

RESEARCH

Open Access



# Characterization of the *Cannabis sativa* glandular trichome epigenome

Lee J. Conneely<sup>1,2,3</sup>, Bhavna Hurgobin<sup>1,2</sup>, Sophia Ng<sup>1,2</sup>, Muluneh Tamiru-Oli<sup>1,2</sup> and Mathew G. Lewsey<sup>1,2,3\*</sup>

## Abstract

**Background** The relationship between epigenomics and plant specialised metabolism remains largely unexplored despite the fundamental importance of epigenomics in gene regulation and, potentially, yield of products of plant specialised metabolic pathways. The glandular trichomes of *Cannabis sativa* are an emerging model system that produce large quantities of cannabinoid and terpenoid specialised metabolites with known medicinal and commercial value. To address this lack of epigenomic data, we mapped H3K4 trimethylation, H3K56 acetylation, H3K27 trimethylation post-translational modifications and the histone variant H2A.Z, using chromatin immunoprecipitation, in *C. sativa* glandular trichomes, leaf, and stem tissues. Corresponding transcriptomic (RNA-seq) datasets were integrated, and tissue-specific analyses conducted to relate chromatin states to glandular trichome specific gene expression.

**Results** The promoters of cannabinoid and terpenoid biosynthetic genes, specialised metabolite transporter genes, defence related genes, and starch and sucrose metabolism were enriched specifically in trichomes for histone marks H3K4me3 and H3K56ac, consistent with active transcription. We identified putative trichome-specific enhancer elements by identifying intergenic regions of H3K56ac enrichment, a histone mark that maintains enhancer accessibility, then associated these to putative target genes using the tissue specific gene transcriptomic data. Bi-valent chromatin loci specific to glandular trichomes, marked with H3K4 trimethylation and H3K27 trimethylation, were associated with genes of MAPK signalling pathways and plant specialised metabolism pathways, supporting recent hypotheses that implicate bi-valent chromatin in plant defence. The histone variant H2A.Z was largely found in intergenic regions and enriched in chromatin that contained genes involved in DNA homeostasis.

**Conclusion** We report the first genome-wide histone post-translational modification maps for *C. sativa* glandular trichomes, and more broadly for glandular trichomes in plants. Our findings have implications in plant adaptation and stress responses and provide a basis for enhancer-mediated, targeted, gene transformation studies in plant glandular trichomes.

**Keywords** *Cannabis sativa*, Glandular trichomes, Specialised metabolism, Multiomics, H3K4me3, H3K27me3, H3K56ac, H2A.Z, Chromatin immunoprecipitation, Gene regulation, Cis-regulatory element

\*Correspondence:

Mathew G. Lewsey  
M.Lewsey@latrobe.edu.au

<sup>1</sup>La Trobe Institute for Sustainable Agriculture and Food, La Trobe University, AgriBio Building, Bundoora, VIC 3086, Australia

<sup>2</sup>Australian Research Council Research Hub for Medicinal Agriculture, La Trobe University, AgriBio Building, Bundoora, VIC 3086, Australia

<sup>3</sup>Australian Research Council Centre of Excellence in Plants for Space, La Trobe University, Bundoora, VIC, Australia



## Background

Histone post-translational modifications and histone variants are features of plant epigenomes that influence gene regulation and, through this, tissue and cell functionality [1–3]. H3K4 trimethylation (H3K4me3) and H3K27 trimethylation (H3K27me3) are the most extensively studied histone post-translational modifications, and have opposing functions on gene expression [4–8]. H3K4me3 is deposited by trithorax complexes in the 5' untranslated region (UTR) of actively transcribed genes where it is required for RNA polymerase II pause-release and transcriptional elongation [9–11]. Conversely, H3K27me3 is deposited by polycomb repressor complex 2 (PRC2) and contributes to the formation of facultative heterochromatin, referred to as polycomb chromatin, that is enriched in the gene body of repressed genes. Furthermore, H3K27me3 PRC2 mediated gene repression operates synergistically with polycomb repressor complex 1 (PRC1) in the monoubiquitylation of the histone variant H2A.Z to repress gene expression [12, 13]. Similar to H3K27me3, H2A.Z in plants is enriched in the gene body of repressed genes [14].

Despite their antagonistic effects on gene expression, H3K4me3 and H3K27me3 may also co-localise on the same nucleosome in a phenomenon known as bivalency. It is hypothesised that bi-valent chromatin containing antagonistic histone post-translational modifications results in poised states of gene expression, whereby plants can rapidly upregulate and fine tune spatio-temporal gene expression in response to various environmental cues [15]. Vernalization is a well-studied example of H3K4me3-H3K27me3 bivalency in plants, whereby temperatures influence the ratio of H3K4me3 to H3K27me3 at a bi-valent locus associated with the floral development repressor gene *FLC* [16]. This modulates *FLC* expression through gene silencing, allowing floral development when temperature increases in spring.

Enhancer elements are short stretches of DNA that contain transcription factor binding motifs of approximately 7–22 bp, found in both intergenic and intragenic regions [17, 18]. Interactions between enhancer elements and gene promoters drive gene transcription, and individual enhancers can regulate single or multiple target genes [19]. Intergenic enhancers may regulate expression of proximal ( $\leq 1.5$  Kbp) genes, or distal genes that are located from thousands to a million base pairs away (termed distal enhancer elements) [20–22]. Distal enhancer elements mediate gene expression via long-range chromatin interactions known as chromatin looping [22, 23]. Enhancers can drive tissue and cell-type specific transcription, making them a useful resource to understand patterns of gene regulation and a useful tool in biotechnology where precise expression in certain cells may be required [24, 25].

The chromatin features specific to enhancers can be used to discover novel regulatory elements [26–28]. For example, plant distal enhancer elements are typically flanked by narrow regions of H3K56 acetylation (H3K56ac) in the tissue or cell type where they are active, a mark that promotes a euchromatic environment and permits transcription factor binding [29, 30]. H3K56ac is also enriched in the promoters of transcribed genes, similar to H3K4me3, where it promotes gene expression [29]. By mapping and analysing the genome wide distribution of H3K56ac, then cross-referencing it with the distribution of genes and intergenic space, undiscovered putative distal enhancer elements can be identified.

*Cannabis sativa* is a predominantly dioecious, dicotyledonous, plant of the Cannabaceae family that has been cultivated for at least 2700 years for both its long durable fibres and its medicinal properties [31–33]. *C. sativa* is known for the production of cannabinoids, a group of species-specific terpenophenolic compounds noted with medicinal and psychoactive properties [34, 35]. The organs that produce specialised metabolites like cannabinoids, the glandular trichomes, have garnered the interest of the plant science community as potential new models for the study of plant cell and tissue development [36]. Most recently there has been interest in engineering of glandular trichomes, due to the remarkable quantities of specialised metabolite they are able to produce and sequester away from other plant tissues (e.g. THC 14.98  $\pm$  2.23% dry weight) [37–40]. However, for glandular trichome engineering to be feasible in *C. sativa* we must understand and be able to manipulate trichome-specific gene expression and specialised metabolism. This would be enabled by mapping epigenomic features, such as histone modifications and variants, and relating them to gene expression characteristics. There is currently a lack of any such data in *C. sativa* or glandular trichomes more broadly.

In this study we generated genome-wide maps of histone post translational modifications H3K4me3, H3K27me3, H3K56ac, and the histone variant H2A.Z for *C. sativa* glandular trichomes, leaves and stems. We made comparisons between tissues to identify trichome specific epigenomic signals. Matched transcriptomic (RNA-seq) datasets were generated for each tissue type to validate histone function in relation to gene expression and to enable functions associated with different epigenome features to be analysed. The relationships of these functions with plant specialised metabolism, abiotic and biotic stress resistance were assessed. Lastly, putative glandular trichome specific distal enhancer elements were predicted by examining H3K56ac data across tissues and drawing comparisons to differential gene expression analysis.

## Results

### Mapping histone marks in three *C. sativa* tissues by chromatin immunoprecipitation sequencing

We first generated genome wide maps of histone post-translational modifications H3K56ac, H3K4me3, H3K27me3, and histone variant H2A.Z, in *C. sativa* tissues. A chromatin immunoprecipitation sequencing (ChIP-seq) protocol was first optimised for cannabis tissues, to avoid interference with procedures by the extreme concentrations of specialized metabolites (Additional File 1: Fig S1). The protocol was then applied to glandular trichome (Fig. 1a), stem (internode), and vegetative leaf tissues and sequencing libraries generated (Additional File 1: Fig. S2, Table S1, Table S2). Alignment of the sequencing data to the *C. sativa* reference genome was successful (Additional File 1: Table S3a). Quality control fingerprint plots were consistent with the expected distributions for narrow peak H3K4me3 and H3K56ac and broad peak H3K27me3 and H2A.Z data, respectively (Additional File 1; Fig. S3). Corresponding RNA-seq libraries were generated and analysed in triplicate for integration with the ChIP-seq data (Additional File 1: Fig. S4, Table S3b).

### Associations of histone marks with gene expression in *C. sativa*

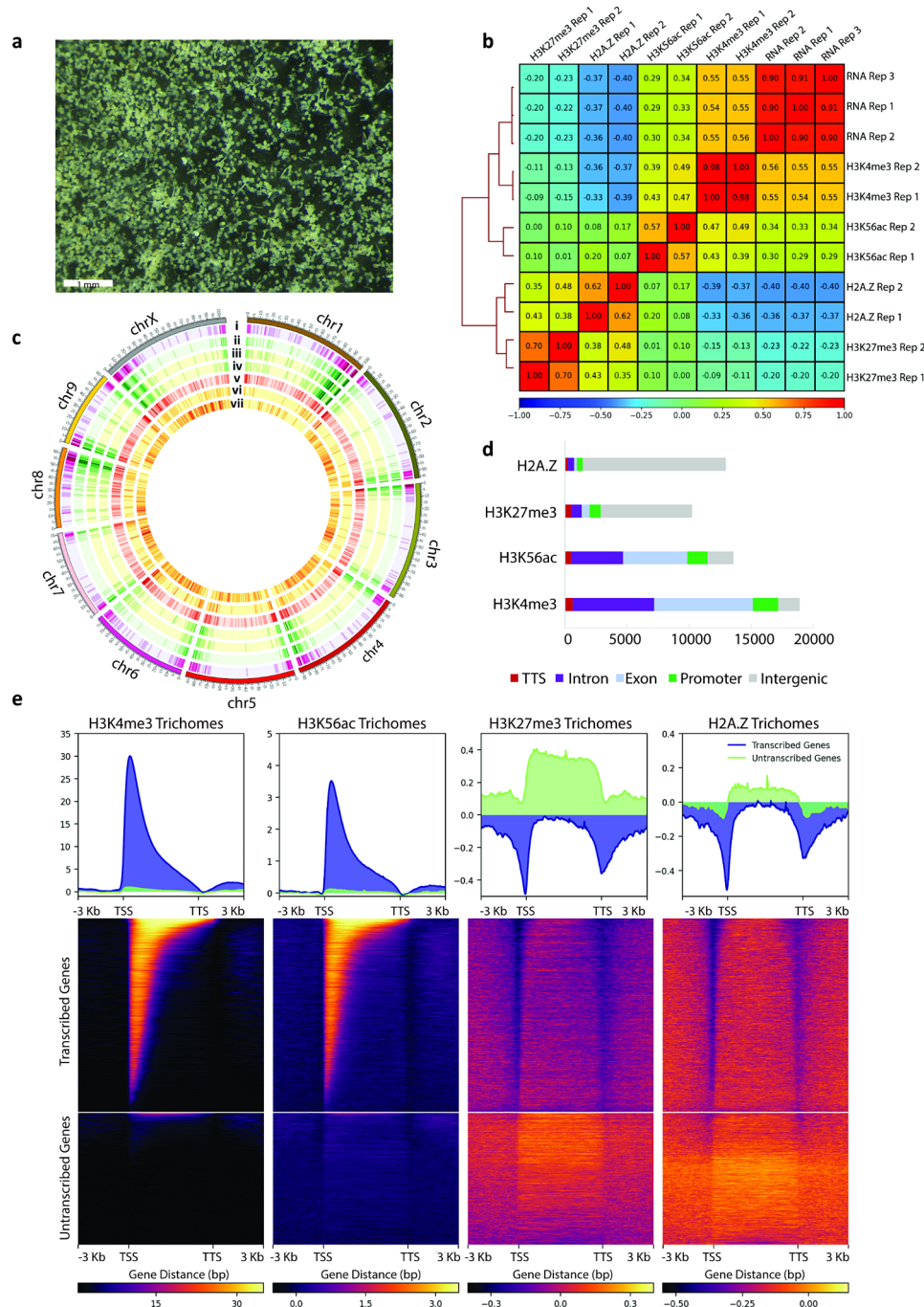
Characteristic association between histone marks and gene expression have been observed across many plant species [41]. We examined if these associations were conserved in *C. sativa* (Fig. 1b-c, Additional File 1: Fig. S5). The distribution of RNA-seq reads – indicative of gene expression - positively correlated with the distributions of ChIP-Seq reads for H3K56ac, H3K4me3 in each tissue (Spearman correlation coefficient, Fig. 1b). Contrastingly, H3K27me3 and H2A.Z ChIP-Seq reads tended to negatively correlate with RNA-seq reads. Consistent with these results, actively transcribed genes were observed in a genome wide visualization overlaying reads for both H3K4me3 and H3K56ac, whilst un-transcribed genes overlaid H3K27me3 and H2A.Z (Fig. 1c). The majority of H3K27me3 and H2A.Z ChIP-seq reads mapped to intergenic regions, whereas the majority of H3K4me3 and H3K56ac reads mapped to gene features (Fig. 1d). H3K4me3 and H3K56ac were enriched at the transcriptional start site of actively transcribed genes, consistent with their roles in facilitating transcription [9, 42, 43]. Conversely, H3K27me3 and H2A.Z were enriched in the gene bodies of un-transcribed genes, consistent with their synergistic roles in polycomb mediated gene repression [14, 44] (Fig. 1e, Additional File 1: Fig. S6). Overall, the distribution of histone marks was consistent with observations in other species [30, 43].

