ATAC-seq Analysis Report

Androgen receptor activation in prostate cancer: ATAC-seq (VCaP vehicle vs DHT, GSE214753)

OmicsDesk

2026-06-09

Table of contents

1 Executive Summary

This report profiles chromatin accessibility in VCaP prostate cancer cells under androgen stimulation, comparing two vehicle controls against two samples treated with 1 nM dihydrotestosterone (DHT). The data (ATAC-seq, GSE214753) were processed end to end: adapter trimming, Bowtie2 alignment, mitochondrial and blacklist filtering, MACS3 peak calling, a fixed-width iterative consensus peak set, differential accessibility, and two transcription-factor interpretation layers (chromVAR motif activity and TOBIAS footprinting). Library quality meets ENCODE ATAC-seq standards across all four samples: alignment rate 97.7 to 97.9 percent, TSS enrichment 7.4 to 8.0 (the ENCODE ideal for human is above 7), and FRiP 0.24 to 0.27. A consensus set of 122,980 reproducible accessible regions was used for all cross-sample comparisons. Androgen stimulation produced a strong and highly directional remodelling of the open chromatin landscape. Of 114,658 tested regions, 9,808 were differentially accessible at FDR below 0.05, and the change was overwhelmingly toward opening: 5,381 regions gained accessibility in DHT against only 256 that lost it. This asymmetry is the expected signature of a ligand activated transcription factor switching on an enhancer programme rather than a non specific stress response. The interpretation layers identify the driver unambiguously. chromVAR scored AR (the androgen receptor) as the single most variable motif programme of 879 tested, moving from strongly closed in vehicle to strongly open in DHT with clean separation between the two conditions. TOBIAS footprinting independently placed the androgen and steroid receptor family at the very top of differential binding, with 813 of 879 motifs significant at p below 0.05. The prostate AR pioneer and co factors FOXA1 and HOXB13 were also recovered. The accessibility data therefore recapitulate androgen receptor activation directly from open chromatin, without requiring matched expression data.

2 Methods

2.1 Sequencing Quality Control

Raw paired-end reads were processed with **fastp** (v0.23+) to remove Nextera adapter sequences and low-quality bases (Q < 20). Reads shorter than 36 bp after trimming were discarded. Per-sample reports were aggregated with **MultiQC** (v1.20+).

2.2 Alignment

Trimmed reads were aligned to the **GRCh38 (Ensembl 111)** reference genome using **Bowtie2** (v2.5+) in very-sensitive mode with `-X 2000` to capture nucleosome-spanning fragments. Only properly-paired, concordant alignments were retained.

2.3 Filtering

Aligned reads were filtered to remove (in order): unmapped reads, secondary alignments, PCR duplicates (samtools markdup), low-quality alignments (MAPQ < 30), and reads mapping to the mitochondrial chromosome. The mitochondrial fraction was recorded as a QC metric *before* removal.

2.4 Peak Calling

Peaks (accessible regions) were called with **MACS3** (v3.0+) using `--nomodel --keep-dup all` and `-f BAMPE` so paired-end fragments are used directly (no shift trick needed for PE data). Q-value threshold: 0.05.

2.5 ATAC-seq QC

Library quality was assessed against the **ENCODE ATAC-seq data standards**, with a per-sample PASS / WARN / FAIL verdict for each metric and an overall call:

- **Fragment size distribution** — healthy libraries show clear nucleosome periodicity (peaks at ~150 bp mononucleosome, ~300 bp dinucleosome).
- **TSS enrichment score** — ratio of signal at TSS center vs flanking ±2 kb. ENCODE thresholds are genome-specific: human (hg38) 7 pass / 5–7 acceptable / <5 fail; mouse (mm10) 15 / 10–15 / <10.
- **FRiP** (Fraction of Reads in Peaks) — 0.3 preferred, 0.2 acceptable.
- **Library complexity** — NRF, PBC1, PBC2 quantify how broadly reads sample the accessible genome (ENCODE: NRF > 0.9, PBC1 > 0.9, PBC2 > 3).
- **Mitochondrial fraction** — should be < 20% in well-prepared libraries. The QC table reports both the raw alignment rate and a **non-mitochondrial alignment rate**, since mito reads inflate the headline alignment number.

