaks
398 of H3K27ac, indicating the presence of a nucleosome and therefore potentially preventing
399 accessibility and subsequent repressor function. This was equally true for Mad/SMAD1
400 motifs which were also enriched in 27% of 96hQ-specific H3K27ac regions. In *Drosophila*,
401 Mad and Brk mediate opposing transcriptional effects in the BMP signalling pathway, in part
402 by competing for binding to overlapping sites at certain enhancers during different
403 development stages (Kirkpatrick et al. 2001). Previous studies have demonstrated
404 enrichment of Brinker sites at regions of H3K27ac in dorsal ectoderm enhancers during
405 development and that they are occupied by Mad (Koenecke et al. 2016), therefore Mad may
406 function in a similar way at queen intronic H3K27ac regions.
407 Taken together, we speculate that 96hW-specific intronic H3K27ac regions bear all the
408 hallmarks of active enhancers. Furthermore, the majority of worker genes that are enriched
409 for intronic H3K27ac are also transcription factors, suggesting further downstream gene
410 expression cascades during worker caste development. 96hQ-specific regions could also be
411 caste-specific enhancers but require further characterisation. Therefore, we conclude that it
412 is highly likely that H3K27ac and CBP play an important role at this key developmental stage
413 in honey bee caste development through differential enhancer activation. This is further
414 supported by elegant studies in the carpenter ant *Camponotus floridanus*, where CBP-
415 catalysed H3K27ac has been shown to be essential in establishing different worker castes
416 and behaviours (Simola et al. 2013, 2016). We conclude that chromatin modifications play a
417 crucial role in defining worker and queen honey bee castes by establishing and orchestrating

418 caste-specific transcriptional networks and that furthermore, it is the worker developmental
419 pathway that is actively switched on from a default queen developmental programme.
420
421 **Methods**
422
423 **Age-matched larvae collection**
424 To obtain worker larvae of known age, frames containing eggs and larvae were removed
425 from healthy hives and placed as soon as possible in an incubator at 35°C and ~80%
426 humidity, in a warm room (at least 30°C). After removal from the incubator, frames were
427 closely examined and an acetate sheet was laid over the patches of newly hatched larvae.
428 The position of the sheet was marked on the edges of the frame to ensure accurate
429 replacement at the time of collection. The locations of individual newly hatched larvae were
430 marked on the sheet and the frames were immediately returned to their original hives. The
431 exact time was recorded. After 96h, the frames were transferred back to a warm room and
432 the larvae were collected with blunt nose soft forceps using the acetate sheet positioned
433 over the frame according to the previously marked reference points. To obtain queen larvae
434 of known age, standard queen raising techniques were used (Evans et al. 2013). Double
435 grafting gave improved results and priming the queen cups with warm (35°C) royal jelly
436 increased yield.
437
438 **Chromatin preparation**
439 Chromatin was extracted from the larval heads (approximately 1.6 mm of the frontal end
440 dissected in PBS) containing brain, optic and retinular ganglia, neurosecretory cells, glands
441 (corpora allata, corpora cardiaca), suboesophageal ganglion, a small number of fat bodies,
442 the maxillae, labium and mandibles, segmented imaginal antennae developing in
443 hypodermal pockets, the openings of silk glands ducts at the tip of the labium-hypopharynx,
444 trachea and cuticle. The rest of the larval body is predominantly occupied by a large
445 digestive system filled with processed food and bacteria (larvae do not defecate), a tracheal

446 network and reproductive parts that at this stage of development are already large in queens
447 and rudimentary in workers. Chromatin was cross-linked in fixation buffer (50 mM HEPES
448 pH 7.9 / 1 mM EDTA pH 8 / 0.5 mM EGTA pH 8 / 100 mM NaCl / 1.8% formaldehyde) for 15
449 minutes. Reactions were quenched by washing twice in stop solution (125 mM glycine /
450 0.01% Triton X-100 / PBS). Afterwards, fixed larval heads were washed four times with wash
451 buffer 1 (10 mM HEPES pH 8 / 10 mM EDTA pH 8 / 0.5 mM EGTA pH 8 / 0.25% Triton X-
452 100), followed by four washes with wash buffer 2 (10 mM HEPES pH 8 / 1 mM EDTA pH 8 /
453 0.5 mM EGTA pH 8 / 0.01% Triton X-100 / 200 mM NaCl). Larval heads were homogenised
454 and centrifuged (1200 ×g / 10 min). All buffers from this step onwards contained Complete
455 protease inhibitors cocktail (Roche) and 2.5 mM sodium butyrate. Pellets were suspended in
456 5 ml of lysis buffer 1 (2.5% glycerol / 50 mM Tris-HCl pH 8.0 / 140 mM NaCl / 0.5% IGEPAL
457 / 0.25% Triton X-100 / 1 mM EDTA) and incubated for 30 min on a rotator mixer at 4˚C.
458 Lysates were then centrifuged (1200 ×g / 10 min) and pellets were suspended in 5 ml of
459 lysis buffer 2 (10 mM Tris-HCl pH 8.0 / 200 mM NaCl / 1 mM EDTA) and incubated for
460 another 30 min on a rotator mixer at 4˚C. Lysates were centrifuged (1200 ×g / 10 min) and
461 pellets were suspended in 900 µl of sonication buffer (10 mM Tris-HCl pH 8.0 / 0.5% SDS / 1
462 mM EDTA). Chromatin was sonicated using a Bioruptor (3 ×15 min cycles [30 s on / 30 s
463 off] at high power). Lysates were cleared by centrifugation (20000 ×g / 10 min) and
464 sonication checked by agarose gel electrophoresis.
465
466 **Chromatin immunoprecipitation and ChIP-seq library preparation**
467 Chromatin was concentrated using a Millipore Amicon concentrator (3 kDa cut-off). 50 µl
468 chromatin aliquots were diluted 10-fold with ChIP dilution buffer (0.01% SDS / 1.1 % Triton
469 X-100 / 1.2 mM EDTA / 16.7 mM Tris-HCl pH 8.0 / 167 mM NaCl). Antibodies (H3K4me3
470 [Active Motif, 39159]; H3K27ac [Active Motif, 39133] and H3K36me3 [Abcam, ab9050]) were
471 added according to manufacturers' instructions and samples were incubated overnight on a
472 rotator mixer at 4˚C. 30 µl of magnetic protein A Dynabeads (Invitrogen) were added to each

reaction and samples were incubated for 4 h on a rotator mixer at 4˚C. Beads were washed twice with 500 µl of wash buffer A (50 mM Tris-HCl pH 8.0 / 150 mM NaCl / 1mM EDTA / 0.1% SDS / 1% NP-40 / 0.5% sodium deoxycholate), and once with 500 µl of wash buffer B (50 mM Tris-HCl pH 8.0 / 500mM NaCl / 1mM EDTA / 0.1% SDS / 1% NP-40 / 0.5% sodium deoxycholate) and wash buffer C (50 mM Tris-HCl pH 8.0 / 250 mM LiCl / 1mM EDTA / 1% NP-40 / 0.5% sodium deoxycholate). After these washes, DNA was eluted from the beads by incubation in 200 µl of elution buffer (1% SDS / 100 mM $NaHCO_3$) for 40 min at 65˚C in a ThermoMixer. 50 µg of RNase A was added and samples were incubated for 15 min at 37˚C. NaCl was then added to a final concentration of 500 mM. Afterwards, the samples were incubated overnight with 40 µg of proteinase K at 65˚C in a ThermoMixer. DNA was purified with GeneJET PCR Purification columns (ThermoFisher). NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) was used to make sequencing libraries from 0.5-1 ng of DNA following manufacturer's instructions. 75bp SE next generation DNA sequencing was carried out by the Queen Mary University of London Genome Centre on the Illumina NextSeq 500 platform.