### Histone marks co-localise in the *C. sativa* genome which may reflect their influence on gene activation or repression

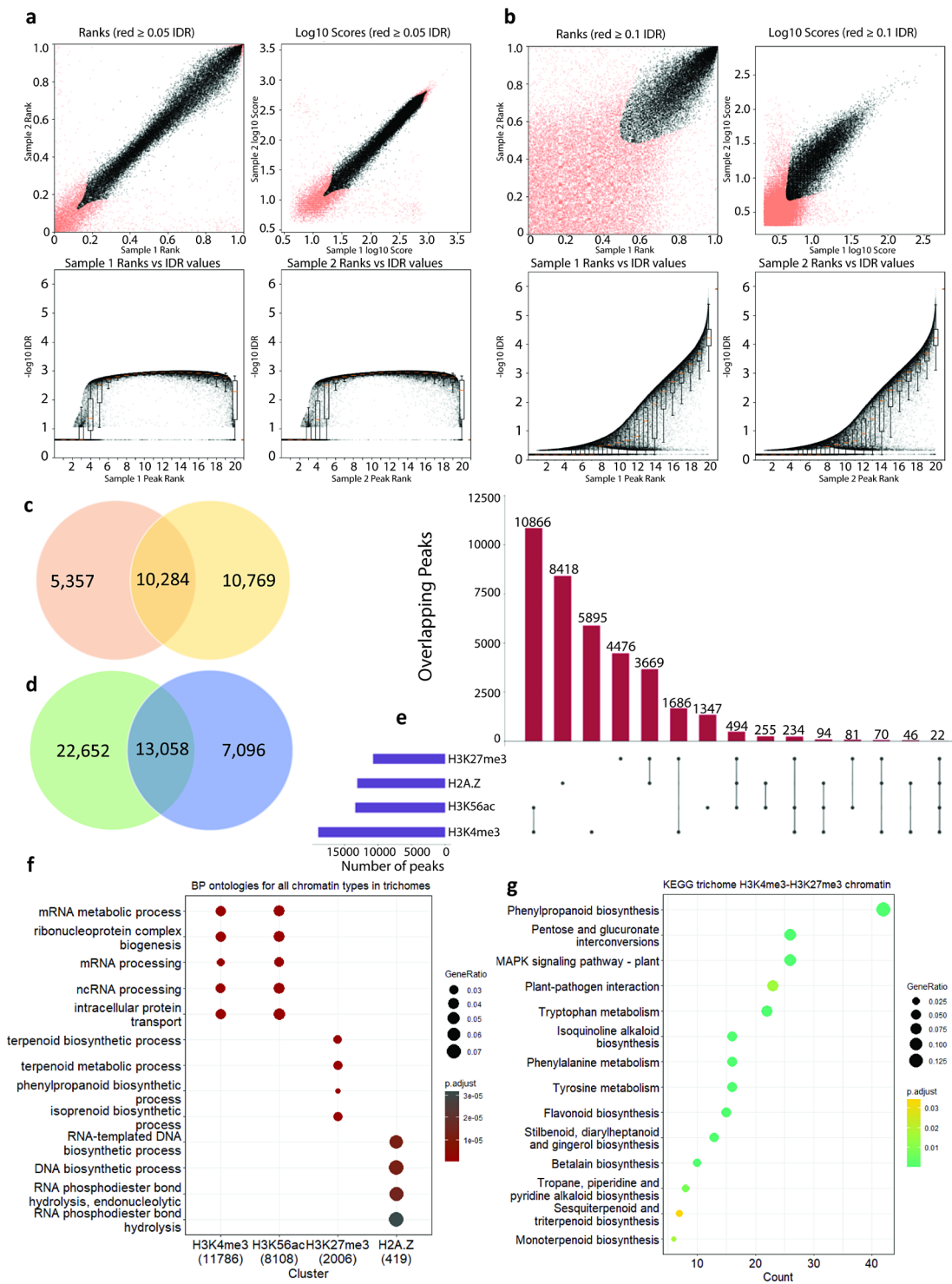
Histone marks can function co-operatively to either promote or repress the expression of genes. H3K56ac often co-localises with H3K4me3 at transcriptional start sites (TSS) and promoters to promote a relaxed chromatin environment conducive for gene expression, while H3K27me3 and mono ubiquitination of H2A.Z co-localise in the body of repressed genes to promote polycomb mediated gene silencing [12, 44]. We investigated the extent of these functional co-occurrences in *C. sativa*. To do so we identified high-confidence loci, genome-wide, that were enriched for each individual mark.

Different strategies were necessary for H3K4me3 and H3K56Kac, because these are found across smaller loci called “narrow” or “mixed” peaks in ChIP-Seq analyses, relative to H3K27me3 and H2A.Z, which are found across large loci termed “broad” peaks. Accordingly, we applied defined gold standard community practises [45]. The Irreproducible Discovery Rate (IDR) was applied to H3K4me3 and H3K56Kac data to ensure the statistical replicability of peak locations across our two biologically independent replicates (IDR cutoff  $\leq 0.05$  for H3K4me3 due to the true narrow peaks, and a less stringent  $\text{IDR} \leq 0.10$  for H3K56ac due to the mixed peak-type). Replicate peak ranks indicated a high degree of replicability (black points) for H3K4me3 data, and acceptable replicability in H3K56ac data (Fig. 2a-b). This resulted in 19,704 (H3K4me3) and 13,590 (H3K56Kac) high confidence peaks for glandular trichomes, 18,668 (H3K4me3) and 14,147 (H3K56ac) high confidence peaks for stem (internode), and 15,135 (H3K4me3) and 1,655 (H3K56ac) high confidence peaks for leaf tissue (Fig. 2a-b, Additional File 1: Fig. S7). Broad domain peak replicability was ensured by retaining only peaks that were found in both biological replicates, yielding 10,284 and 13,058 high confidence peaks for H3K27me3 and H2A.Z in *C. sativa* glandular trichomes respectively (Fig. 2c-d). The genomic locations of high-confidence peaks and their associated features were then annotated for all histone marks (Additional Files 2–4). Sequence reads and high confidence peaks for each histone mark were also loaded into a public web browser ([https://jbrowse.latrobe.edu.au/cannabis\\_trichome\\_epigenome/](https://jbrowse.latrobe.edu.au/cannabis_trichome_epigenome/)) to allow individual loci to be examined.

We then investigated potential co-localisation in the genome of histone marks, focusing on glandular trichomes (Fig. 2e). The co-localization of H3K4me3 and H3K56ac was statistically significant (Fisher exact test  $p < 1 \times 10^{-5}$ ; 10,866 overlapping peaks versus 5,895 H3K4me3 and 1,347 H3K56ac non-overlapping peaks) (Additional File 1: Fig. S7b-d). The co-localisation of H3K27me3 and H2A.Z was also statistically significant (Fisher exact test  $p < 1 \times 10^{-5}$ ; 3,669 overlapping peaks), as



**Fig. 1** Assessment of data quality and genome feature associations of histone marks in *Cannabis sativa* glandular trichomes. **(a)** Light microscopy image (1 mm scale bar) of *C. sativa* glandular trichomes collected using a 25  $\mu$ m mesh size during floral tissue ice-water extraction method. **(b)** Spearman correlation matrix of bam files for 3 biological replicates of glandular trichome RNA-seq data and two biological replicates each of glandular trichome ChIP-seq data for H3K4me3, H3K56ac, H3K27me3, and H2A.Z. **(c)** Distribution of features across the *C. sativa* genome; (i) gene density plot (ii) glandular trichome transcribed-gene density plot (iii) glandular trichome H3K4me3 peak density plot (iv) glandular trichome H3K56ac peak density plot (v) glandular trichome untranscribed-gene density plot (vi) glandular trichome H3K27me3 peak density plot and (vii) glandular trichome H2A.Z density plot. **(d)** Distribution of H2A.Z, H3K27me3, H3K56ac, and H3K4me3 across genome features. **(e)** Distribution of H3K4me3, H3K56ac, H3K27me3, and H2A.Z reads across all transcribed genes and untranscribed genes in glandular trichomes. Scale bars indicate relative enrichment



**Fig. 2** Identification of high confidence, replicable, histone mark loci, and their general functional associations in glandular trichomes. Replicable peaks (black points) for **(a)** H3K4me3 narrow peaks and **(b)** H3K56ac mixed peak type in glandular trichomes were determined by applying irreproducible discovery rate (IDR) analysis, comparing peak ranks by significance value between sample panels show good replicability plotted by both sample peak ranks (top panels) and corresponding  $-\log_{10}$  q-value of peak enrichment vs. sample rank (bottom panels) indicates the strength of enrichment ( $-\log_{10}$  q-value) across ranked list. H3K4me3 showed extremely high replicability with the majority of peaks displaying strong enrichment across lower ranks (bottom panel) yielding 19,704 replicable peaks, and H3K56ac and showed good replicability yielding 13,590 replicable peaks. Replicable broad peaks determined via the intersection of replicates for **(c)** H3K27me3 and **(d)** H2A.Z yield 10,284 and 13,058 replicable peaks respectively for glandular trichomes. **(e)** Upset plot of the replicable peak type data showing the relationships, by means of physical intersection of the various chromatin types throughout the glandular trichome epigenome. **(f)** Biological process ontologies enriched in glandular trichomes for each H3K4me3, H3K56ac, H3K27me3, and H2A.Z. **(g)** KEGG ontology analysis of supposed bivalent chromatin H3K4me3-H3K27me3 in glandular trichomes

was co-localisation of H3K4me3 and H3K27me3 (Fisher exact test  $p < 1 \times 10^{-5}$ ; 1,686 overlapping peaks). We further cross-referenced replicable peaks for each histone mark against our trichome RNA-seq dataset – we partitioned genes according to their transcriptional status (transcribed genes  $\geq 1$  TPM) and then recorded genes that contained each of the four histone marks (Additional File 1: Table S4).

#### Cross-validation of the expression-associated histone mark H3K4me3 across independent public transcriptome datasets

We made use of public data to cross-validate our ChIP-seq maps. Several laboratories have performed transcriptomic analyses of *C. sativa* trichomes [46–48]. Some variation between datasets would be expected for both technical and biological reasons, such as different *C. sativa* strains, varying trichome purification methods, differences in sequence depth and sequencing design such as single and paired end sequencing methods. Nonetheless, we can reasonably expect that the population of expressed genes should substantially correlate between trichome datasets, because the fundamental biological processes operating in glandular trichomes of any *C. sativa* strain will not differ. It has been confidently established in many plant species that H3K4me3 is directly associated with the TSSs of expressed genes [5, 9, 49, 50]. This consequently provided an opportunity to cross-validate the accuracy of our ChIP-seq maps across the independent transcriptomic datasets. We extracted the population of genes underlying glandular trichome specific H3K4me3 loci in our ChIP-seq dataset, then quantified their expression in each of 33 transcriptome datasets (Additional File 1: Fig. S7e). We observed positive correlation of the expression of this population of genes across all trichome specific transcriptome datasets, contrasting with lower correlation to *C. sativa* stem and leaf specific transcriptomes. This indicates that our ChIP-seq method robustly identifies genes expressed trichome-specifically, consistent with the known biology of H3K4me3.