2.6 Peak Annotation

Peaks were annotated to genomic features using **ChIPseeker** (v1.36+) against the UCSC knownGene TxDb (`TxDb.Hsapiens.UCSC.hg38.knownGene` for human, `TxDb.Mmusculus.UCSC.mm10.knownGene` for mouse; a TxDb built from the project GTF for other organisms) for human. Promoter regions were defined as ±3 kb from TSS.

2.7 Consensus Peak Set

Per-sample MACS3 peaks were merged into a single reproducible reference using a fixed-width iterative consensus (Corces/TCGA-atlas style): summits were extended to a common width and ranked by signal, then overlapping peaks were resolved iteratively (strongest kept), retaining only regions reproducible across at least two samples. All cross-sample quantification (PCA, differential accessibility, motif activity) uses this consensus set.

2.8 Differential Accessibility

Reads were re-counted under the consensus peaks and tested for differential accessibility with **DiffBind** (v3+); counting used `summits = 250` (peaks re-centred to ±250 bp windows) with library-size normalization. The default statistical engine is **DESeq2** (edgeR is available for low-replicate or weak-signal designs); a `batch` metadata column, when present and multi-level, is added as a blocking factor. Peaks were called differentially accessible at **FDR < 0.05 and |log2 fold-change| > 1** for the configured condition contrast.

2.9 Dimensionality Reduction (PCA)

Principal-component analysis was run on the log-transformed consensus-peak count matrix to summarise sample-to-sample structure; the first two components are reported with their percent of total variance.

2.10 Differential Volcano

Differential peaks were visualised as a volcano plot (**ggplot2** with **ggrepel** labels). The most significant peaks were annotated to their nearest gene with **ChIPseeker** (UCSC TxDb) and labelled by gene symbol.

2.11 Signal Visualisation

Genome-wide coverage tracks were generated with **deepTools** `bamCoverage` (RPKM-normalised bigWig, 10 bp bins). Aggregate signal at peak centres (±2 kb) was computed with `computeMatrix` and rendered as heatmaps and average profiles with `plotHeatmap` / `plotProfile`.

2.12 Transcription-Factor Motif Activity (chromVAR)

Per-transcription-factor accessibility was scored with **chromVAR** + **motifmatchr** against the **JASPAR2024** CORE vertebrate motif set, with GC-bias correction on the matching BSgenome. chromVAR returns a bias-corrected accessibility-deviation Z-score per motif per sample, plus a variability ranking across samples — a set-level read of which TF programmes are active.

2.13 Transcription-Factor Footprinting (TOBIAS)

Transcription-factor footprints were profiled with **TOBIAS**: `ATACorrect` removes Tn5 insertion bias, `ScoreBigwig` computes a continuous footprint score, and `BINDetect` quantifies, per motif (JASPAR2024), how much footprint signal differs between conditions at base resolution. This per-site footprint view complements the set-level chromVAR activity above.

2.14 Pathway Enrichment

Genes with promoter-proximal peaks underwent over-representation testing in **GO Biological Process**, **KEGG**, and **Reactome** using **clusterProfiler** and **ReactomePA** (BH-adjusted $p < 0.05$).

3 Quality Control

3.1 Per-Sample QC Metrics

How to read this table: Each row is one sample. **TSS enrichment** measures how strongly ATAC reads concentrate at gene start sites — ENCODE expects 7 for human (15 for mouse); values below the warn threshold indicate a noisy library. **FRiP** (Fraction of Reads in Peaks) above 0.2 means the library captured open-chromatin signal cleanly. **NRF / PBC1 / PBC2** quantify how broadly the library samples the genome (close to 1 = good, low values = a few PCR duplicates dominate). **Mito fraction** should be below 20 %; the **non-mito** alignment rate strips mitochondrial reads from the headline alignment number. The **ENCODE** column is the overall PASS / WARN / FAIL verdict (worst metric wins). Use it to spot any sample to flag before reading the downstream results.