RNA isolation and RNA-seq library preparation

Larval heads were dissected and individually snap frozen in liquid nitrogen. Total RNA from individual heads was isolated using the TRIzol method, followed by the use of RNA Clean-up & Concentration kit (Zymo Research). mRNA was isolated with poly(A) mRNA Magnetic Isolation Module (NEB) from 1 µg of total RNA. RNA-seq libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) following manufacturer's instructions. 100bp PE next generation sequencing was carried out by the Queen Mary University Genome Centre on the Illumina NextSeq 500 platform.

ChIP-seq analysis

The genome assembly Amel_4.5 (GCF_000002195.4) was downloaded from the NCBI and indexed using Bowtie 2 (v2.2.8) (Langmead and Salzberg 2012). ChIP-seq samples were

mapped to this indexed genome using Bowtie 2 with default parameters. Detailed mapping statistics for each sample is available in Supplemental Table S7. Reads were counted into windows of width 100bp with a spacing of 50bp (each window therefore overlaps 2 other windows) using csaw (Lun and Smyth 2015). Duplicate reads were included in the analysis, mapping quality was restricted to >=20 and each read was extended to 250bp. Input reads were counted using the same parameters except each region was expanded to 5000bp (+/- 2450bp). Windows which could not be expanded (e.g. those at the end of contigs) were removed from analysis. The count value of each window (C_i) for each sample, i, was then normalized to the background counts (C_b) using **equation 1**

$$NC_i = \frac{C_i N_b}{C_b N_i}$$

Equation 1

where N_i is the total number of reads sequenced for each sample, i, and N_b is the total number of reads sequenced in the appropriate input sample. To determine differential windows, we computed moderated t-statistics using empirical Bayes moderation of the standard errors with the *limma* R package (Ritchie et al. 2015).

RNA-seq analysis

The cDNA of reference transcripts and ncRNA were downloaded from EnsemblMetazoa in FASTA format using genome version GCA_000002195.1. These two FASTA files were concatenated and supercontigs were removed using linux command grep with the following string “supercontig|'"$genome_version"':[^1-9XMY]”. Kallisto (Bray et al. 2016) was used to build an index for further mapping using default parameters. Each sample’s FASTQ file was mapped using Kallisto quant with default parameters except for increasing the number of bootstrap samples to 100 and setting the strand specific nature of the reads using parameters “-b 100 --rf-stranded”. Detailed mapping statistics for each sample is available in Supplemental Table S8. To determine differential expression, the resulting files from the

mapping were used with the R program sleuth (Pimentel et al. 2017). Default parameters were used throughout analysis. Sleuth uses a likelihood ratio test and hence we tested for those genes whose abundance is significantly better explained when caste is included in the model compared to a reduced model in which a single parameter is fitted for each gene.

Gene Ontology analysis

Drosophila melanogaster GO terms were downloaded from FlyBase [http://flybase.org]. Each *D. melanogaster* gene was mapped to its *A. mellifera* orthologue using HymenopteraMine [http://hymenopteragenome.org/hymenopteramine/begin.do]. GO analysis was performed using the R package; topGO (Alexa and Rahnenfuhrer 2016).

Transcription Factor motif analysis

Transcription factor motif analysis was performed using CentriMo (Bailey and MacHanick 2012) which is part of the MEME suite tools**.** Individual peaks of H3K27ac were extended to 5kbp from the centre of the called peak.

Data Access

ChIP-seq and RNA-seq data from this study have been submitted to the National Centre for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under accession number [GEO:GSE110642].

Acknowledgements

This research utilised the Bart's and the London Genome Centre and Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT (http://doi.org/10.5281/zenodo.438045). This work was funded by a grant to P.J.H. from the Biotechnology and Biological Sciences Research Council (BBSRC; BB/L023164/1). M.W. received funding from the William Harvey International Translational Research Academy COFUND Marie Curie Actions (WHRI-ACADEMY; PCOFUND-GA-2013-608765). Work in R.M. lab was supported by the

Australian Research Council grant DP160103053. We thank Paul Helliwell for bee keeping support and Dr Özgen Deniz for critical reading of the manuscript.

Authors' contributions

M.W. performed the chromatin, RNA and library preparation work. R.L. carried out all the computational analyses. J.M. carried out all the larval collections, dissections and tissue fixing. D.C. performed some of the ChIP-seq library preparation. R.M. and P.J.H. conceived and designed the study and supervised the work. R.L. and R.M. participated in drafting the manuscript. P.J.H. wrote the manuscript.

Disclosure declaration

The authors declare that they have no competing interests.

References

Alexa A, Rahnenfuhrer J. 2016. topGO: Enrichment Analysis for Gene Ontology.

Ashby R, Forêt S, Searle I, Maleszka R. 2016. MicroRNAs in Honey Bee Caste Determination. *Sci Rep* **6**: 18794.

Badenhorst P, Xiao H, Cherbas L, Kwon SY, Voas M, Rebay I, Cherbas P, Wu C. 2005. The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. *Genes Dev* **19**: 2540–2545.

Bailey TL, MacHanick P. 2012. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res* **40**: 1–10.

Barchuk AR, Cristino AS, Kucharski R, Costa LF, Simões ZL, Maleszka R. 2007. Molecular determinants of caste differentiation in the highly eusocial honeybee Apis mellifera. *BMC Dev Biol* **7**: 70.