#### Histone marks are associated with different functional categories of genes in glandular trichomes

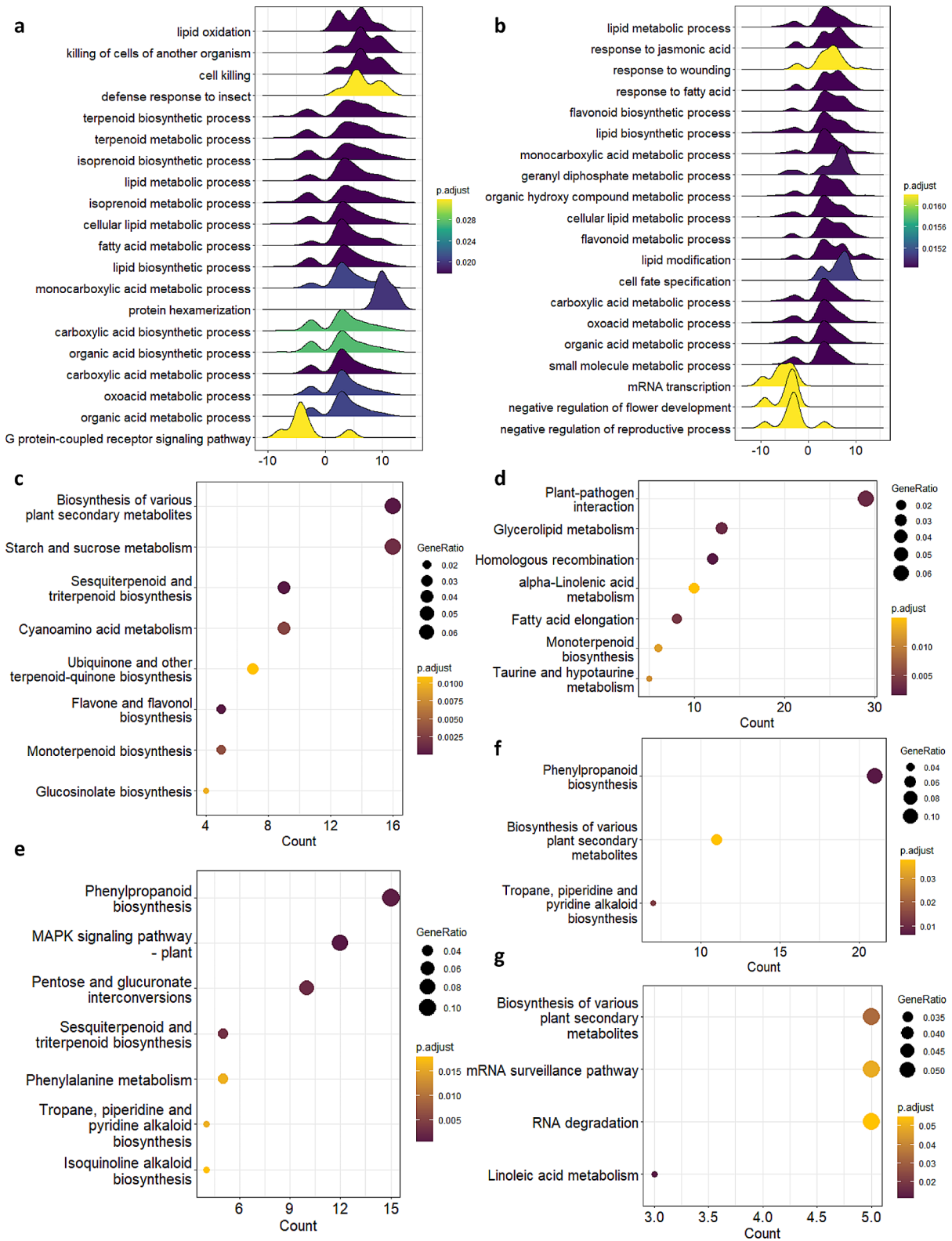
Next, we asked what the function of genes associated with different histone marks in glandular trichomes were. To gain a global understanding, we assessed all loci bearing histone marks in glandular trichomes; we did not focus on loci only present specifically in glandular trichomes. Gene Ontology (GO) enrichment analysis of biological processes (BP) was performed on genes found within regions bearing each of the four histone marks in glandular trichomes (Fig. 2f). The genes associated with H3K4me3 (11,786 genes) and H3K56ac (8,108 genes)

were primarily enriched for the same housekeeping functions, including mRNA metabolic process, and mRNA processing. This likely reflects the conserved role of H3K4me3 and H3K56ac in promoting gene expression, and the fact that our analyses considered all expressed genes rather than a trichome-specific subset. H3K27me3 and H2A.Z loci contained fewer genes and 419, respectively). Genes associated H3K27me3 (2,006 genes) were enriched for functions related to plant specialised metabolism, such as terpenoid metabolic process and isoprenoid biosynthetic process. This was notable, as H3K27me3 is associated with repression of gene expression. Genes associated with H2A.Z (419 genes) were enriched for functions associated with DNA homeostasis and transcriptional control, in line with previous observations made in plants and animals (Additional File 5) [51, 52].

Our observations that specialised metabolism genes were enriched within H3K27me3 regions, and that H3K27me3 was frequently co-located with H3K4me3 in glandular trichomes, prompted us to investigate whether bi-valent chromatin states may be a common feature of regulation of specialised metabolism in glandular trichomes (Fig. 2e). Bi-valent H3K27me3/H3K18ac loci are involved in regulation of biosynthesis of the specialised metabolite camalexin in *A. thaliana* [53]. We examined this by using the Kyoto Encyclopedia for Genes and Genomes (KEGG) curated pathways for *C. sativa* cs10 (GCF\_900626175.2) to analyse pathways enriched amongst genes located in regions where H3K27me3/H3K4me3 overlapped. Pathways enriched amongst these genes were associated with plant specialised metabolism and defence, including phenylpropanoid biosynthesis, MAPK signalling, plant-pathogen interaction, flavonoid biosynthesis, and sesquiterpenoid biosynthesis (Fig. 2g, Additional File 6).

#### Differential epigenomic analysis shows glandular trichome specific enrichment of histone marks associated with specialised metabolism

We reasoned that by focusing our analysis on glandular trichome specific gene expression and histone marks, we could gain deeper insight into gene regulation in glandular trichomes. To achieve this we first identified differentially expressed genes in glandular trichomes versus leaf or stem tissues. There were 1,765 (trichome vs. leaf) and 1,987 (trichome vs. stem) differentially expressed genes ( $\log_2 \text{FC} > 2$ , BH  $p_{\text{adj}} < 0.05$ ) (Additional File 7). Next, we assessed the functions of genes differentially expressed between these tissues using gene set enrichment analysis (GSE). Genes differentially expressed in glandular trichomes, compared with both stem and leaf, were highly enriched for biological processes including lipid metabolism, specialised metabolism, and organic acid biosynthesis in both cases (Fig. 3a-b, Additional File 8). These



**Fig. 3** Analysis of glandular trichome specific gene expression and histone marks yields functional insights. **(a)** Pairwise gene set enrichment analysis (GSE) leveraging Leaf and **(b)** Stem RNA datasets. KEGG ontology analysis of glandular trichome specific chromatin types **(c)** H3K4me3 **(d)** H3K56ac **(e)** H3K4me3-H3K27me3 **(f)** H3K27me3 and **(g)** H2A.Z

genes were also enriched for molecular functions including delta-9 tetrahydrocannabinolate synthase activity, cannabidiolate synthase activity, germacrene-a synthase activity, and other specialised metabolite related functions (Additional File 1: Fig. S8a-b, Additional File 8).

We next examined the functions of genes that were associated with glandular trichome specific regions of the different histone marks. To accomplish this, we identified the glandular trichome specific enriched loci for each histone mark from our high confidence peak lists (i.e. loci not present in any other tissue), then identified genes underlying these peaks. Functions enriched amongst these genes were assessed by KEGG ontology analysis (Additional File 9). Genes within glandular trichome specific H3K4me3 or H3K56ac regions were enriched for starch and sucrose metabolism, sesquiterpenoid and triterpenoid biosynthesis, plant-pathogen interactions, and monoterpenoid biosynthesis (Fig. 3c-d). Genes within glandular trichome specific H3K4me3-H3K27me3 bi-valent regions were enriched for MAPK signalling, phenylpropanoid biosynthesis, and sesquiterpenoid and triterpenoid biosynthesis (Fig. 3e). Genes within glandular trichome specific H3K27me3 chromatin or H2A.Z chromatin were enriched in KEGG ontologies related to biosynthesis of various plant specialised metabolites, phenylpropanoid biosynthesis, and mRNA surveillance (Fig. 3f-g). This analysis implies that specialised metabolism may be differentially regulated by specific histone marks in glandular trichomes (Additional File 9).

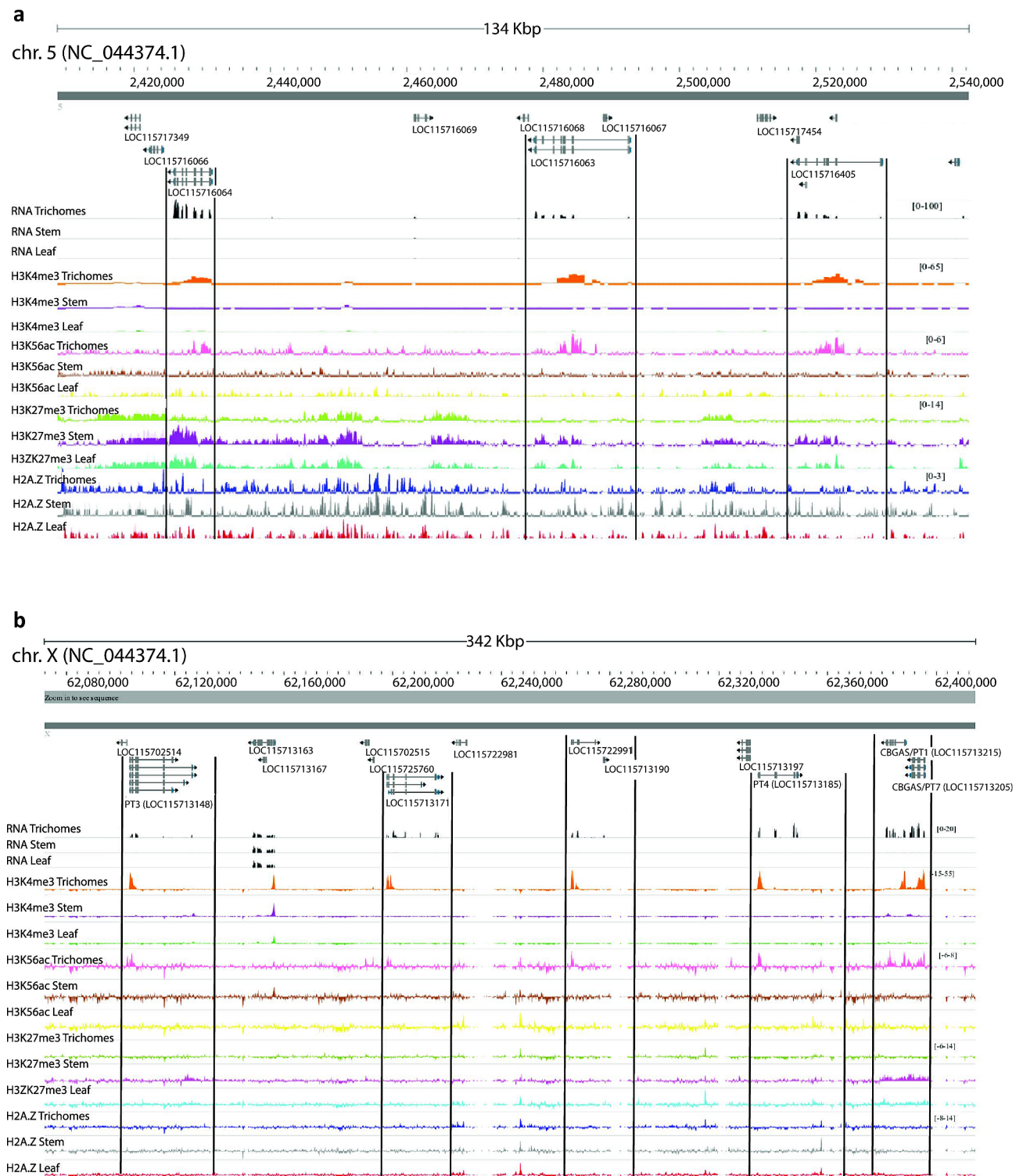
We then examined whether there were tissue specific differences in the histone marks associated with well characterised genes expressed specifically in glandular trichomes. For example, phytocannabinoid biosynthesis 2-acylphloroglucinol 4-prenyltransferase (also called cannabigerolic acid synthase, CBGAS/PT4; LOC115713185), ABC-transporter B family member 2 (LOC115716265), and alpha-humulene synthase (LOC115724563) were all expressed specifically in glandular trichomes and associated with glandular trichome specific enrichment of the transcription-associated histone marks H3K4me3 and H3K56ac in their promoter regions (Additional File 1: Fig. S9a) [9, 54]. The silencing marks H3K27me3 and H2A.Z were found in the gene body of alpha humulene synthase (LOC115724563) in stem and leaf tissues, consistent with the lack of expression of this gene in those tissues. H3K27me3 enrichment was less at alpha humulene synthase (LOC115724563) in glandular trichomes, where the gene is expressed, but it was not completely absent. Phytocannabinoid biosynthesis 2-acylphloroglucinol 4-prenyltransferase (LOC115713185) and ABC-transporter B family member 2 (LOC115716265) were not associated with H3K27me3 or H2A.Z in stem or leaf, despite their lack of expression in those tissues. Taken together, these observations indicate that genes

expressed trichome specifically are associated with trichome specific local chromatin landscapes but may also indicate further complexity or multivalency to chromatin states and the regulation of expression at individual genes in glandular trichomes (Additional File 1: Fig. S9a).

#### **Glandular trichome specific epigenomic regulation of putative gene clusters in *Cannabis sativa***

Genes encoding components of specialized metabolite pathways are frequently found in plant gene clusters (GCs) [55]. These can be the consequence of local duplications such as gene family expansion in terpene synthases or, alternatively, unrelated genes that constitute or modulate a metabolic pathway may be co-located thereby forming a biosynthetic gene cluster (BGC) [56–61]. Genes in either cluster type are frequently expressed tissue or cell type specifically [62, 63]. The epigenomic properties of such clusters are poorly defined in plants and understanding them better may offer a unique and refined strategy to introduce entire biosynthetic gene pathways, discretely into a genome, in a tissue or cell type specific manner for applications in plant metabolic engineering.