Table 1: ATAC-seq QC metrics per sample, scored against ENCODE ATAC-seq standards (TSS 7 human / 15 mouse, FRiP 0.2, NRF 0.9, Mito 20%). The ENCODE column is the overall PASS/WARN/FAIL verdict. The full per-sample table (raw read counts, per-metric verdicts, PBC1, PBC2) ships in `results_workbook.xlsx`.

Sample	Reads (M)	Aligned	Non-mito	Mito	FRiP	TSS	NRF	ENCODE
VEH_rep1	65.9	97.8%	79.3%	18.9%	0.25	8.01	0.78	WARN
VEH_rep2	61.3	97.8%	77.5%	20.7%	0.24	7.43	0.77	WARN
DHT_rep1	99.9	97.7%	82.0%	16.1%	0.27	7.50	0.78	WARN
DHT_rep2	81.2	97.9%	82.0%	16.2%	0.26	7.48	0.79	WARN

3.2 Fragment Size Distribution

How to read this figure: The x-axis is fragment length in base pairs; the y-axis is the count of fragments at that length. A healthy ATAC library has a clear ladder: a tall peak below ~100 bp (nucleosome-free regions, where Tn5 cut on both sides of an open patch) followed by progressively smaller bumps near 200, 400 and 600 bp (mono-, di- and tri-nucleosomal fragments). Loss of this periodicity, or a single broad peak, suggests over- or under-tagmentation and a noisier signal downstream.

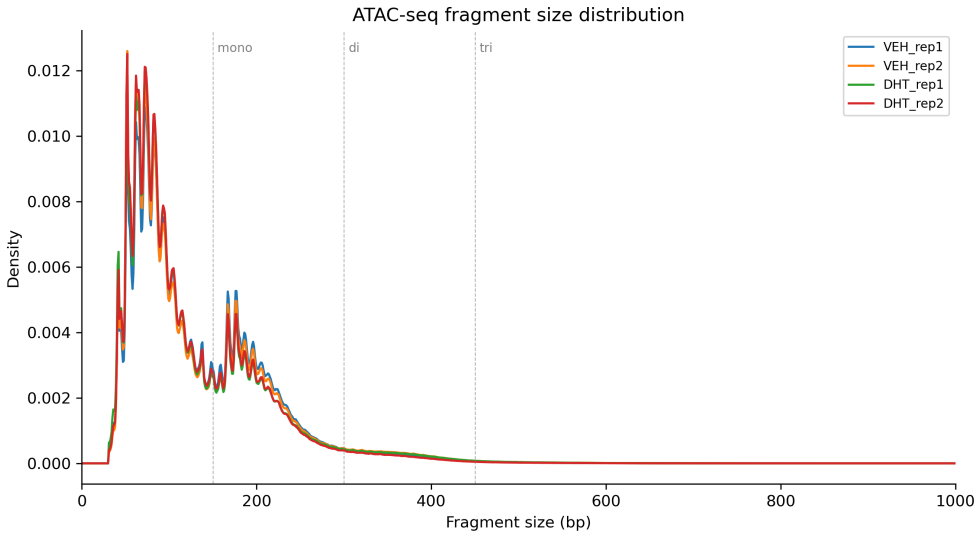

Figure 1: Fragment-size distribution of properly-paired ATAC-seq fragments. Each curve is one sample; the x-axis is fragment length (bp) and the y-axis is the density of fragments at that length. Dashed vertical lines mark the expected mono- (150 bp), di- (300 bp) and tri-nucleosome (450 bp) positions. In signal-rich libraries the sub-100 bp nucleosome-free peak dominates, so the higher-order nucleosomal bumps are comparatively small; what matters for quality is that the periodicity is present, not its absolute height.

What this means for your data: Across 4 sample(s), the median nucleosome-free fraction (<100 bp) is 49% and the median mononucleosome fraction (100-200 bp) is 34%. Clean ATAC libraries show clear periodicity at the dashed lines.

3.3 TSS Enrichment Profile

How to read this figure: Each line is one sample. The x-axis runs from −2 kb to +2 kb around every annotated transcription start site (TSS); the y-axis is ATAC signal normalised to each sample's own flanking background, so the height of the central peak *is* the TSS enrichment score (also shown in the legend) and the dotted line at y = 1 is the background level. Healthy libraries show a sharp peak at the TSS rising well above 1. Low or absent peaks point to soluble-chromatin contamination or low signal-to-noise.