Barolo S, Stone T, Bang AG, Posakony JW. 2002. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to

582 suppressor of hairless. *Genes Dev* **16**: 1964–1976.
583 Boija A, Mahat DB, Zare A, Holmqvist PH, Philip P, Meyers DJ, Cole PA, Lis JT, Stenberg P,
584 Mannervik M. 2017. CBP Regulates Recruitment and Release of Promoter-Proximal
585 RNA Polymerase II. *Mol Cell* **68**: 491–503.
586 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq
587 quantification. *Nat Biotechnol* **34**: 525–527.
588 Calo E, Wysocka J. 2013. Modification of Enhancer Chromatin: What, How, and Why? *Mol*
589 *Cell* **49**: 825–837.
590 Cameron RC, Duncan EJ, Dearden PK. 2013. Biased gene expression in early honeybee
591 larval development. *BMC Genomics* **14**: 1–12.
592 Campbell G, Tomlinson A. 1999. Transducing the Dpp Morphogen Gradient in the Wing of
593 Drosophila: Regulation of Dpp Targets by brinker. *Cell* **96**: 553–562.
594 Chen X, Hu Y, Zheng H, Cao L, Niu D, Yu D, Sun Y, Hu S, Hu F. 2012. Transcriptome
595 comparison between honey bee queen- and worker-destined larvae. *Insect Biochem*
596 *Mol Biol* **42**: 665–673.
597 Chinnadurai G. 2007. Transcriptional regulation by C-terminal binding proteins. *Int J*
598 *Biochem Cell Biol* **39**: 1593–1607.
599 Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J,
600 Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active
601 from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**:
602 21931–21936.
603 Deignan L, Pinheiro MT, Sutcliffe C, Saunders A, Wilcockson SG, Zeef LAH, Donaldson IJ,
604 Ashe HL. 2016. Regulation of the BMP Signaling-Responsive Transcriptional Network
605 in the Drosophila Embryo. *PLoS Genet* **12**: 1–23.
606 Dickman MJ, Kucharski R, Maleszka R, Hurd PJ. 2013. Extensive histone post-translational
607 modification in honey bees. *Insect Biochem Mol Biol* **43**: 125–137.
608 Evans JD, Schwarz RS, Chen YP, Budge G, Cornman RS, De la Rua P, de Miranda JR,
609 Foret S, Foster L, Gauthier L, et al. 2013. Standard methods for molecular research in

Apis mellifera. *J Apic Res* **52**: 1–54.

Evans JD, Wheeler DE. 2001. Expression profiles during honeybee caste determination. *Genome Biol* **2**: RESEARCH0001.

Evans JD, Wheeler DE, Universitaria C, Rica C. 1999. Differential gene expression between developing queens and workers in the honey bee, Apis mellifera. *Proc Natl Acad Sci U S A* **96**: 5575–80.

Fang M, Li J, Blauwkamp T, Bhambhani C, Campbell N, Cadigan KM. 2006. C-terminal-binding protein directly activates and represses Wnt transcriptional targets in Drosophila. *EMBO J* **25**: 2735–2745.

Foret S, Kucharski R, Pellegrini M. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc Natl Acad Sci U S A* **109**: 4968–4973.

Gordon D. 2002. The organization of work in social insect colonies. *Complexity* **380**: 43–46.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* **130**: 77–88.

Hasson P, Müller B, Basler K, Paroush Z. 2001. Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. *EMBO J* **20**: 5725–5736.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318.

Hrassnigg N, Crailsham K. 2007. Differences in drone and worker physiology in honeybees (Apis mellifera). *Apidologie* **36**: 255–277.

Jarvis J. 1981. Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science (80-)* **212**: 571–573.

Jaźwińska A, Kirov N, Wieschaus E, Roth S, Rushlow C. 1999. The Drosophila gene brinker reveals a novel mechanism of Dpp target gene regulation. *Cell* **96**: 563–573.

Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SYR, Ge K.

2011. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* **30**: 249–262.

Kirkpatrick H, Johnson K, Laughon A. 2001. Repression of Dpp Targets by Binding of Brinker to Mad Sites. *J Biol Chem* **276**: 18216–18222.

Koenecke N, Johnston J, Gaertner B, Natarajan M, Zeitlinger J. 2016. Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol* **17:** 196.

Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**: 1827–1830.

Kucharski R, Maleszka J, Maleszka R, Rodin S, Riggs A, Lynch M, Conery J, Flores K, Wolschin F, Corneveaux J, et al. 2016. A possible role of DNA methylation in functional divergence of a fast evolving duplicate gene encoding odorant binding protein 11 in the honeybee. *Proc Biol Sci* **283**: 718–729.

Kwon SY, Grisan V, Jang B, Herbert J, Badenhorst P. 2016. Genome-Wide Mapping Targets of the Metazoan Chromatin Remodeling Factor NURF Reveals Nucleosome Remodeling at Enhancers, Core Promoters and Gene Insulators. *PLoS Genet* **12**: 1–26.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Li W, Notani D, Rosenfeld MG. 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* **17**: 207–223.

Lun ATL, Smyth GK. 2015. Csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**: e45.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes: Differential methylation of brain DNA in queens and workers. *PLoS Biol* **8**: e1000506

Maleszka R. 2018. Beyond Royalactin and a master inducer explanation of phenotypic plasticity in honey bees. *Commun Biol* 1–7.

Morel V, Lecourtois M, Massiani O, Maier D, Preiss A, Schweisguth F. 2001. Transcriptional repression by Suppressor of Hairless involves the binding of a Hairless-dCtBP complex in Drosophila. *Curr Biol* **11**: 789–792.

Nagel AC, Krejci A, Tenin G, Bravo-Patiño A, Bray S, Maier D, Preiss A. 2005. Hairless-mediated repression of notch target genes requires the combined activity of Groucho and CtBP corepressors. *Mol Cell Biol* **25**: 10433–10441.

Nalepa CA. 2015. Origin of termite eusociality: Trophallaxis integrates the social, nutritional, and microbial environments. *Ecol Entomol* **40**: 323–335.

Négre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, et al. 2011. A cis-regulatory map of the Drosophila genome. *Nature* **471**: 527–531.

Okada M, Hirose S. 1998. Chromatin remodeling mediated by Drosophila GAGA factor and ISWI activates fushi tarazu gene transcription in vitro. *Mol Cell Biol* **18**: 2455–2461.

Paroush Z, Finley RL, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. 1994. Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**: 805–815.

Pearson KL, Hunter T, Janknecht R. 1999. Activation of Smad1-mediated transcription by p300/CBP. *Biochim Biophys Acta - Gene Struct Expr* **1489**: 354–364.

Philip P, Boija A, Vaid R, Churcher AM, Meyers DJ, Cole PA, Mannervik M, Stenberg P. 2015. CBP binding outside of promoters and enhancers in Drosophila melanogaster. *Epigenetics and Chromatin* **8**: 1–12.

Phippen TM, Sweigart AL, Moniwa M, Krumm A, Davie JR, Parkhurst SM. 2000. Drosophila C-terminal binding protein functions as a context-dependent transcriptional co-factor and interferes with both Mad and Groucho transcriptional repression. *J Biol Chem* **275**: 37628–37637.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* **14**: 687–690.