We consequently examined the expression status of genes in putative *C. sativa* gene clusters and their associated histone marks, focusing upon two previously predicted clusters that may have glandular trichome specific activity [64]. The first putative BGC (#17), involved in monoterpene biosynthesis, exhibited glandular trichome specific gene expression, and association with transcriptionally active histone marks (H3K4me3 and H3K56ac), including limonene synthase/TPS1 (LOC115716064), and two myrcene synthases (LOC115716063, LOC115716405) (Fig. 4a). Contrastingly, in leaf and stem tissue these same genes were associated with the repressive mark H3K27me3. The second putative BGC (#28) had trichome specific gene expression of class V chitinases *CHIT5* and *CHIT5*-like (LOC115724705 and LOC115695573), germacrene-A synthase (LOC115695573), and ferredoxin-NADP reductase embryo like (LOC115724444) but did not exhibit clear relationships between histone marks H3K4me3 and H3K56ac and transcription. The chromatin state of germacrene-A synthase (LOC115695573) was indistinguishable between either tissue at the epigenomic level, except for more pronounced H2A.Z deposition near the promoter region. Ferredoxin-NADP reductase embryo-like (LOC115724444) was expressed solely in the glandular trichomes, yet there was no discernible difference in either H3K4me3 or H3K56ac deposition between glandular trichomes and other tissues. There was however enrichment of H2A.Z in the gene body of LOC115724444 in glandular trichomes. Furthermore, throughout the length of BGC #28 there was more pronounced



**Fig. 4** Specialised metabolism gene clusters have glandular trichome histone marks and gene expression. **(a)** Monoterpene synthase gene cluster (cluster no. 17) shows glandular trichome specific enrichment of H3K4me3 and H3K56ac chromatin and corresponding gene expression in LOC115716064 limonene synthase, LOC115716063 myrcene synthase, and LOC115716405 myrcene synthase and inverse depletion and expression in leaf and stem datasets with corresponding H3K27me3 enrichment in stem and leaf datasets. **(b)** A 342 Kbp region of X chromosome containing putative novel cannabinigerolic acid synthase/prenyltransferase gene cluster. This had histone marks H3K4me3 and H3K56ac specifically in glandular trichomes, conducive to gene expression, and corresponding trichome specific expression of PT3 (LOC115713148), LOC115713171, LOC115722991, PT4 (LOC115713185), CBGAS/PT1 (LOC115713215), and CBGAS/PT7 (LOC115713205)

enrichment of H2A.Z in trichomes comparatively, a feature which has been associated with tissue specific BGC regulation in *A. thaliana* (Additional File 1: Fig. S9b) [65].

We also observed a novel putative cluster of six aromatic prenyltransferases with trichome specific histone marks and gene expression (Fig. 4b, Additional File 7). This included the cannabigerolic acid synthases PT1, PT4, and PT7 (LOC115713215, LOC114713185, and LOC115713205) cannflavin biosynthesis gene PT3 (LOC115713148) as well as two additional uncharacterised CBGAS-like genes (LOC115713171 and LOC115722991) in a 347 Kbp region of the X chromosome [66, 67]. These genes had glandular trichome specific enrichment of H3K4me3 and H3K56ac in their promoters compared to other intervening genes, consistent with their active transcription in that tissue.

#### Identification of putative glandular trichome specific regulatory motifs using H3K56ac enrichment

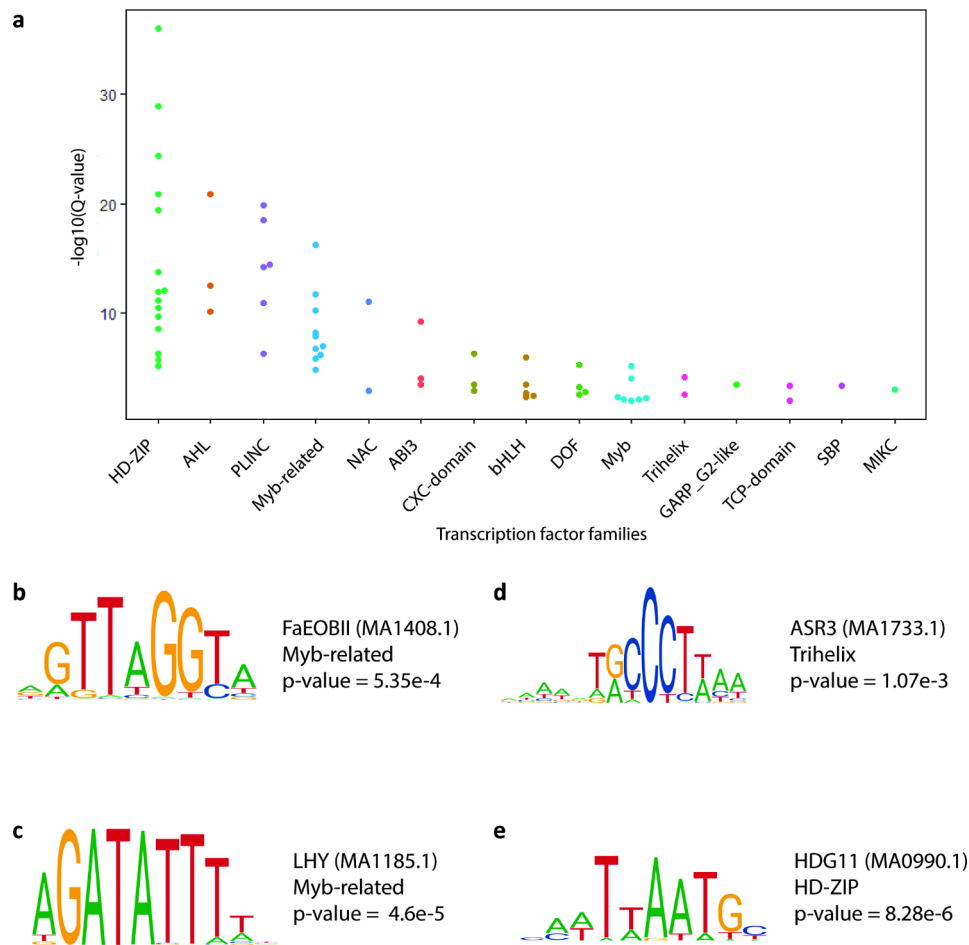
Cis-regulatory elements (CREs) are features within genomes which contain DNA sequence motifs that TFs can bind to. TFs drive gene expression through interactions with target gene promoters, either because the CRE is within a promoter itself, or through long-range interactions between a promoter and an enhancer that contains the CRE. Active plant CREs are often flanked by H3K56 acetylated chromatin [29]. We reasoned that we could mine putative CREs with glandular trichome specific activity by identifying enriched DNA sequence motifs within H3K56ac loci specific to glandular trichomes. The 1,048 H3K56ac loci specific to glandular trichomes were significantly enriched for 65 DNA sequence motifs ( $q < 0.05$ , JASPAR CORE database). The enriched motifs included MYB, MYB-related, NAC, and TCP transcription factor binding motifs (Fig. 5a, Additional File 10).

We examined the cognate transcription factors that may recognise these motifs and their likely functions, to better understand how they may be related to glandular trichome biology. For example, the MYB-related *Fragaria ananassa* emission of benzoid II (FaEOBII) cis element was enriched in putative CREs (Additional File 10). The corresponding best blast P hit for its associated transcription factor in the cs10 reference assembly was the MYB-related transcription factor CsEOBI (LOC115699015). CsEOBI is highly expressed in the glandular trichomes compared to either leaf ( $\log_2FC = 7.49$ , BH  $p_{adj} = 0.000105$ ) or stem ( $\log_2FC = 8.08$ , BH  $p_{adj} = 1.21E-05$ ). FaEOBII is known to positively regulate production of volatile organic compounds (Fig. 5b) [68]. Interestingly, LHY and CCA1 motifs were also enriched in putative CREs. The cognate transcription factors for these motifs work cooperatively to regulate circadian rhythms (Fig. 5c) [69]. The trihelix family member ASR3 motif was enriched, and its associated transcription

factor negatively regulates pattern triggered immunity (PTI) that mitigates unnecessary hypersensitivity responses (Fig. 5d). HDG11 motifs were also enriched in glandular trichome specific H3K56ac chromatin. HDG11 has been shown to negatively regulate trichome branching in other plant species (Additional File 10).

#### Predicting individual enhancer-gene interactions specific to glandular trichomes

Enhancer elements regulate gene expression by long range interactions with gene promoters, over kilobases to megabases. Enhancers contain CREs within which TFs bind, driving the long-range interactions, and so often carry H3K56ac histone marks. We therefore identified putative enhancer-gene target pairs active in glandular trichomes by making use of our CRE analysis results. We focused only on H3K56ac loci greater than 1.5 Kbp from the nearest TSS, to exclude gene-proximal promoters. These H3K56ac loci were then matched with the nearest gene that was expressed specifically in glandular trichomes, identifying 87 putative enhancer-target gene pairs (Table 1, Additional File 11). Strikingly, this strategy captured H3K56ac loci associated with the expression of hallmark glandular trichome associated genes. These included genes involved in terpenoid specialised metabolism, such as sesquiterpene biosynthesis; alpha-humulene synthase (LOC115724563), monoterpene biosynthesis; myrcene synthase (LOC115716405) and limonene synthase (LOC115716064), and triterpene biosynthesis; beta-amyrin synthase (LOC115719748). They also included genes involved in cannabinoid specialised metabolism, such as biosynthesis of the cannabinoid precursor hexanoate Butanoate-CoA ligase AAE1 (LOC115713865), as well as cannabidiolic acid synthase (LOC115697762). Fatty acid biosynthetic genes implicated in the upstream production of fatty alkyl side groups, required for cannabinoid production, were also found (Lineolate 13 S-lipoxygenase, LOX1, LOC115719612; Delta (12)-oleate desaturase, FAD1, LOC115719329), as were specialised metabolite transporters (ABC transporter B family member 2, LOC115716265; ABC transporter G family member 20, LOC115711415), and a transcription factor involved in trichome maturation (MYB106, LOC115701410). Interestingly, several of the enhancer-gene target predictions identified genes involved in biotic stress resistance that have not yet been reported in studies of glandular trichomes, including disease resistance protein RMP1 (LOC115704534) that is currently being investigated in wheat for its properties in resistance to powdery mildew, and an unusual protease neprosin (LOC115720846) found in carnivorous plants.



**Fig. 5** Putative glandular trichome specific cis regulatory elements are enriched in transcription factor binding motifs involved in specialised metabolism, circadian rhythm, and stress resistance. **(a)** Plot showing various transcription factor families associated with mined motifs in glandular trichome specific cis-elements. Each point represents the enrichment significance score of a transcription factor motif found in our dataset, plotted above its cognate family. **(b)** FaEOBII (MA1408.1) motif enriched in putative trichome specific cis-elements is associated with aroma volatile production. **(c)** LHY (MA1185.1) and CCA1 (MA0972.1) motifs involved in circadian rhythm homeostasis is enriched in trichome specific cis elements. **(d)** The pattern triggered immunity related ASR3 (MA1733.1) motif and **(e)** negative regulator of trichome branching associate regulatory motif HDG11(MA0990.1) are enriched in trichome specific cis regulatory elements

**Discussion**

*C. sativa* is a plant with significant biotechnological potential owing to its specialised metabolite productivity from large capitate stalked glandular trichomes. Despite our reliance on plant specialised metabolite products in industry, food, and medicine, our understanding of glandular trichomes is sparse. Currently, only a single publication in plants investigates epigenomic regulation of plant glandular trichome activity [94]. Here, we have provided the first H3K4me3, H3K56ac, H3K27me3, and H2A.Z chromatin maps in *C. sativa* glandular trichomes, and in plant glandular trichomes more broadly. Additionally, we demonstrate that classical eukaryotic functions of histone post translational modifications are functionally conserved in the *C. sativa* epigenome (Fig. 1).

Our chromatin landscape maps have been assembled using gold standard ENCODE guidelines to provide

confidence in the observations made in this study. We provide these as a web browser resource for easy data re-use and for readers to inspect individual loci of interest ([https://jbrowse.latrobe.edu.au/cannabis\\_trichome\\_epi\\_genome/](https://jbrowse.latrobe.edu.au/cannabis_trichome_epi_genome/)). A key theme we observe in this study is that chromatin modifications often co-localise, presumably to co-operatively regulate gene expression. For example, H3K56ac and H3K4me3 often co-localize at the promoters and TSSs of actively transcribed genes, whilst H3K27me3 and H2A.Z co-localize in the gene bodies of silenced genes (Fig. 2).