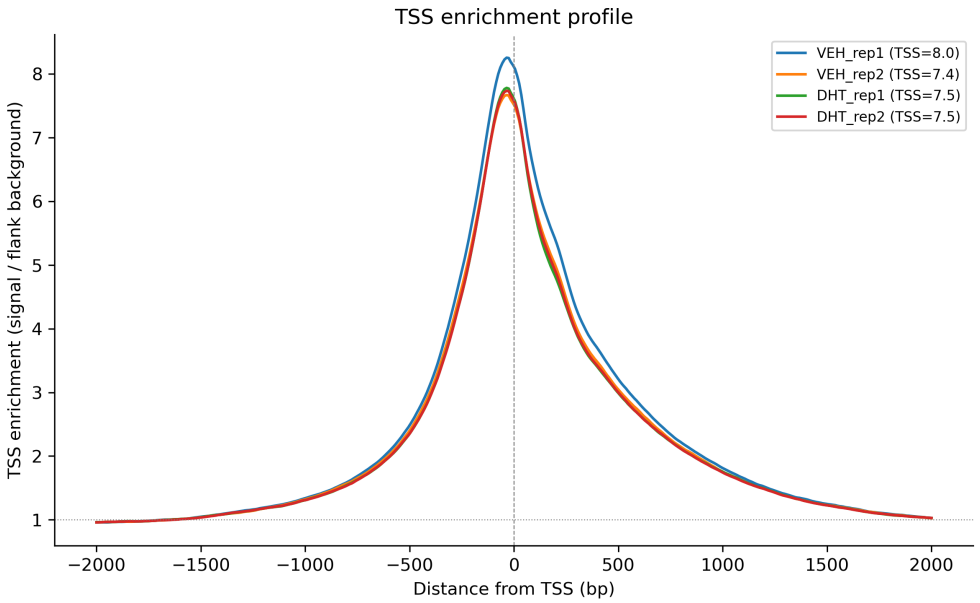

Figure 2: Mean ATAC-seq signal centred on transcription start sites (TSS ± 2 kb), one curve per sample. Signal is normalised to each sample's flanking background, so the central peak height equals the TSS enrichment score (shown in the legend) and the dotted line at y = 1 marks the background level.

What this means for your data: Median TSS enrichment across 4 sample(s): 7.5. 0/4 sample(s) fall below the ENCODE hg38 threshold of 6. Best: VEH_rep1 (8.0); worst: VEH_rep2 (7.4).

4 Peak Calling Results

How to read this table: Each row is one sample. **Total peaks** is the number of accessible regions called by MACS3 at q < 0.05 — for human ATAC the typical range is 50–150 k peaks depending on cell type and depth. **Median width** is the typical peak length in base pairs; ATAC peaks are usually 200–500 bp wide. Wide deviations between samples often track with library quality (low-complexity libraries call fewer, broader peaks).

Table 2: Number and median width of accessible regions per sample.

sample	Total peaks	Median width (bp)
DHT_rep1	157754	338
DHT_rep2	150997	310
VEH_rep1	130436	319
VEH_rep2	121504	322

5 Genomic Distribution of Accessible Regions

5.1 Feature Distribution

How to read this figure: Each slice / bar shows the share of accessible peaks falling into one genomic feature class (promoter, 5 /3 UTR, exon, intron, distal intergenic). For most cell types you expect a sizeable promoter slice (gene-regulatory regions near TSSs), a large intronic + distal intergenic share (putative enhancers), and a small UTR / exonic share. Heavy skew toward intergenic regions points to a strongly distal-enhancer-driven program; skew toward promoters points to a more constitutive accessibility profile.