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Tong IL, Bell GW, Walker K,

694 Rolfe PA, Herbolsheimer E, et al. 2005. Genome-wide map of nucleosome acetylation
695 and methylation in yeast. *Cell* **122**: 517–527.
696 Poortinga G, Watanabe M, Susan SM. 1998. Drosophila CtBP: A Hairy-interacting protein
697 required for embryonic segmentation and Hairy-mediated transcriptional repression.
698 *EMBO J* **17**: 2067–2078.
699 Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique
700 chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**:
701 279–283.
702 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers
703 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic*
704 *Acids Res* **43**: e47.
705 Rosenberg MI, Parkhurst SM. 2002. Drosophila Sir2 is required for heterochromatic
706 silencing and by euchromatic Hairy/E(Spl) bHLH repressors in segmentation and sex
707 determination. *Cell* **109**: 447–458.
708 Simola DF, Graham RJ, Brady CM, Enzmann BL, Desplan C, Ray A, Zwiebel LJ, Bonasio R,
709 Reinberg D, Liebig J, et al. 2016. Epigenetic (re)programming of caste-specific behavior
710 in the ant Camponotus floridanus. *Science* **351**: aac6633.
711 Simola DF, Ye C, Mutti NS, Doleza K, Bonasio R, Liebig J, Reinberg D, Berger SL. 2013. A
712 chromatin link to caste identity in the carpenter ant Camponotus floridanus. *Genome*
713 *Res* **23**: 486–496.
714 Spannhoff A, Kim YK, Raynal NJM, Gharibyan V, Su MB, Zhou YY, Li J, Castellano S,
715 Sbardella G, Issa JPJ, et al. 2011. Histone deacetylase inhibitor activity in royal jelly
716 might facilitate caste switching in bees. *EMBO Rep* **12**: 238–243.
717 Takata T, Ishikawa F. 2003. Human Sir2-related protein SIRT1 associates with the bHLH
718 repressors HES1 and HEY2 and is involved in HES1- and HEY2-mediated
719 transcriptional repression. *Biochem Biophys Res Commun* **301**: 250–257.
720 Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri
721 PC, Harte PJ. 2009. CBP-mediated acetylation of histone H3 lysine 27 antagonizes

722 Drosophila Polycomb silencing. *Development* **136**: 3131–3141.
723 Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C,
724 Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers.
725 *Nature* **457**: 854–858.
726 Waltzer L, Bienz M. 1999. A function of CBP as a transcriptional co-activator during Dpp
727 signalling. *EMBO J* **18**: 1630–1641.
728 Wang Y, Ma L, Zhang W, Cui X, Wang H, Xu B. 2016. Comparison of the nutrient
729 composition of royal jelly and worker jelly of honey bees (Apis mellifera). *Apidologie* **47**:
730 48–56.
731 Weaver N. 1966. Physiology of Caste Determination. *Annu Rev Entomol* **11**: 79–102.
732 Wedd L, Kucharski R, Maleszka R. 2016. Differentially methylated obligatory epialleles
733 modulate context-dependent LAM gene expression in the honeybee Apis mellifera.
734 *Epigenetics* **11**: 1–10.
735 Winston ML. 1991. *The Biology of the Honeybee*. Second. Harvard University Press,
736 Cambridge, Massachusetts.
737 Zhang H, Levine M, Ashe HL. 2001. Brinker is a sequence-specific transcriptional repressor
738 in the Drosophila embryo. *Genes Dev* **15:** 261–266.
739
740
741 **Figure Legends**
742
743 **Figure 1**. H3K4me3, H3K27ac and H3K36me3 are associated with transcribed regions in
744 honey bee castes. *(A)* Plots of the average ChIP-seq enrichment above input around the
745 TSS (+/- 2kbp) of genes profiled across 96hW (upper panel) and 96hQ (lower panel). The
746 green and yellow lines represent the two replicates performed for each ChIP-seq experiment
747 and the grey line represents the input. *(B)* The expression distribution (shown by a coloured
748 half-bean) of transcripts enriched by > 3-fold change above input for all possible
749 combinations of H3K4me3, H3K27ac and H3K36me3, compared to a random sampling of

genes (grey half-bean). The mean of either distribution is shown by a solid black line. 96hW is shown in the top panel and 96hQ in the bottom panel. n is the number of transcripts.

Figure 2. At 96h, worker and queen larvae show caste-specific differences in the enrichment of H3K4me3, H3K27ac and H3K36me3 that correlate with differential gene expression. *(A)* A volcano plot of the difference in enrichment between 96hW and 96hQ castes against the negative log p-value for H3K4me3, H3K27ac and H3K36me3. In grey are regions which fall below the genome wide threshold of significance (p > 0.01). In blue (96hW) and red (96hQ) are those regions which reach genome wide significance (p <= 0.01) and have a larger than 3-fold difference in enrichment above input between castes. *(B)* The negative log p-value for the top 5 biological process GO terms for those genes which show increased enrichment for H3K4me3, H3K27ac or H3K36me3 in 96hW compared to 96hQ (left panel) and for those which show an increased enrichment in 96hQ compared to 96hW (right panel). *(C)* A volcano plot of the log fold change (LogFC) in transcript expression by RNA-seq between 96hW and 96hQ castes against the negative log p-value. In grey are transcripts which fall below the genome wide threshold of significance (p > 0.01) and in blue are those transcripts which reach genome wide significance (p <= 0.01) and are more expressed in 96hW and in red are those transcripts which reach genome wide significance (p <= 0.01) and are more expressed in 96hQ. *(D)* A scatter plot of the difference in significant ChIP-seq enrichment between 96hQ and 96hW (x-axis) against the LogFC of transcript expression between 96hQ and 96hW castes (y-axis). *(E)* The negative log p-value for the top five biological process GO terms for those transcripts which show both increased expression and increased enrichment in H3K4me3, H3K27ac or H3K36me3 in 96hQ compared to 96hW (left panel) and for those which show both increased expression and increased enrichment in 96hQ compared to 96hW (right panel).

Figure 3. Profiles of two genes that show a significant difference in both ChIP-seq enrichment and transcript expression between castes. *(A)* Pyruvate kinase (*PYK*; 552007).

Regions are shown which reach genome wide significance (p < = 0.01) and have a larger than 3-fold difference in enrichment over input between 96hW (blue) and 96hQ (red). The expression profile is shown in the right panel. *(B)* Ten-eleven translocation (*TET*; 412878). Regions are shown which reach genome wide significance (p < = 0.01) and have a larger than 3-fold difference in enrichment over input between 96hW (blue) and 96hQ (red) for both TET transcripts, XM_016915488 (upper transcript) and XM_006561197 (lower transcript). The expression profile is also shown for transcripts XM_016915488 (left panel) and XM_006561197 (right panel).