Our analysis highlights tissue-specific gene expression and chromatin states consistent with focal points of specialised metabolism, carbon trafficking, and plant abiotic and biotic stress resistance, as would be expected given the functions of glandular trichomes. For example, we observed that glandular trichome specific putative

**Table 1** Closest neighbouring genes that are expressed glandular trichome specifically to *C. sativa* glandular trichome specific gene distal H3K56ac loci, showing expression – a list of putative tissue specific enhancer gene interactions. \* indicates intragenic putative cis-regulatory elements

Category	Gene ID	Description	Relative expression trichome/stem		Relative expression trichome/leaf		Putative enhancer -H3K56ac locus	Refs.
			log2FC	p.adj	log2FC	p.adj		
Cannabinoid biosynthesis	LOC115697762	Cannabidiolic acid synthase (CBDAS)	7.32	2.24E-09	9.15	7.3E-09	Chr7 30,985,961 – 30,986,744	[70, 71]
	LOC115697886	Cannabichromenic acid synthase-like	8.174	7.38E-06	7.58	6.8E-05	Chr7 26,086,586 – 26,087,174	
Sesquiterpenoid biosynthesis	LOC115724563	Alpha-humulene synthase	4.841	1.63E-19	7.145	3.90E-31	Chr6 78,980,722 – 78,981,538	[72]
Triterpene biosynthesis	LOC115719748	Beta-amyrin synthase	9.13	5.42E-09	8.54	1.23E-07	Chr2 81,831,322 – 81,832,508	[73]
Monoterpenoid biosynthesis	LOC115716064	(-)-Limonene synthase	7.595	4.12E-28	11.28	1.98E-48	Chr5 2,427,955 – 2,429,181*	[74]
	LOC115716405	Myrcene synthase	6.792	2.17E-13	8.807	2.17E-13	Chr5 2,520,951 – 2,522,151*	
Aroma volatile biosynthesis	LOC115717307	Methionine gamma-lyase	4.84	9.23E-43	5.464	1.7E-56	Chr5 86,422,670 – 86,424,161	[75]
	LOC115718871	Pyruvate decarboxylase 1-like	7.886	3.53E-57	12.88	9.72E-20	Chr2 5,491,001 – 5,491,980	[76]
	LOC115716705	branched-chain amino acid aminotransferase 1	5.107	7.07E-13	7.384	1.62E-07	Chr5 7,467,915 – 7,469,631*	[77]
Hexanoate biosynthesis	LOC115713865	Butanoate—CoA ligase (AAE1)	2.191	0.000929	2.955	4.3E-09	ChrX 70,642,217 – 70,643,833	[78–81]
	LOC115695656	2-alkenal reductase (NADP (+) dependent)	6.654	1.34E-23	10.656	1.55E-12	Chr6 17,801,253 – 17,802,062*	
Oxylipin pathway	LOC115719329	Delta (12)-oleate desaturase (FAD1)	6.335	1.63E-39	9.606	3.56E-64	Chr2 87,098,771 – 87,099,548	
	LOC115719612	Lineolate 13 S-lipoxygenase (LOX1)	5.03	1.45E-33	2.148	0.0048	Chr2 93,912,176 – 93,919,286	
Fatty acid biosynthesis	LOC115705717	3-ketoacyl-CoA synthase 19	3.946	6.33E-08	8.833	1.87E-10	Chr1 22,744,264 – 22,745,167*	[82, 83]
Desaturase	LOC115724633	Fatty acid desaturase 4-like 1	6.688	1.86E-46	9.642	1.62E-25	Chr6 70,113,632 – 70,114,540	[84]
Protease	LOC115720846	Protein neprosin	4.43	1.26E-18	4.52	1.75E-17	Chr2 12,637,583 – 12,638,999	[85]
Powdery mildew resistance	LOC115704534	Disease resistance protein (RMP1)	9.23	3.47E-10	9.61	6.85E-11	Chr1 89,873,411 – 89,875,839*	[86, 87]
Innate immunity	LOC115722088	Epsin-3	8.157	2E-06	4.48	9.7E-05	Chr9 142,307 – 143,004	[88]
Metabolite transport	LOC115716265	ABC transporter B family member 2	4.133	1.13E-20	3.258	9.61E-11	Chr5 83,485,739 – 83,487,630*	[89]
	LOC115711415	ABC transporter G family member 20	5.18	8.92E-17	4.39	1.92E-11	Chr3 66,573,005 – 66,573,376	[90, 91]
Trichome maturation	LOC115701410	MYB106	4.93	2.97E-59	3.64	6.63E-34	Chr8 62,597,900 – 62,598,481	[92]
Phloem protein	LOC115723346	Sieve element occlusion B	6.46	4.93E-43	8.21	2.3E-52	Chr9 32,017,566 – 32,018,467*	[93]

H3K4me3-H3K27me3 bi-valent chromatin is enriched in specialised metabolism, defence signalling, and innate immunity associated genes. We speculate that complex chromatin landscapes like these have the potential to integrate environmental cues to dynamically allocate carbon resources as needed during defence response elicitation like that reported in camalexin biosynthesis (Fig. 3) [53]. It should be noted, however, that overlaps between H3K4me3-H3K27me3 loci in leaf or stem tissue might reflect the heterogenous composition of those tissues, as

opposed to a bona-fide signal of bi-valency. This limitation is less probable in the glandular trichomes samples, which were isolated to a high degree of purity largely composed of disc cells.

Previous work in *A. thaliana* suggests unique chromatin landscapes regulate plant BGC expression, so examining these landscapes provides an opportunity to better define *C. sativa* BGCs and understand how their activity may be regulated [95, 96]. We investigated our previously predicted putative *C. sativa* BGCs for evidence

of trichome specific gene expression and chromatin composition [64]. We observed that two putative BGCs (#17, terpene-related; #28, terpene and fungal resistance related) were expressed specifically in glandular trichomes. Most genes in both BGCs displayed enrichment of promoter H3K4me3 and H3K56ac specifically in glandular trichomes, consistent with active gene expression. Observations made in *A. thaliana* suggest enrichment of H2A.Z is functionally associated with active BGCs, and enrichment of H3K27me3 at those BGCs in tissues where they are not expressed [97]. Contrasting with this, we did not observe enrichment of H2A.Z in a tissue specific manner that might explain trichome specific expression of either BGC. However, we did observe enrichment of H3K27me3 in BGC #17 monoterpene genes in stem and leaf tissue, where these genes were not expressed. Nonetheless, our data prompts further research into the association of distinct H2A.Z and H3K27me3 chromatin states at active and repressed BGCs. Future investigation might also consider the capacity of H2A.Z, a histone variant, to be post-translationally modified itself [44].

We made use of the association between H3K56ac and CREs to identify putative *C. sativa* glandular trichome specific TFs and binding motifs [29, 30, 98]. This strategy yielded candidates involved in aroma volatile production associated (FaEOBII, JASPAR ID MA1408.1), trichome development (HDG11; MA0990.1), and circadian rhythms (LHY; MA1185.1, CCA1; MA0972.1). Daytime progression has recently been implicated in cannabinoid production which might explain the enrichment of circadian rhythm associated motifs in our dataset (Fig. 5) [99]. Furthermore, we hypothesised that we could leverage our H3K56ac and gene expression datasets to mine the glandular trichome epigenome for putative enhancer-gene target pairs. This yielded several putative enhancer-gene pairs where the genes have known glandular trichome specific functions, including CBDAS (LOC115697762), terpenoid synthases (LOC115724563, LOC115719748, LOC115716064, LOC115716405), and metabolite transporters (LOC115716265, LOC115711415) [70, 91, 100, 101]. We note that our approach would favour identification of interactions at kilobase scale and would not allow us to identify enhancer elements operating at very long range, over intervening genes. Further study of these would require assays like Hi-C to capture the full scope of enhancer-target gene interactions. Nonetheless, the putative enhancer-gene interactions we identify here in *C. sativa* presents a significant biotechnological resource that may provide a framework for engineering glandular trichome gene expression in the future.

## Conclusion

In this study we provide novel insights into the chromatin landscape of glandular trichomes. Our work shows how chromatin features and function are conserved in *C. sativa* compared with more deeply characterised plant species, from feature types to correlation with gene expression, deposition and localisation within genes, and association with distal regulatory regions. Moreover, we show the advantages of studying such chromatin features where, for example, genes postulated to exist in previous studies have been predicted by our multimodal approach, and genome wide sequence information can be mined in tissue specific manners to identify cis-regulatory elements not yet observed in *C. sativa*. This work has laid a solid foundation for our understanding of glandular trichome chromatin dynamics with respect to tissue specific gene expression and cis-regulatory elements, and may to glandular trichome metabolic engineering efforts in years to come.

## Methods

### Plant material and growth conditions

Cuttings were taken from female genotype (MW6-15) [81]. MW6-15 was derived from an industrial hemp line (accession #6) previously donated to the university by Southern Cross University [102]. Cuttings were made using a sterile blade and at a 45-degree angle immediately below the 3rd node from the apical meristem. The cuttings were partially pruned – removing leaf tips and large leaves, taking care not to remove all leaves. The cuttings were then dipped in 3.0 g/L indole butyric acid rooting gel hormone (Growth Technology © Clonex purple) and allowed to sit for 30 s. Cuttings were individually transferred to Grodan Rockwool cubes (36 mm x 36 mm x 40 mm) previously saturated with half strength CANNA veg fertiliser (20 mL A + 20 mL B/ 10 L water), and then transferred to propagator humidity domes (4 clones to one propagator) with the vent fully closed. Clones were monitored daily – half strength CANNA veg fertiliser (20 mL A + 20 mL B/ 10 L water) was added to the base of the propagator when required and vents were opened in small daily increments until completely opened and grown under vegetative lighting conditions of 18 h on 6 h off using Philips Master TL-D Super 80 low-pressure mercury discharge lamps. Clones were kept in propagators until established roots could be observed (approximately 2 weeks) at the base of the rockwool cubes. Potting mix containing 1:1:1 peat moss, perlite, and vermiculite supplemented with 1 g/L dolomite was prepared and divided into 1 L pots. Pots were then saturated with full strength CANNA veg fertiliser (40 mL A + 40 mL B/ 10 L water). Clones were then transferred to appropriately labelled 1 L pots and allowed to grow under vegetative lighting conditions for 4 weeks. Plants were

monitored daily and watered regularly using full strength CANNA veg during the vegetative growth stage. Following 4 weeks of vegetative growth plants were re-potted into appropriately labelled 10 L pots using the above potting mix recipe. Potting mix was saturated with full strength CANNA flora (40 mL A + 40 mL B/ 10 L water) prior to re-potting of the clones. The lighting cycle was then varied to 12 h on and twelve hours off to induce flowering. Clones were grown under flowering conditions for 6 weeks and watered daily with full strength CANNA flora. Clones were harvested during the late stage of flowering as defined by a colour change, from white to brown, of greater than two thirds of all stigmas on the plant.

#### **Tissue harvest and trichome isolation for RNA sequencing**

For RNA extraction, fresh samples were collected in three biological replicates from mature fan leaves, 3 cm internodal stem segments, and female inflorescences. Triplicates were obtained from three individual plants. Trichomes were isolated from female inflorescences using a method modified from a protocol previously described [103]. Briefly, ~5 g samples were transferred to 50 mL Falcon™ tubes and about 10 mL of liquid nitrogen was added to each tube. The tubers were loosely capped and vortexed until the trichomes were fully removed. After removing plant debris by inverting and tapping the tubes, trichome-enriched samples were carefully transferred to 2 mL Eppendorf tubers in liquid nitrogen for further processing.

#### **Tissue harvest and trichome isolation for chromatin immunoprecipitation**

Samples of stem (internode) and vegetative leaves were taken from each clone using a sterile blade. Each sample was transferred to an appropriately labelled 50 mL falcon tube, sealed, and immediately flash frozen in a Dewar containing liquid nitrogen. Samples were then stored at -80 °C.

Glandular trichomes were isolated by ice-water extraction – all inflorescences were harvested from an individual clone and cut into small 5 cm x 5 cm pieces using sterile secateurs. Nylon bags with 25 µm, 45 µm, 73 µm, 120 µm, and 160 µm mesh sizes were sequentially placed inside one another starting the smallest, outermost, 25 µm mesh bag. The setup was then placed in a 5 L beaker and then filled with 1:3 ice-cold milli-q water: crushed ice. The cut inflorescences were then added to the beaker/nylon-mesh set-up and stirred using a large metal spatula for ten minutes. The nylon-mesh bags were then removed from the beaker and the contents of the beaker were then filtered through a nylon mesh with 25 µm pore size. Glandular trichomes were then retrieved, *via* pipetting, from the surface of the 25 µm nylon mesh and transferred into appropriately labelled falcon tubes. Glandular

trichomes were immediately frozen in liquid nitrogen and then transferred to cold storage at -80 °C.

#### **Tissue preparation and chromatin cross-linking**

Glandular trichomes, stem (internode), and vegetative leaves were retrieved from cold storage (-80 °C) and approximately 1 g of each tissue was ground into a fine powder using liquid nitrogen and a pre-chilled mortar and pestle. The ground tissues were then transferred into appropriately labelled 15 mL falcon tubes followed by the addition of 12.5 mL of nuclear isolation/ cross-linking buffer (60 mM HEPES at pH 8.0, 1 M Sucrose, 5 mM KCl, 5 mM EDTA at pH 8.0, 0.6% (v/v) Triton X-100, 1 mM PMSF (Sigma-Aldrich 93482-50ML-F), 1 mM pepstatin A (Sigma-Aldrich P5318-5MG), 1 mini-complete tablet (Sigma-Aldrich 11836170001) per 10 mL buffer) on ice (Additional File 12: Table A1). The contents of each tube were stirred until a homogenous suspension had formed. Next, 360 µL of 37% formaldehyde was added to each of the sample suspensions - cross-linking was achieved through incubation for 25 min at room temperature and gentle rotation. Cross-linking was halted through the addition of 875 µL 2 M glycine to each of the sample suspensions followed by 25 min incubation at room temperature and gentle rotation.