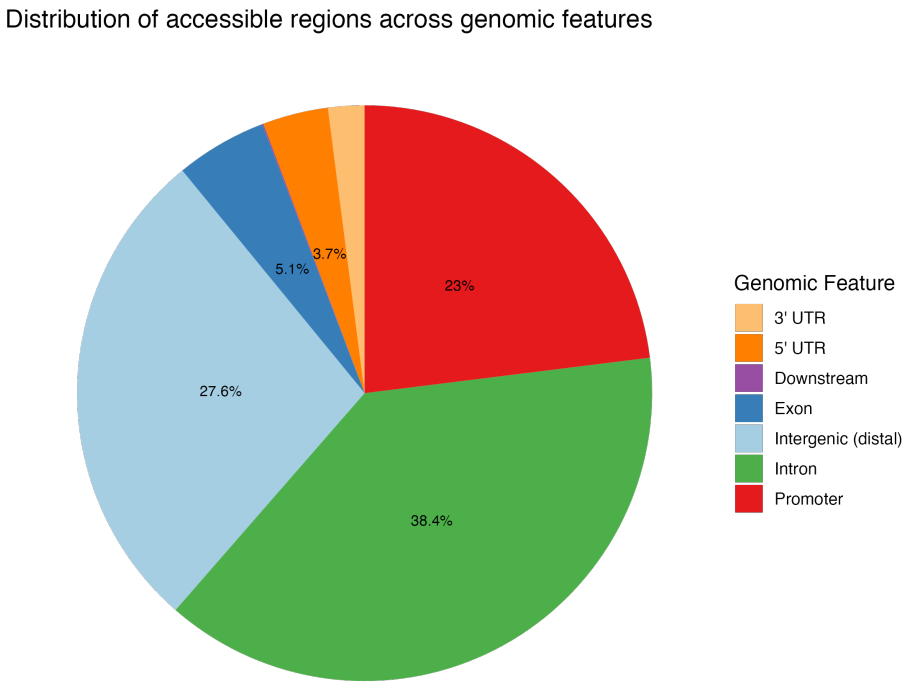

Figure 3: Genomic-feature distribution of all called accessible regions. Each slice gives the percentage of peaks annotated by ChIPseeker to a feature class (promoter, 5 /3 UTR, exon, intron, distal intergenic). Promoter peaks are gene-proximal regulatory regions; intronic and distal-intergenic peaks are candidate enhancers.

What this means for your data: 560691 peaks annotated across 4 samples; 23.0% fall in promoter regions and 27.6% in intergenic regions. The most frequently annotated target gene is AR.

5.2 Per-Sample Annotation

How to read this figure: Each column is one sample; each coloured segment is the percentage of that sample's peaks falling into a given feature class. Use it to spot samples that drift from the cohort — e.g. one library with a much larger promoter share than the rest can indicate over-fragmentation, and a strong shift in the intergenic share across conditions can be the first signal of a real enhancer-level effect.