Figure 4. Intronic H3K27ac regions define the worker caste and are enriched for transcription factor binding sites. *(A)* A bar plot showing the percentage of unique H3K4me3, H3K27ac and H3K36me3 ChIP-seq regions and their location relative to the nearest TSS in 96hW (blue) and 96hQ (red). *(B)* A bar plot showing the percentage of unique H3K4me3, H3K27ac and H3K36me3 ChIP-seq regions within intergenic or intragenic locations. Intragenic distributions are further defined as either exonic or intronic. Blue bar represents regions with higher enrichment in 96hW and red bar are those regions with higher enrichment in 96hQ. *(C)* In the top panel, the vertical blue line shows the number of differentially expressed genes that contain at least one peak of intronic H3K37ac and are more expressed in 96hW compared to 96hQ. A background distribution (shown in grey) was calculated by randomly selecting an identical number of genes and calculating how many of these contain at least one peak of intronic H3K27ac. This was repeated 1000 times. In the bottom panel, the vertical red line shows the number of differentially expressed genes that contain at least one peak of intronic H3K27ac and are more expressed in 96hQ compared to 96hW. A background distribution (shown in grey) was calculated as previously and repeated 1000 times. *(D)* A bar plot of the adjusted p-value for enrichment of transcription factor binding motifs located within caste-specific intronic H3K27ac regions. *(E)* A motif probability graph showing the probability of transcription factor binding motifs in relation to caste-specific intronic H3K27ac (centered at 0bp).

Figure 1

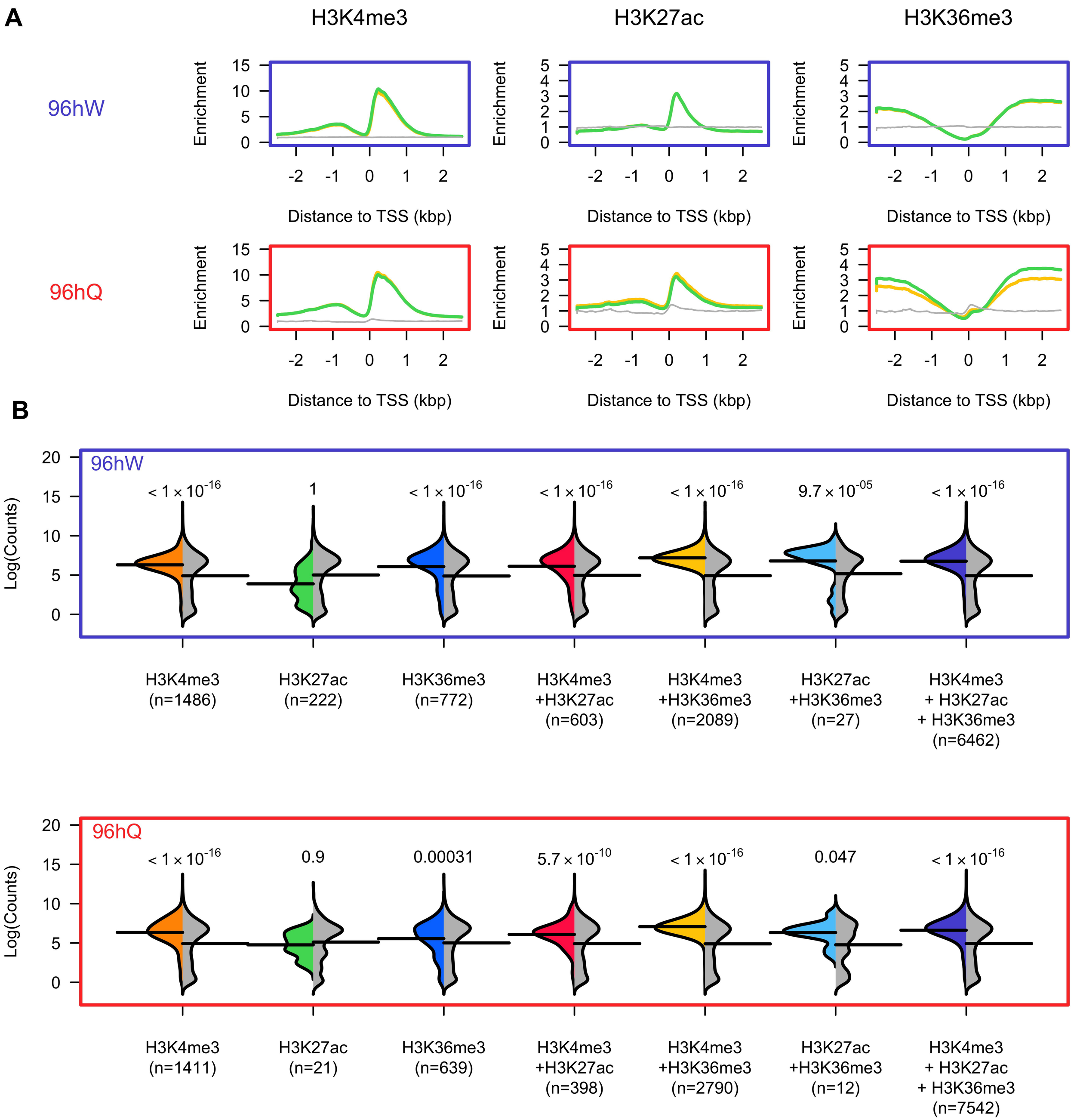

A
H3K4me3
H3K27ac
H3K36me3
96hW
96hQ
Enrichment
Distance to TSS (kbp)
B
Log(Counts)
96hW
< 1 × 10-16
1
< 1 × 10-16
< 1 × 10-16
< 1 × 10-16
9.7 × 10-05
< 1 × 10-16
H3K4me3 (n=1486)
H3K27ac (n=222)
H3K36me3 (n=772)
H3K4me3 +H3K27ac (n=603)
H3K4me3 +H3K36me3 (n=2089)
H3K27ac +H3K36me3 (n=27)
H3K4me3 + H3K27ac + H3K36me3 (n=6462)
96hQ
< 1 × 10-16
0.9
0.00031
5.7 × 10-10
< 1 × 10-16
0.047
< 1 × 10-16
H3K4me3 (n=1411)
H3K27ac (n=21)
H3K36me3 (n=639)
H3K4me3 +H3K27ac (n=398)
H3K4me3 +H3K36me3 (n=2790)
H3K27ac +H3K36me3 (n=12)
H3K4me3 + H3K27ac + H3K36me3 (n=7542)