#### **Nuclei isolation**

Nuclei were isolated from the sample suspensions through passive filtration into a 50 mL falcon tube using a 40 µm nylon mesh sieve. The filtrate was then centrifuged at 4000 RPM for 20 min at 4 °C. The supernatants were carefully removed, and the soft nuclei pellets were resuspended, by pipetting, using 1 mL of extraction buffer (0.25 M Sucrose, 10 mM Tris-HCl at pH 8.0, 10 mM MgCl<sub>2</sub>, 1% (v/v) Triton X-100, 1 mM EDTA at pH 8.0, 5 mM β-mercaptoethanol, 1 mM PMSE, 1 mM Pepstatin A, 1 mini-complete table per 10mL) (Additional File 12: Table A2) and then transferred to a 2 mL Eppendorf tube. The walls of each 50 mL falcon tube were washed twice with 100 µL of extraction buffer and transferred to its corresponding Eppendorf tube to collect any residual nuclei. Tubes were then sealed and centrifuged at 11.4k RPM for 10 min at 4 °C and the supernatant discarded.

#### **Chromatin shearing**

Each of the pellets were then resuspended in 300 µL of nuclei lysis buffer (50 mM Tris-HCl at pH 8.0, 10 mM EDTA at pH 8.0, 1% (w/v) SDS, 1 mM PMSE, 1 mM Pepstatin A, 1 mini-complete table per 10 mL buffer) (Additional File 12: Table A3). Samples were then sonicated using a Diagenode Bioruptor® on high setting, 30 s on, 30 s off, for 15 min at 4 °C. Samples were then centrifuged at 5000 RPM for 10 min at 4 °C to pellet unwanted cellular/nuclear debris. The supernatant, containing sheared

chromatin, was then transferred to new Eppendorf tubes on ice. All samples were then placed in cold storage at  $-80^{\circ}\text{C}$ .

### Chromatin immunoprecipitation

In preparation for immunoprecipitation 60  $\mu\text{L}$  of Dynabeads-Protein A (Thermo Fisher Scientific, #10002D) per sample per histone mark was washed with 30  $\mu\text{L}$  of ChIP dilution buffer (1.1% (v/v) Triton X-100, 1.2 mM EDTA at pH 8.0, 16.7 mM Tris-HCl at pH 8.0, 167 mM NaCl, 1 mM PMSE, 1 mM Pepstatin A, 1 mini-complete tablet per 10 mL buffer) (Additional File 12: Table A4). Samples were pre-cleared by adding 90  $\mu\text{L}$  of the washed Dynabeads-Protein A and incubating for 4.5 h, rotating, at  $4^{\circ}\text{C}$ . Eppendorf tubes were labelled with the appropriate sample name and histone mark and 60  $\mu\text{L}$  aliquots of the washed Dynabeads-Protein A were made into each tube followed by the addition of 2.5  $\mu\text{L}$  (2.5  $\mu\text{g}$ ) of either Anti-H3K4me3 (Millipore<sup>®</sup>, 07-473), Anti-H3K56ac (Millipore<sup>®</sup>, 07-677-1), Anti-H3K27me3 (Millipore<sup>®</sup>, 07-449), or Anti-H2A.Z (Millipore<sup>®</sup>, 07-594) to their correspondingly labelled tubes. The tubes were then sealed and incubated at  $4^{\circ}\text{C}$ , rotated, for 1 h.

Following pre-clearing, samples were then placed on a magnetic stand (Thermo Fisher Scientific, #AM10027) for 2 min until all Dynabead-Protein A precipitate out of solution. Next, 300  $\mu\text{L}$  of the supernatant, of each tissue type, was then transferred to a new tube to use as an input control. One input control was made for each of the tissue types used in the experiment. The INPUT controls were then immediately placed in storage at  $-20^{\circ}\text{C}$ . The remaining 3.6 mL of supernatant was then made into  $4 \times 900$   $\mu\text{L}$  aliquots and transferred to the previously labelled, tubes, containing the Dynabeads-Protein A/antibody slurry and incubated at  $4^{\circ}\text{C}$ , rotating, for 90 min. The samples were then placed on a magnetic stand for 2 min and the supernatant was removed. On the stand, 1 mL of low salt wash buffer (150 mM NaCl, 0.1% (w/v) SDS, 1% (w/v) TritonX-100, 2 mM EDTA at pH 8.0, 20 mM Tris-HCl at pH 8.0) was added to each tube (Additional File 12: Table A5). The tubes were then sealed, and the contents were resuspended by inverting and then placed on the magnetic stand for 2 min until all the Dynabeads-Protein A/antibody/chromatin complexes precipitate out of solution and the supernatant discarded. The wash step was repeated one more time using the low salt wash buffer, then again using the high salt wash buffer (500 mM NaCl, 0.1% SDS, 1% TritonX-100, 2 mM EDTA at pH 8.0, 20 mM Tris-HCl at pH 8.0) (Additional File 12: Table A6), followed by a final wash using the LiCl wash buffer (0.25 M LiCl, 1% (v/v) NP-40 (Sigma-Aldrich NP40S-100 mL), 1% Sodium Deoxycholate, 1 mM EDTA at pH 8.0, 10 mM Tris-HCl at pH 8.0) (Additional File 12: Table A7). The beads were then washed using 1 mL of TE

buffer (10 mM Tris-HCl at pH 8.0, 1 mM EDTA at pH 8.0) (Additional File 12: Table A8), allowed to precipitate on the magnetic stand for 2 min, followed by aspiration of the supernatant. Each sample was then removed from the magnetic stand and the pellets were resuspended with 150  $\mu\text{L}$  of SDS elution buffer (1% SDS, 0.1 M  $\text{NaHCO}_3$ ) (Additional File 12: Table A9) and incubated at  $65^{\circ}\text{C}$  for 15 min. The samples were then placed on the magnetic stand for 2 min and the supernatant was then transferred to a newly labelled 1.5 mL tube.

A second elution of the beads was performed using 150  $\mu\text{L}$  of SDS elution buffer. The supernatants were combined in 1.5 mL tube for a final volume of 300  $\mu\text{L}$ . A master mix solution containing 12  $\mu\text{L}$  5 M NaCl, 30  $\mu\text{L}$  Dithiothreitol (DTT), and 30  $\mu\text{L}$  1 M  $\text{NaHCO}_3$  per sample was made up for all samples, including INPUT controls. 72  $\mu\text{L}$  of the master mix was aliquoted into each of the samples and INPUT controls and allowed to incubate overnight at  $65^{\circ}\text{C}$ . Samples and INPUT controls were retrieved from overnight incubation and 6  $\mu\text{L}$  0.5 M EDTA, 12  $\mu\text{L}$  1 M Tris pH 7.0, and 2  $\mu\text{L}$  proteinase K was added to each tube. Tubes were then incubated at  $45^{\circ}\text{C}$  for 1 h. Samples and INPUT controls were then cleaned up by adding 350  $\mu\text{L}$  chloroform/ isoamyl alcohol (24:1), vortexing, and centrifuging using a bench top centrifuge at max speed for 25 min at room temperature. The top layer was then carefully, without disturbing the interphase, transferred to a new tube. 2  $\mu\text{L}$  of glycogen, 60  $\mu\text{L}$  of 3 M sodium acetate pH 5.2, and 900  $\mu\text{L}$  of chilled ethanol (95%) was added to each sample and INPUT control and then allowed to precipitate overnight at  $-20^{\circ}\text{C}$ . Samples and INPUT controls were then centrifuged at 16,000 g, for 30 min, at  $4^{\circ}\text{C}$ . The DNA pellet was then washed two times using 1 mL of ethanol (70%). The washed DNA pellets were then allowed to dry in a fume hood. The dry DNA pellets were then resuspended in 50  $\mu\text{L}$  TE buffer. Double stranded DNA (dsDNA) content was quantified with the Qubit dsDNA high sensitivity assay using the Qubit 4 fluorometer (See Additional File 1). ChIP-seq libraries were then generated using the Accel-NGS 2 S Plus DNA library Kit (Swift biosciences), following the manufacture's recommendations using 10 pg – 250 ng of input dsDNA per sample.

### Assessing library quality – tapestation

The quality of the sequencing libraries was assayed using a 2200 TapeStation (Agilent) and D1000 screen tape. Samples were prepared using 1  $\mu\text{L}$  of each sample library and diluted using 3  $\mu\text{L}$  of D1000 sample buffer for a final volume of 4  $\mu\text{L}$ . Tubes were then sealed and vortexed for 30 s. The samples were then spun down in a microfuge for 1 min to ensure residual droplets on the tube walls were collected at the base of each tube. Sample were then run on the TapeStation, and the results recorded.

### Chromatin immunoprecipitation sequencing operations

Samples were sequenced on an Illumina next-seq according to manufacturer's instructions. We applied the ENCODE consortium guidelines for broad (H3K37me3, H3K56ac, H2A.Z) and narrow peak (H3K4me3) type data in humans of >45 million reads and >20 million reads are recommended respectively. For *C. sativa* the target reads were linearly scaled approximately 4 times to >10 million reads and >5 million reads respectively to account for the genome size discrepancy between human and *C. sativa*.

### RNA sequencing library preparation and sequencing operations

All samples were homogenised using Geno/Grinder 2010 (SPEX SamplePrep) and total RNA was isolated from homogenised samples using the Sigma Spectrum Plant Total RNA kit (Sigma) supplemented with the On-Column DNase I Digestion step to remove genomic DNA. RNA was eluted in 50 µL EB buffer. RNA concentration was measured using a Nanodrop spectrophotometer. RNA-seq library generation was performed using Illumina TruSeq Stranded mRNA Library Prep and indexed using Truseq RNA UD Indexes (96 indexes, 96 samples). Individual libraries were quality checked using Qubit dsRNA HS Assay kit and Agilent TapeStation (D1000) before being pooled into one sample for sequencing using Illumina NextSeq 500/550 High Output kit v2 (75 cycles).

### ChIP-seq data processing

Raw sequencing reads were quality trimmed using Trim Galore version 0.6.3 and the following parameters were used for paired end reads -paired -trim1 -fastqc [104]. The -fastqc option was selected to provide a quality report of the trimmed reads after trimming.

Quality trimmed reads were then aligned to the cs10 version 2 reference genome GCF\_900626175.2 using Bowtie2 version 2.3.5.1 using the default parameters for paired end reads. Bowtie2 output SAM files were processed using Samtools version 1.9. SAM files were first converted to BAM files using the Samtools view -Sb function. Samtools view output BAM files were then sorted using the Samtools sort function [105]. The aligned reads were then filtered for PCR duplicates. Using the Picard version 2.2.2 suite of tools PCR duplicates were marked and subsequently removed from the aligned files using the MarkDuplicates function with the following parameters: TAG\_DUPLICATE\_SET\_MEMBERS=true TAGGING\_POLICY=All REMOVE\_DUPLICATES=true ASSUME\_SORT\_ORDER=coordinate READ\_NAME\_REGEX=null INDEX=true.

### RNA-seq data processing

The quality of the raw RNA-seq data was assessed using FastQC v0.11.9 [106]. The data was aligned in single-end

mode against the *Cannabis sativa* reference genome (cv. cs10; accession number GCF\_900626175.2) using HISAT2 v2.1.0 (default parameters) [107]. The mapped reads were sorted using Samtools v1.9 and transcript per million (TPM) counts were generated using StringTie v2.1.3b [105, 108]. Gene-level quantification was performed on the mapped reads using featureCounts from the Subread v2.0.0 package [109]. Principal component analysis (PCA) plots were generated in R v4.1 using the normalised read counts from the DESeq2 v1.44.0 R package [110].

### Peak calling

Narrow peak type data was called for each biological replicate for the narrow peak type H3K4me3 and for the mixed peak type data H3K56ac using model based analysis of ChIP-seq (MACS2 2.2.7) using the parameters bam paired end function -f BAMPE and mappable genome size -g 736,579,359 -B and a relaxed p value cut off -p 0.1 as indicated for downstream irreproducible discovery rate (IDR) analysis [111]. Statistically significant, replicable peaks were then estimated using the irreproducible discovery rate (IDR 2.0.4.2) [112]. An IDR threshold or false discovery rate of 0.05 was applied for transcription factor-like H3K4me3 narrow peaks.

There was no community standardised approach for handling mixed peak types like H3K56ac. We consequently handled the dataset as a narrow peak type dataset, as we were primarily interested in the narrow domain functionality of H3K56ac [113]. As IDR is considered particularly stringent in calling reproducible peaks for narrow peak type data, and there is inherent broad domain noise in H3K56ac data, we increased the IDR or false discovery rate cutoff threshold to 0.1 for H3K56ac datasets.

The ENCODE consortium does not have guidelines on how to process biological replicates for H3K27me3 and H2A.Z broad domains. We consequently took a conservative approach to identifying reproducible broad domains between biological replicates for both H2A.Z and H3K27me3. We used spatial clustering for identification of ChIP-seq regions (SICER), epic2 0.0.52 specifically designed for broad domain peak type data using default parameters. Prior to epic2 calling of broad domains we scaled each replicate and its corresponding input control, as epic2 does not have auto-scaling functionality like MACS2 software, using the deepTools 3.5.1 bamcoverage option and --scalefactor parameter. We then identified the replicable broad domains between the replicates using BEDTools 2.3.0 intersect option requiring that broad domains share at least 30% reciprocal overlaps -f 0.3 -r.