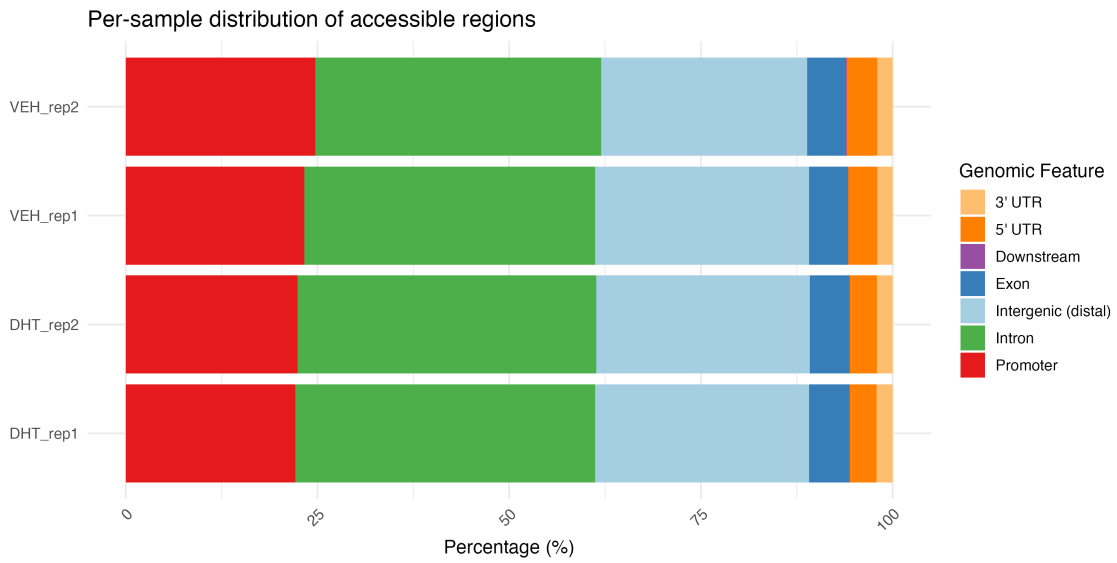

Figure 4: Per-sample genomic-feature distribution of accessible regions. Each bar is one sample; coloured segments give the percentage of that sample's peaks in each feature class. A consistent profile across samples indicates uniform library quality.

What this means for your data: The feature split is consistent across all samples, so no library shows the promoter or intergenic skew that would flag over-fragmentation or a failed sample; the conditions differ in *which* regions are open (see Differential Accessibility), not in their global feature composition.

5.3 Top Accessible Genes

How to read this table: Genes are ranked by the number of accessible peaks annotated to them (any feature class). Long genes and gene-rich regions naturally accumulate more peaks, so this list is best used as a "where the chromatin is busy" overview rather than a hit list — for differential biology look at the DiffBind section below and the per-peak annotations in `results_workbook.xlsx`.

Table 3: Top 20 genes ranked by number of nearby accessible regions.

Gene	n_peaks
AR	700
TECRL	585
SNTG1	512
COBL	476
DEFB130B	441
GRIK2	430
LINC02077	422
LINC02055	410
LINC02008	394
TMTC2	390
FAM135B	384
ADGRL3-AS1	382
CSMD3	375
NRXN1	369
PRKN	358
ROBO1	353
LINC01419	352
LINC00290	350
TRPS1-AS1	349
ARAP2	333

6 Signal at Peak Centers

How to read this figure: Each row is one merged consensus peak; each column is a position in the ±2 kb window around the peak centre, and cell colour is ATAC signal (darker = more reads). A well-resolved ATAC peak appears as a tight central column of signal that fades outward; broad uniform colour across the whole window indicates a low signal-to-noise call. Unlike the TSS enrichment profile in the QC section (which looks only at gene starts), this view summarises the signal shape across *all* accessible regions genome-wide, one panel per sample.

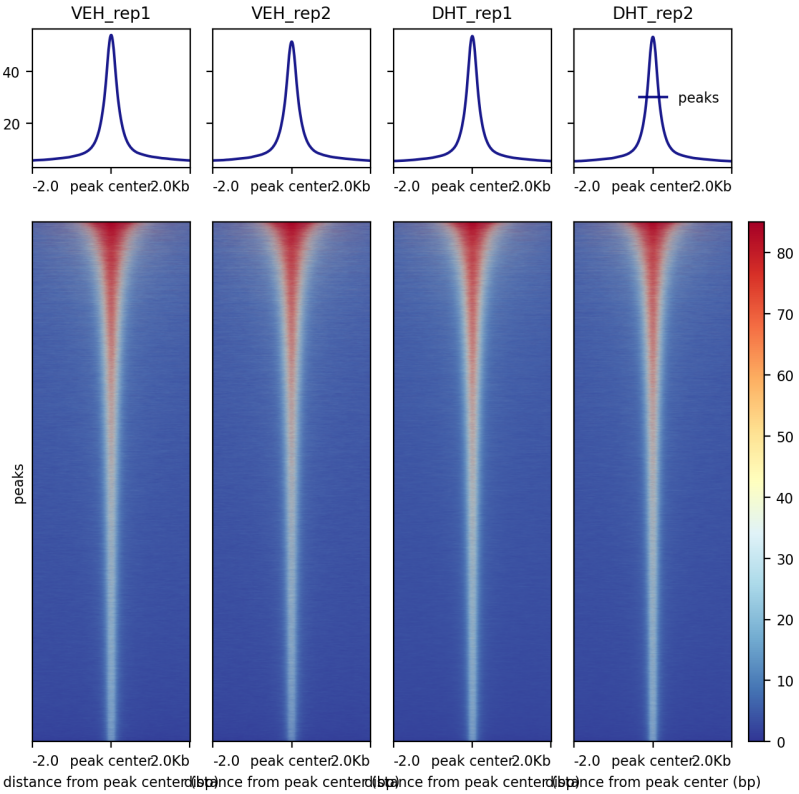

Figure 5: Heatmap of ATAC-seq signal at consensus peak centres (± 2 kb). Rows are individual peaks ordered by signal; the horizontal axis is genomic position relative to the peak centre; colour encodes ATAC read density (deepTools, RdYlBu colour map). Each sample is shown as a separate panel labelled along the top.