Figure 2

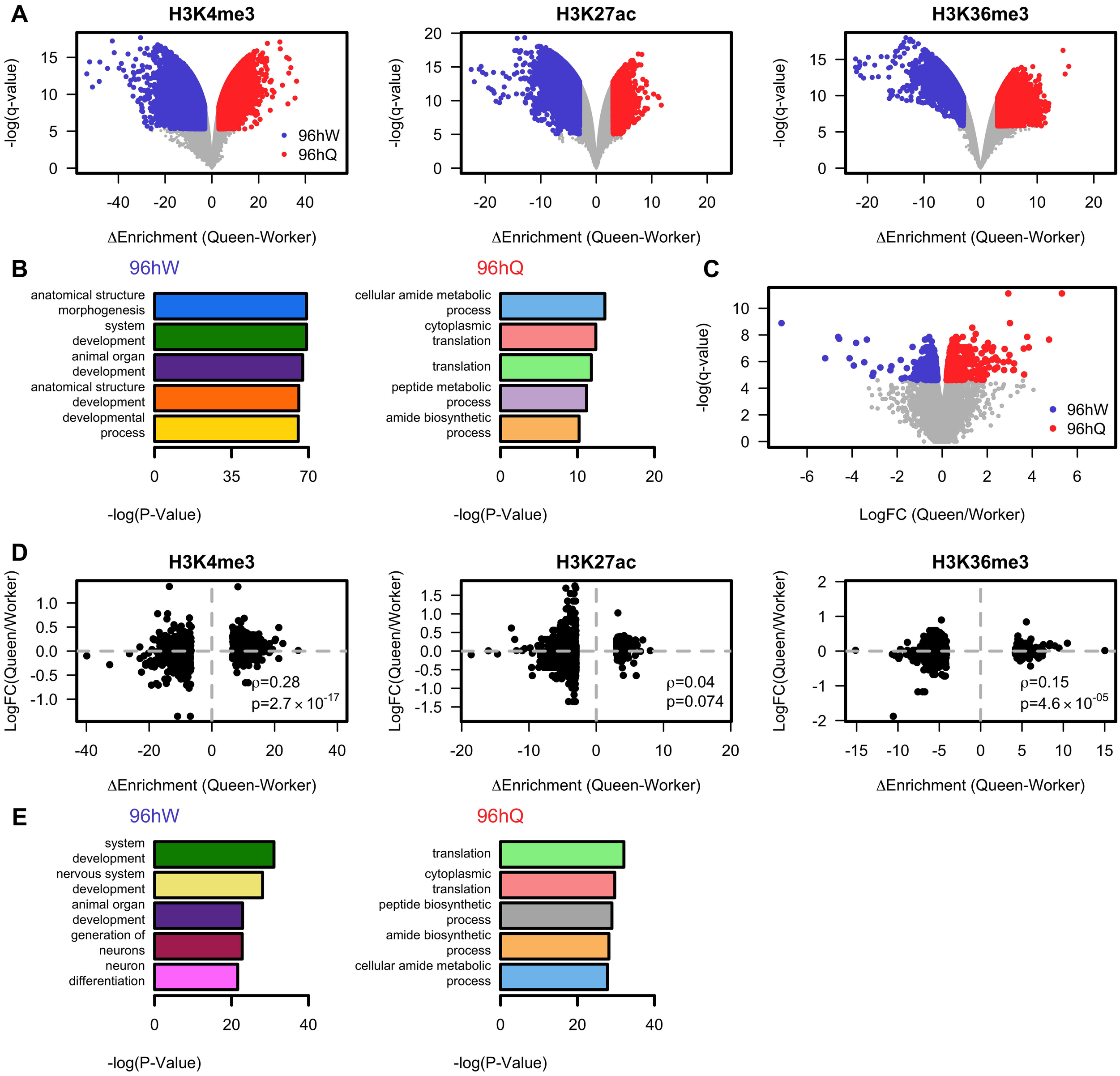
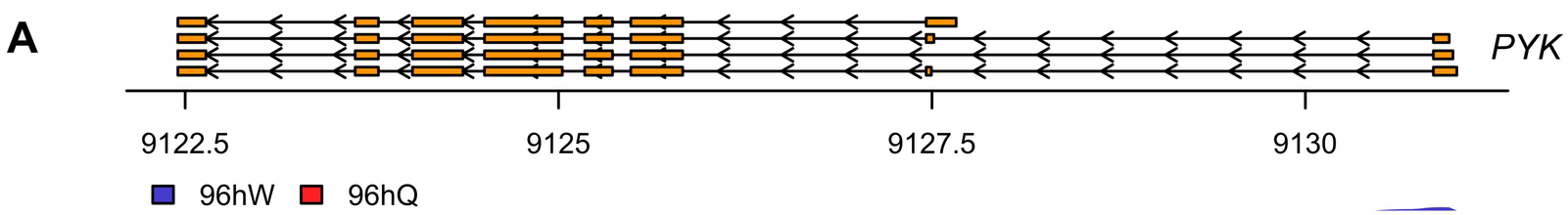
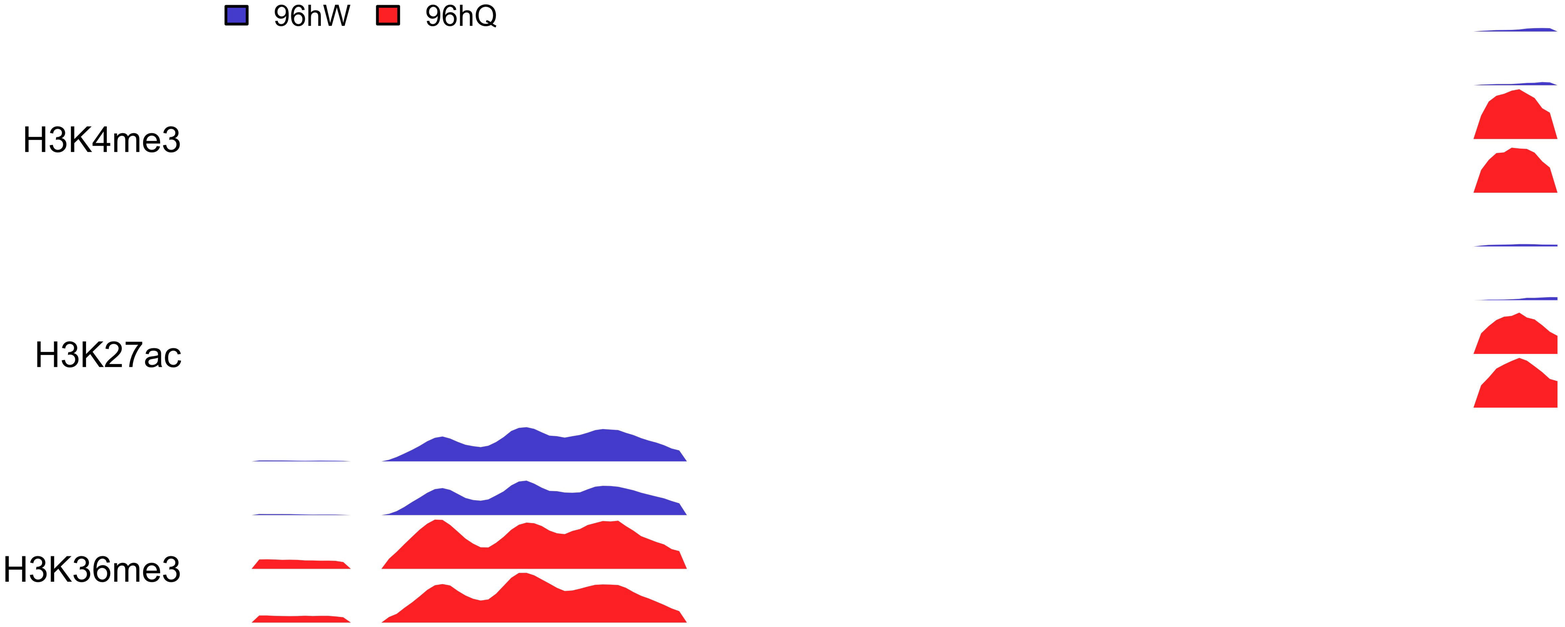
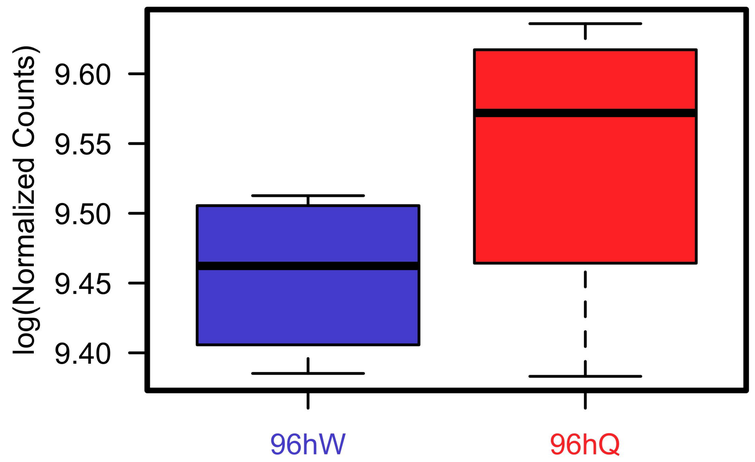