### Gene and KEGG ontology analysis

To identify genes associated with various chromatin states we used BEDTools 2.3.0 intersect function to find genes genome-wide that overlapped with replicable peaks for each histone mark/modification. We then used these gene lists to perform Gene ontology and Kyoto encyclopedia of genes and genomes (KEGG) ontology analysis. We used the R-Package ClusterProfiler 4.0 [114] to perform gene ontology analysis using an in-house gene ontology database for the cs10 reference assembly (GCF\_900626175.2) created using PANNZER2 [115]. Similarly, we used ClusterProfiler 4.0 to perform KEGG pathway enrichment using the available *C. sativa* KEGG ontologies for the cs10 reference assembly (GCF\_900626175.2). For glandular trichome specific analysis we used BEDTools intersect -v function to report peaks that only occurred in glandular trichomes compared to stems and with that output we applied the same operation with the leaf dataset to identify glandular trichome specific peaks for each chromatin type. The genes intersecting the glandular trichome specific chromatin peaks were then subject to ontology analysis as above.

### Differential gene expression and gene set enrichment analysis

Differential gene expression was carried out on the RNA-seq datasets using the R-package DESeq2 with default parameters and selecting differentially expressed genes a >2 log fold change and q-value > 0.05 [110]. The DESeq2 output results were then brought forward for gene set enrichment analysis using the ClusterProfiler 4.0 package ridgeplot function using the PANNZER2 cs10 (GCF\_900626175.2) gene ontology database.

### Cis-regulatory motif analysis

Glandular trichome specific distal ( $\geq 1.5$  Kbp from TSS) H3K56ac peaks were determined using the BEDTools 2.3.0 intersect -v function. The nucleotide sequences within peaks were extracted using the BEDTools 2.3.0 getfasta option. Similarly, the BEDTools 2.3.0 getfasta option was used to determine the nucleotide sequences of all glandular trichome H3K56ac peaks which were then used as background control for motif discovery. Motif discovery was carried out using MEME suite tools [116]. Glandular trichome specific H3K56ac chromatin sequences were enriched for motifs using the simple enrichment analysis (SEA) with default settings, cross-referencing the JASPAR CORE (2022) plants non-redundant motif database [116]. The significantly ( $q < 0.05$ ) motifs were then plotted by transcription factor family vs.  $-\log_{10}(Q\text{-value})$  using the beeswarm R-package.

### Distal enhancer and target prediction leveraging multiomic datasets

Using the list of glandular trichome specific H3K56ac peaks we then determined gene distal peaks occurring  $\geq 1.5$  kb from gene TSSs using the hypergeometric optimization of motif enrichment (HOMER) Perl script getDistalPeaks.pl using the parameters -gid -d 1500, then re-run using the parameter -targets to produce a list of the nearest gene-distal H3K56ac peaks [117]. We then used a custom python script to match the list of putative target genes to their corresponding DESeq2 differentially expressed gene ID in glandular trichome vs. stem and glandular trichome vs. leaf DESeq2 outputs. The list of genes was then filtered to include only those genes that show transcriptomic evidence of strong glandular trichome specific gene expression ( $\log_2FC > 2$ ;  $p\text{-adj} < 0.05$ ) in both trichome vs. stem and trichome vs. leaf datasets.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05787-x>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8  
Supplementary Material 9  
Supplementary Material 10  
Supplementary Material 11  
Supplementary Material 12

### Acknowledgements

LJC received a PhD scholarship from La Trobe University Graduate Research School. We thank Asha Haslem for providing sequencing services at the La Trobe University Genomics Platform.

### Author contributions

L.J.C. and M.G.L. Conceptualized and designed the study. L.J.C. performed the chromatin immunoprecipitation of each histone mark, prepared the sequencing libraries, and performed genomic alignment of the reads. M.O. extracted RNA from each of the samples, S.N. prepared the RNA sequencing libraries, and B.H. performed alignment of the RNA sequencing reads to genome and prepared public web browser. L.J.C. performed normalization and peak calling of the ChIP-seq reads, differential gene expression of the RNA-seq reads, gene ontology analysis of both RNA-seq and ChIP-seq data, motif discovery enrichment, and integration of RNA-seq and ChIP-seq data. L.J.C. wrote the original manuscript draft and M.G.L. edited the original draft manuscript. L.J.C. wrote the final version. All authors read and approved the final manuscript.

### Funding

Work in the Lewsey lab is funded by the Australian Research Council Industrial Transformation Hub in Medicinal Agriculture (IH180100006).

**Data availability**

Chromatin immunoprecipitation sequencing data was made publicly available on the sequencing read archive (SRA) under BioProject ID PRJNA1128358 complimentary RNA sequencing datasets were registered under BioProject ID PRJNA1128734. Web browser containing all alignments, peak calls, peak replicability, and differential epigenomic analysis [https://jbrow.se.latrobe.edu.au/cannabis\\_trichome\\_epigenome/](https://jbrow.se.latrobe.edu.au/cannabis_trichome_epigenome/).

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare no competing interests.

Received: 12 July 2024 / Accepted: 5 November 2024

Published online: 14 November 2024

**References**

- Zhao T, Zhan Z, Jiang D. Histone modifications and their regulatory roles in plant development and environmental memory. *J Genet Genomics*. 2019;46(10):467–76.
- Halsall JA, Andrews S, Krueger F, Rutledge CE, Ficz G, Reik W, et al. Histone modifications form a cell-type-specific chromosomal bar code that persists through the cell cycle. *Sci Rep*. 2021;11(1):3009.
- Ikeuchi M, Iwase A, Sugimoto K. Control of plant cell differentiation by histone modification and DNA methylation. *Curr Opin Plant Biol*. 2015;28:60–7.
- Zhang A, Wei Y, Shi Y, Deng X, Gao J, Feng Y et al. Profiling of H3K4me3 and H3K27me3 and their roles in gene subfunctionalization in allotetraploid cotton. *Front Plant Sci*. 2021;12.
- Hussey SG, Loots MT, van der Merwe K, Mizrahi E, Myburg AA. Integrated analysis and transcript abundance modelling of H3K4me3 and H3K27me3 in developing secondary xylem. *Sci Rep*. 2017;7(1):3370.
- Yamaguchi N, Matsubara S, Yoshimizu K, Seki M, Hamada K, Kamitani M, et al. H3K27me3 demethylases alter HSP22 and HSP17.6 C expression in response to recurring heat in Arabidopsis. *Nat Commun*. 2021;12(1):3480.
- Zhao N, Zhang K, Wang C, Yan H, Liu Y, Xu W et al. Systematic analysis of differential H3K27me3 and H3K4me3 deposition in callus and seedling reveals the epigenetic regulatory mechanisms involved in callus formation in rice. *Front Genet*. 2020;11.
- Prevalent putative chromatin. Bivalency and partial resetting of H3K27me3 in plant sperm. *Nat Genet*. 2023;55(4):536–7.
- Wang H, Fan Z, Shliha PV, Miele M, Hendrickson RC, Jiang X, et al. H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature*. 2023;615(7951):339–48.
- Ornelas-Ayala D, Cortés-Quiñones C, Olvera-Herrera J, García-Ponce B, Garay-Arroyo A, Álvarez-Buylla ER, et al. A Green Light to switch on genes: revisiting trithorax on plants. *Plants*. 2023;12(1):75.
- Ding Y, Avramova Z, Fromm M. Two distinct roles of ARABIDOPSIS HOMOLOG OF TRITHORAX1 (ATX1) at promoters and within transcribed regions of ATX1-Regulated genes. *Plant Cell*. 2011;23(1):350–63.
- Blackledge NP, Klose RJ. The molecular principles of gene regulation by polycomb repressive complexes. *Nat Rev Mol Cell Biol*. 2021;22(12):815–33.
- Berry S, Dean C, Howard M. Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor activity. *Cell Syst*. 2017;4(4):445–e578.
- Gómez-Zambrano Á, Merini W, Calonje M. The repressive role of Arabidopsis H2A.Z in transcriptional regulation depends on AtBMI1 activity. *Nat Commun*. 2019;10(1):2828.
- Espinas NA, Saze H, Saijo Y. Epigenetic control of defense signaling and priming in plants. *Front Plant Sci*. 2016;7.
- Gao Z, Li Y, Ou Y, Yin M, Chen T, Zeng X, et al. A pair of readers of bivalent chromatin mediate formation of Polycomb-based memory of cold in plants. *Mol Cell*. 2023;83(7):1109–e244.
- Hua P, Badat M, Hanssen LLP, Hentges LD, Crump N, Downes DJ, et al. Defining genome architecture at base-pair resolution. *Nature*. 2021;595(7865):125–9.
- Meng F, Zhao H, Zhu B, Zhang T, Yang M, Li Y, et al. Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in Arabidopsis thaliana. *Plant Cell*. 2021;33(6):1997–2014.
- Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015;16(3):144–54.
- Yang Y, Dai Z, Dai X. Insights into active intragenic enhancers. *Biochem Biophys Res Commun*. 2019;515(3):423–8.
- Yan W, Chen D, Schumacher J, Durantini D, Engelhorn J, Chen M, et al. Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat Commun*. 2019;10(1):1705.
- Li E, Liu H, Huang L, Zhang X, Dong X, Song W, et al. Long-range interactions between proximal and distal regulatory regions in maize. *Nat Commun*. 2019;10(1):2633.
- Huang Y, An J, Sircar S, Bergis C, Lopes CD, He X, et al. HSFA1a modulates plant heat stress responses and alters the 3D chromatin organization of enhancer-promoter interactions. *Nat Commun*. 2023;14(1):469.
- Panigrahi A, O'Malley BW. Mechanisms of enhancer action: the known and the unknown. *Genome Biol*. 2021;22(1):108.
- Fu L-Y, Zhu T, Zhou X, Yu R, He Z, Zhang P, et al. ChIP-Hub provides an integrative platform for exploring plant regulome. *Nat Commun*. 2022;13(1):3413.
- Marand AP, Zhang T, Zhu B, Jiang J. Towards genome-wide prediction and characterization of enhancers in plants. *Biochimica et Biophysica Acta (BBA) - Gene Regul Mech*. 2017;1860(1):131–9.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108–12.
- Yang X, Yan J, Zhang Z, Lin T, Xin T, Wang B, et al. Regulation of plant architecture by a new histone acetyltransferase targeting gene bodies. *Nat Plants*. 2020;6(7):809–22.
- Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants*. 2019;5(12):1250–9.
- Zhou C, Yuan Z, Ma X, Yang H, Wang P, Zheng L, et al. Accessible chromatin regions and their functional interrelations with gene transcription and epigenetic modifications in sorghum genome. *Plant Commun*. 2021;2(1):100140.
- Ghosh D, Chaudhary N, Shanker K, Kumar B, Kumar N. Monoecious Cannabis sativa L. discloses the organ-specific variation in glandular trichomes, cannabinoids content and antioxidant potential. *J Appl Res Med Aromatic Plants*. 2023;35:100476.
- Crocq MA. History of cannabis and the endocannabinoid system. *Dialogues Clin Neurosci*. 2020;22(3):223–8.
- Pain S. A potted history. *Nature*. 2015;525(7570):S10–1.
- Bilbao A, Spanagel R. Medical cannabinoids: a pharmacology-based systematic review and meta-analysis for all relevant medical indications. *BMC Med*. 2022;20(1):259.
- Casajuana Köguel C, López-Pelayo H, Balcells-Olivero MM, Colom J, Gual A. Psychoactive constituents of cannabis and their clinical implications: a systematic review. *Adicciones*. 2018;30(2):140–51.
- Schuurink R, Tissier A. Glandular trichomes: micro-organs with model status? *New Phytol*. 2020;225(6):2251–66.
- Feng Z, Bartholomew ES, Liu Z, Cui Y, Dong Y, Li S, et al. Glandular trichomes: new focus on horticultural crops. *Hortic Res*. 2021;8(1):158.
- Mahmoud SS, Maddock S, Adal AM. Isoprenoid metabolism and engineering in glandular trichomes of lamiaceae. *Front Plant Sci*. 2021;12.
- Dhankhar R, Regmi K, Kawatra A, Gulati P. Trichomics: trichomes as natural chemical factories. In: Swamy MK, Kumar A, editors. *Phytochemical Genomics: Plant Metabolomics and Medicinal Plant Genomics*. Singapore: Springer Nature Singapore; 2022. pp. 379–402.
- Schwabe AL, Johnson V, Harrelson J, McGlaughlin ME. Uncomfortably high: testing reveals inflated THC potency on retail Cannabis labels. *PLoS ONE*. 2023;18(4):e0282396.
- Lloyd JPB, Lister R. Epigenome plasticity in plants. *Nat Rev Genet*. 2022;23(1):55–68.
- Eberharder A, Becker PB. Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep*. 2002;3(3):224–9.
- Forozani M, Vandal MP, Smith AP. H3K4 trimethylation dynamics impact diverse developmental and environmental responses in plants. *Planta*. 2021;253(1):4.

44. Liu S, Trejo-Arellano MS, Qiu Y, Eklund DM, Köhler C, Hennig L. H2A ubiquitination is essential for polycomb repressive complex 1-mediated gene regulation in *Marchantia polymorpha*. *Genome Biol.* 2021;22(1):253.
45. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglu S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813–31.
46. Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, Laflamme-Yonkman J, et al. Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *Plant J.* 2020;101(1):37–56.
47. Braich S, Baillie RC, Jewell LS, Spangenberg GC, Cogan NOI. Generation of a Comprehensive Transcriptome Atlas and Transcriptome dynamics in Medicinal Cannabis. *Sci Rep.* 2019;9(1):16583.
48. Zager JJ, Lange I, Srividya N, Smith A, Lange BM. Gene Networks underlying cannabinoid and Terpenoid Accumulation in Cannabis. *Plant Physiol.* 2019;180(4):1877–97.
49. Chen W, Tamada Y, Yamane H, Matsushita M, Osako Y, Gao-Takai M, et al. H3K4me3 plays a key role in establishing permissive chromatin states during bud dormancy and bud break in apple. *Plant J.* 2022;111(4):1015–31.
50. Haider S, Farrona S. Decoding histone 3 lysine methylation: insights into seed germination and flowering. *Curr Opin Plant Biol.* 2024;81:102598.
51. Giaimo BD, Ferrante F, Herchenröther A, Hake SB, Borggrefe T. The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin.* 2019;12(1):37.
52. Jarillo JA, Piñeiro M. H2A.Z mediates different aspects of chromatin function and modulates flowering responses in *Arabidopsis*. *Plant J.* 2015;83(1):96–109.
53. Zhao K, Kong D, Jin B, Smolke CD, Rhee SY. A novel bivalent chromatin associates with rapid induction of camalexin biosynthesis genes in response to a pathogen signal in *Arabidopsis*. *eLife.* 2021;10:e69508.
54. Topal S, Vasseur P, Radman-Livaja M, Peterson CL. Distinct transcriptional roles for histone H3-K56 acetylation during the cell cycle in yeast. *Nat Commun.* 2019;10(1):4372.
55. Méteignier L-V, Nützmann H-W, Papon N, Osbourn A, Courdavault V. Emerging mechanistic insights into the regulation of specialized metabolism in plants. *Nat Plants.* 2023;9(1):22–30.
56. Qiao D, Tang M, Jin L, Mi X, Chen H, Zhu J, et al. A monoterpene synthase gene cluster of tea plant (*Camellia sinensis*) potentially involved in constitutive and herbivore-induced terpene formation. *Plant Physiol Biochem.* 2022;184:1–13.
57. Reichardt S, Budahn H, Lamprecht D, Riewe D, Ulrich D, Dunemann F, et al. The carrot monoterpene synthase gene cluster on chromosome 4 harbours genes encoding flavour-associated sabinene synthases. *Hortic Res.* 2020;7(1):190.
58. Chen H, Köllner TG, Li G, Wei G, Chen X, Zeng D, et al. Combinatorial evolution of a terpene synthase gene cluster explains terpene variations in *Oryza 1* [OPEN]. *Plant Physiol.* 2019;182(1):480–92.
59. Liu Z, Cheema J, Vigouroux M, Hill L, Reed J, Paajanen P, et al. Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat Commun.* 2020;11(1):5354.
60. Bharadwaj R, Kumar SR, Sharma A, Sathishkumar R. Plant metabolic gene clusters: evolution, organization, and their applications in synthetic biology. *Front Plant Sci.* 2021;12.
61. Polturak G, Osbourn A. Defense-related phenylpropanoid biosynthetic gene clusters in rice. *Sci Bull (Beijing).* 2022;67(1):13–6.
62. Wegel E, Koumproglou R, Shaw P, Osbourn A. Cell type-specific chromatin decondensation of a metabolic gene cluster in Oats. *Plant Cell.* 2009;21(12):3926–36.
63. Zhan C, Shen S, Yang C, Liu Z, Fernie AR, Graham IA, et al. Plant metabolic gene clusters in the multi-omics era. *Trends Plant Sci.* 2022;27(10):981–1001.
64. Conneely LJ, Berkowitz O, Lewsey MG. Emerging trends in genomic and epigenomic regulation of plant specialised metabolism. *Phytochemistry.* 2022;203:113427.
65. Nützmann H-W, Osbourn A. Regulation of metabolic gene clusters in *Arabidopsis thaliana*. *New Phytol.* 2015;205(2):503–10.
66. Sands LB, Haiden SR, Ma Y, Berkowitz GA. Hormonal control of promoter activities of Cannabis sativa prenyltransferase 1 and 4 and salicylic acid mediated regulation of cannabinoid biosynthesis. *Sci Rep.* 2023;13(1):8620.
67. Rea KA, Casaretto JA, Al-Abdul-Wahid MS, Sukumaran A, Geddes-McAlister J, Rothstein SJ, et al. Biosynthesis of cannflavins A and B from Cannabis sativa L. *Phytochemistry.* 2019;164:162–71.
68. Medina-Puche L, Molina-Hidalgo FJ, Boersma M, Schuurink RC, López-Vidriero I, Solano R, et al. An R2R3-MYB transcription factor regulates Eugenol Production in Ripe Strawberry Fruit receptacles. *Plant Physiol.* 2015;168(2):598–614.
69. He Y, Yu Y, Wang X, Qin Y, Su C, Wang L. Aschoff's rule on circadian rhythms orchestrated by blue light sensor CRY2 and clock component PRR9. *Nat Commun.* 2022;13(1):5869.
70. van Velzen R, Schranz ME. Origin and evolution of the cannabinoid oxidocyclase gene family. *Genome Biol Evol.* 2021;13(8).
71. Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Timothy Motley S, et al. A new Cannabis genome assembly associates elevated cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol.* 2021;230(4):1665–79.
72. Booth JK, Page JE, Bohlmann J. Terpene synthases from Cannabis sativa. *PLoS ONE.* 2017;12(3):e0173911.
73. Aminfar Z, Rabiei B, Tohidfar M, Mirjalili MH. Identification of key genes involved in the biosynthesis of triterpenic acids in the mint family. *Sci Rep.* 2019;9(1):15826.
74. Booth JK, Yuen MMS, Jancsik S, Madilao LL, Page JE, Bohlmann J. Terpene synthases and terpene variation in Cannabis sativa. *Plant Physiol.* 2020;184(1):130–47.
75. Maoz I, Lewinsohn E, Gonda I. Amino acids metabolism as a source for aroma volatiles biosynthesis. *Curr Opin Plant Biol.* 2022;67:102221.
76. Wang M, Zhang L, Boo KH, Park E, Drakakaki G, Zakharov F. PDC1, a pyruvate/ $\alpha$ -ketoacid decarboxylase, is involved in acetaldehyde, propanal and pentanal biosynthesis in melon (*Cucumis melo* L.) fruit. *Plant J.* 2019;98(1):112–25.
77. Bizzio LN, Tieman D, Munoz PR. Branched-chain volatiles in fruit: a molecular perspective. *Front Plant Sci.* 2022;12.
78. Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. *Plant J.* 2012;71(3):353–65.
79. Gülcik T, Möller BL. Phytocannabinoids. Origins and biosynthesis. *Trends Plant Sci.* 2020;25(10):985–1004.
80. Borrego EJ, Robertson M, Taylor J, Schultzhause Z, Espinoza EM. Oxylipin biosynthetic gene families of Cannabis sativa. *PLoS ONE.* 2023;18(4):e0272893.
81. Welling MT, Deseo MA, O'Brien M, Clifton J, Bacic A, Doblin MS. Metabolomic analysis of methyl jasmonate treatment on phytocannabinoid production in Cannabis sativa. *Front Plant Sci.* 2023;14.
82. Pascal S, Bernard A, Sorel M, Pervent M, Vile D, Haslam RP, et al. The Arabidopsis cer26 mutant, like the cer2 mutant, is specifically affected in the very long chain fatty acid elongation process. *Plant J.* 2013;73(5):733–46.
83. Joubès J, Raffaele S, Bourdenx B, Garcia C, Laroche-Traineau J, Moreau P, et al. The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling. *Plant Mol Biol.* 2008;67(5):547–66.
84. Gao J, Ajjawi I, Manoli A, Sawin A, Xu C, Froehlich JE, et al. FATTY ACID DESATURASE4 of Arabidopsis encodes a protein distinct from characterized fatty acid desaturases. *Plant J.* 2009;60(5):832–9.
85. Ting T-Y, Baharin A, Ramzi AB, Ng C-L, Goh H-H. Neprosin belongs to a new family of glutamic peptidase based on in silico evidence. *Plant Physiol Biochem.* 2022;183:23–35.
86. Nie YB, Ji WQ. Cloning and characterization of disease resistance protein RPM1 genes against powdery Mildew in wheat Line N9134. *Cereal Res Commun.* 2019;47(3):473–83.
87. Liang X, Xu H, Zhu S, Zheng Y, Zhong W, Li H, et al. Genetically dissecting the Novel Powdery Mildew Resistance Gene in wheat breeding line PBDH1607. *Plant Dis.* 2022;106(8):2145–54.
88. Collins CA, LaMontagne ED, Anderson JC, Ekanayake G, Clarke AS, Bond LN, et al. EPSIN1 modulates the plasma membrane abundance of FLAGELLIN SENSING2 for effective Immune Responses1 [OPEN]. *Plant Physiol.* 2020;182(4):1762–75.
89. Lane TS, Rempe CS, Davitt J, Staton ME, Peng Y, Soltis DE, et al. Diversity of ABC transporter genes across the plant kingdom and their potential utility in biotechnology. *BMC Biotechnol.* 2016;16(1):47.
90. Stukkens Y, Bultreys A, Grec Sb, Trombik T, Vanham D, Boutry M. NpPDR1, a Pleiotropic Drug Resistance-Type ATP-Binding Cassette Transporter from Nicotiana glauca, plays a major role in Plant Pathogen Defense. *Plant Physiol.* 2005;139(1):341–52.
91. Conneely LJ, Mauleon R, Mieog J, Barkla BJ, Kretschmar T. Characterization of the Cannabis sativa glandular trichome proteome. *PLoS ONE.* 2021;16(4):e0242633.
92. Camoirano A, Alem AL, Gonzalez DH, Viola IL. Arabidopsis thaliana TCP15 interacts with the MIXTA-like transcription factor MYB106/NOECK. *Plant Signal Behav.* 2021;16(10):1938432.
93. Ernst AM, Jekat SB, Zielonka S, Müller B, Neumann U, Rüping B, et al. Sieve element occlusion (SEO) genes encode structural phloem proteins

- involved in wound sealing of the phloem. *Proc Natl Acad Sci U S A*. 2012;109(28):E1980–9.
94. Zhou L, Huang Y, Wang Q, Guo D. Chromatin accessibility is associated with artemisinin biosynthesis regulation in *artemisia annua*. *Molecules*. 2021;26(4).
  95. Nützmann H-W, Doerr D, Ramírez-Colmenero A, Sotelo-Fonseca JE, Wegel E, Di Stefano M et al. Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc Nat Acad Sci*. 2020;117(24):13800–9.
  96. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*. 2017;45(W1):W55–63.
  97. Yu N, Nützmann HW, MacDonald JT, Moore B, Field B, Berriri S, et al. Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res*. 2016;44(5):2255–65.
  98. Local A, Huang H, Albuquerque CP, Singh N, Lee AY, Wang W, et al. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat Genet*. 2018;50(1):73–82.
  99. Dimopoulos N, Guo Q, Purdy SJ, Nolan M, Halimi RA, Mieog JC et al. From dawn 'til dusk: daytime progression regulates primary and secondary metabolism in *Cannabis glandular trichomes*. *J Exp Bot*. 2024; erae148.
  100. Booth JK, Yuen MMS, Jancsik S, Madilao LL, Page JE, Bohlmann J. Terpene synthases and terpene variation in *cannabis sativa*1 [OPEN]. *Plant Physiol*. 2020;184(1):130–47.
  101. Fu X, Zheng H, Wang Y, Liu H, Liu P, Li L, et al. AaABCG20 transporter involved in cutin and wax secretion affects the initiation and development of glandular trichomes in *Artemisia annua*. *Plant Sci*. 2024;339:111959.
  102. Welling MT, Deseo MA, Bacic A, Doblin MS. Untargeted metabolomic analyses reveal chemical complexity of dioecious *Cannabis* flowers. *Aust J Chem*. 2021;74(6):463–79.
  103. Vincent D, Rochfort S, Spangenberg G. Optimisation of protein extraction from medicinal cannabis mature buds for bottom-up proteomics. *Molecules*. 2019;24(4).
  104. Krueger F, TrimGalore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. 2019 [ [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
  105. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2).
  106. Andrew S, FastQC. A quality control tool for high throughput sequence data; 2010. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
  107. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60.
  108. StringTie assembles transcriptomes. *Nat Methods*. 2015;12(4):288.
  109. Liao Y, Smyth GK, Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41(10):e108.
  110. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
  111. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
  112. Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, et al. Practical guidelines for the Comprehensive analysis of ChIP-seq data. *PLoS Comput Biol*. 2013;9(11):e1003326.
  113. Jeon H, Lee H, Kang B, Jang I, Roh TY. Comparative analysis of commonly used peak calling programs for ChIP-Seq analysis. *Genomics Inf*. 2020;18(4):e42.
  114. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov*. 2021;2(3):100141.
  115. Törönen P, Medlar A, Holm L. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Res*. 2018;46(W1):W84–8.
  116. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(Web Server issue):W202–8.
  117. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.