A
H3K4me3
H3K27ac
H3K36me3
-log(q-value)
96hW
96hQ
ΔEnrichment (Queen-Worker)
B
96hW
anatomical structure morphogenesis
system development
animal organ development
anatomical structure development
developmental process
-log(P-Value)
96hQ
cellular amide metabolic process
cytoplasmic translation
translation
peptide metabolic process
amide biosynthetic process
C
-log(q-value)
LogFC (Queen/Worker)
D
H3K4me3
H3K27ac
H3K36me3
LogFC(Queen/Worker)
ρ=0.28
p=2.7 × 10-17
ρ=0.04
p=0.074
ρ=0.15
p=4.6 × 10-05
ΔEnrichment (Queen-Worker)
E
96hW
system development
nervous system development
animal organ development
generation of neurons
neuron differentiation
-log(P-Value)
96hQ
translation
cytoplasmic translation
peptide biosynthetic process
amide biosynthetic process
cellular amide metabolic process

Figure 3

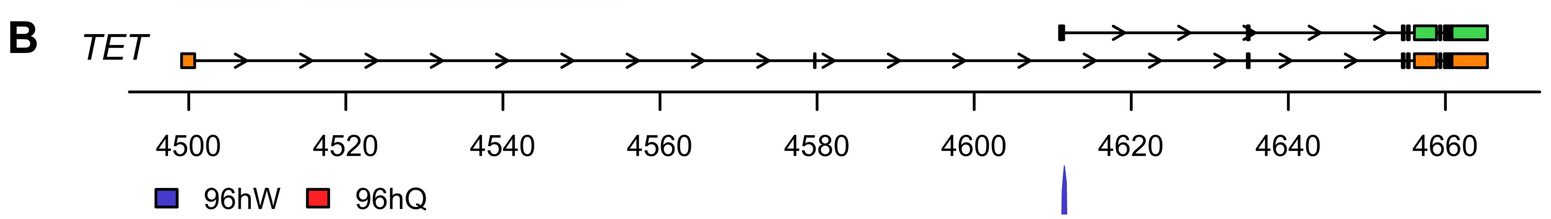
A

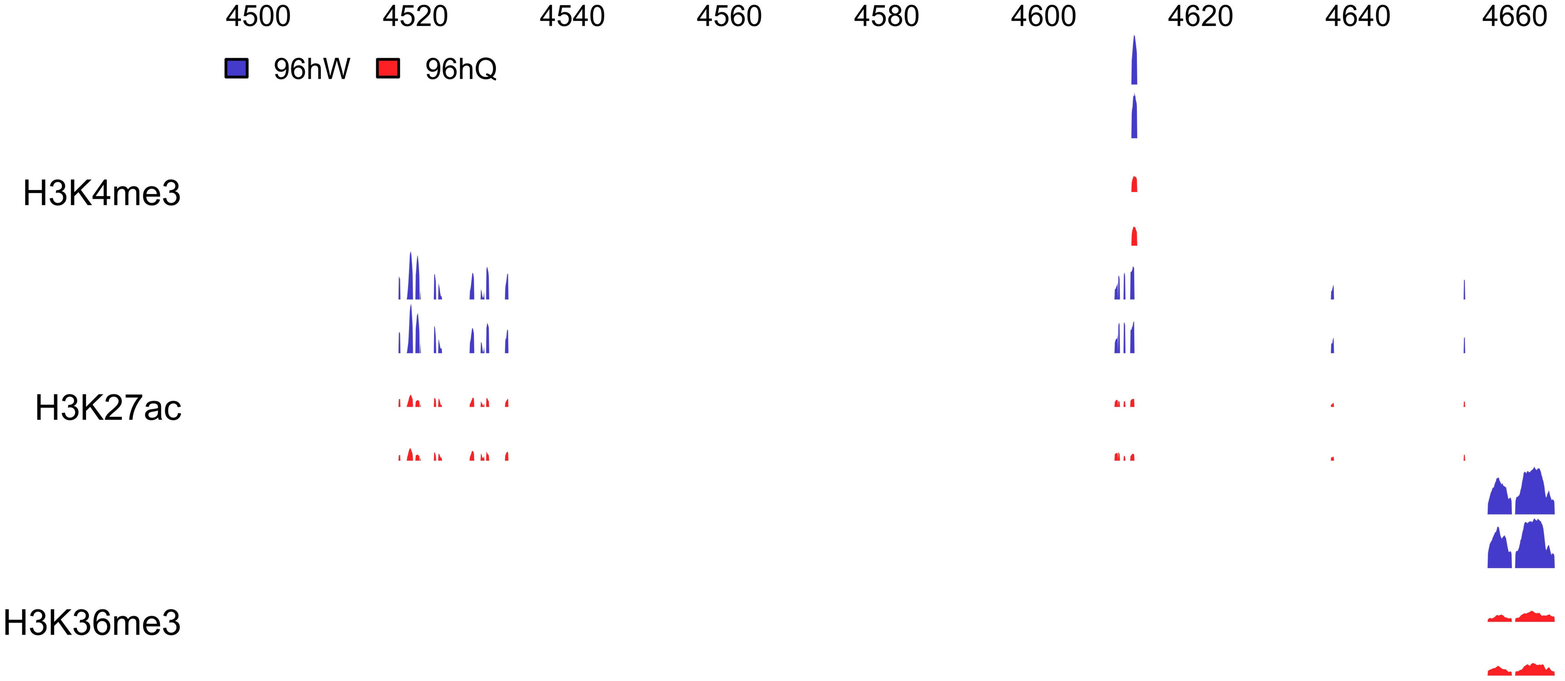

PYK
9122.5
9125
9127.5
9130
96hW
96hQ

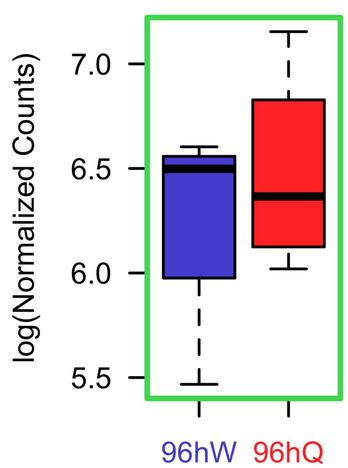

H3K4me3
H3K27ac
H3K36me3


log(Normalized Counts)
9.60
9.55
9.50
9.45
9.40
96hW
96hQ

B


TET
4500
4520
4540
4560
4580
4600
4620
4640
4660
96hW
96hQ


H3K4me3
H3K27ac
H3K36me3


log(Normalized Counts)
7.0
6.5
6.0
5.5
96hW
96hQ

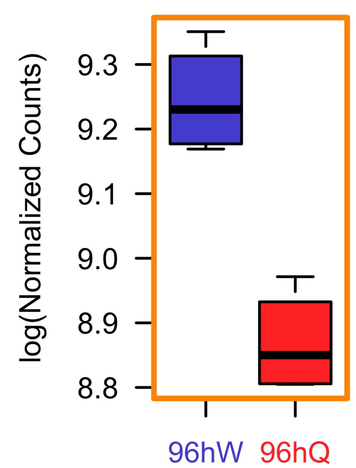
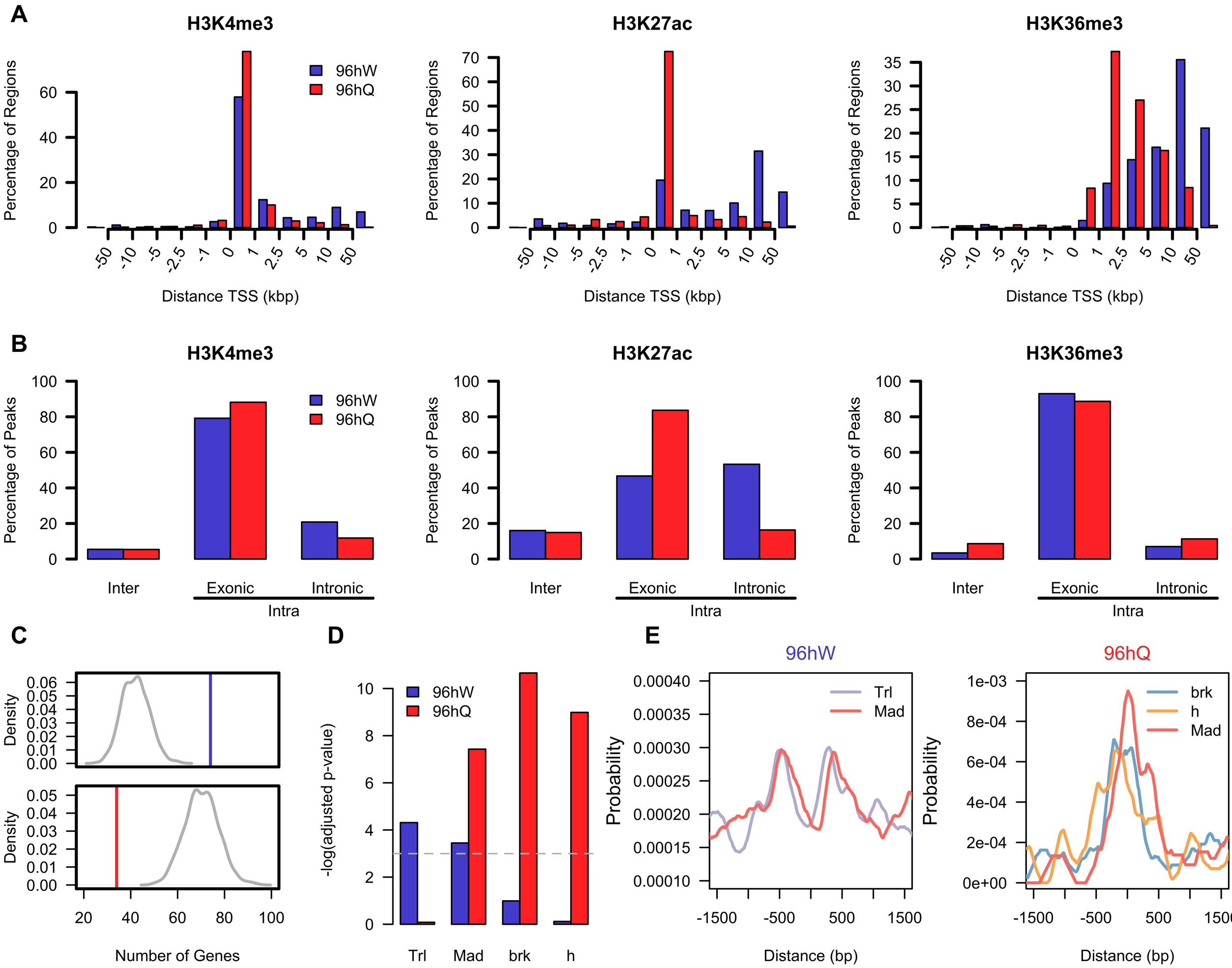

log(Normalized Counts)
9.3
9.2
9.1
9.0
8.9
8.8
96hW
96hQ

Figure 4


A
H3K4me3
H3K27ac
H3K36me3
Percentage of Regions
96hW
96hQ
Distance TSS (kbp)
B
H3K4me3
H3K27ac
H3K36me3
Percentage of Peaks
Inter
Exonic
Intronic
Intra
C
Density
Number of Genes
D
-log(adjusted p-value)
Trl
Mad
brk
h
E
96hW
96hQ
Probability
Distance (bp)