What this means for your data: Signal forms a tight central enrichment that decays symmetrically into the flanks in every sample, confirming the called peaks capture genuine focal accessibility rather than diffuse background.

7 Differential Accessibility

DiffBind was used to test for differentially accessible peaks between conditions (consensus peak set, DESeq2 mode). Sections render only when a contrast was configured.

7.1 Differential PCA

How to read this figure: Each dot is one sample. The two axes — PC1 and PC2 — are the largest sources of variance in the consensus-peak count matrix. Samples from the same condition should cluster together and separate from the other condition; if they don't, the biological signal is weaker than batch / technical variation and the differential calls below should be read with caution.

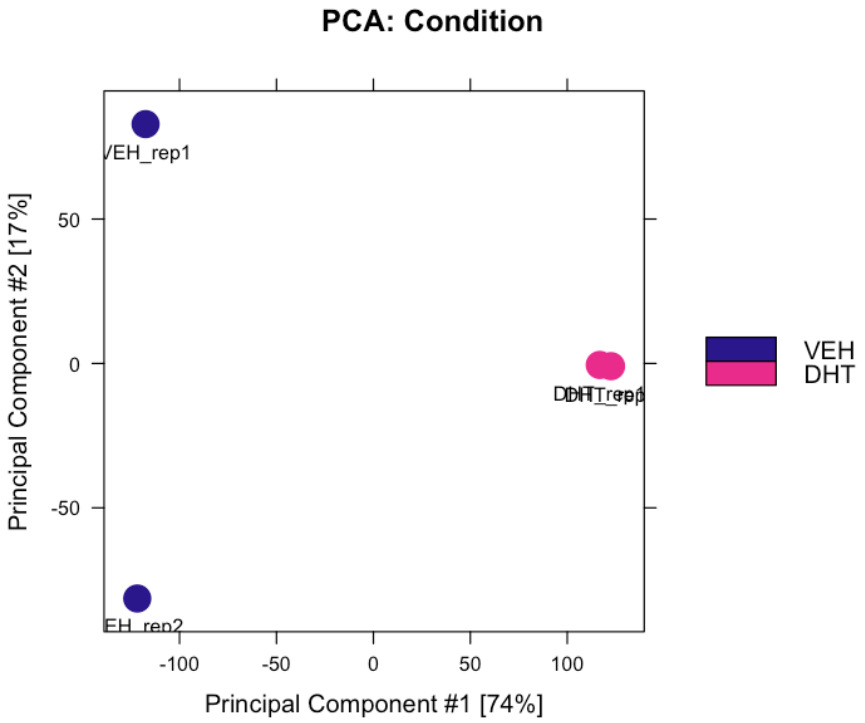

Figure 6: Principal-component analysis of the consensus-peak count matrix. Each point is one sample, coloured by condition and labelled by sample ID; the axes are the first two principal components with the percentage of total variance each explains.

What this means for your data: PC1 (73.9% of variance) cleanly separates DHT from VEH along the horizontal axis (PC1), with the replicates of each condition grouping tightly, so the treatment contrast is the dominant source of variation and the differential calls below are well powered. PC2 carries 16.9% of the variance.

7.2 Differential Volcano

How to read this figure: Each dot is one consensus peak. The horizontal axis (log2 Fold Change) shows how much accessibility moved between conditions (positive = more open in the first group, negative = more open in the second). The vertical axis (−log padj) is the statistical confidence — higher = stronger evidence. Peaks in the **top-right corner** gained accessibility, **top-left** lost it. Dashed lines mark the significance and effect-size thresholds, and the most significant peaks are labelled with their **nearest gene** so you can read the hits by name.

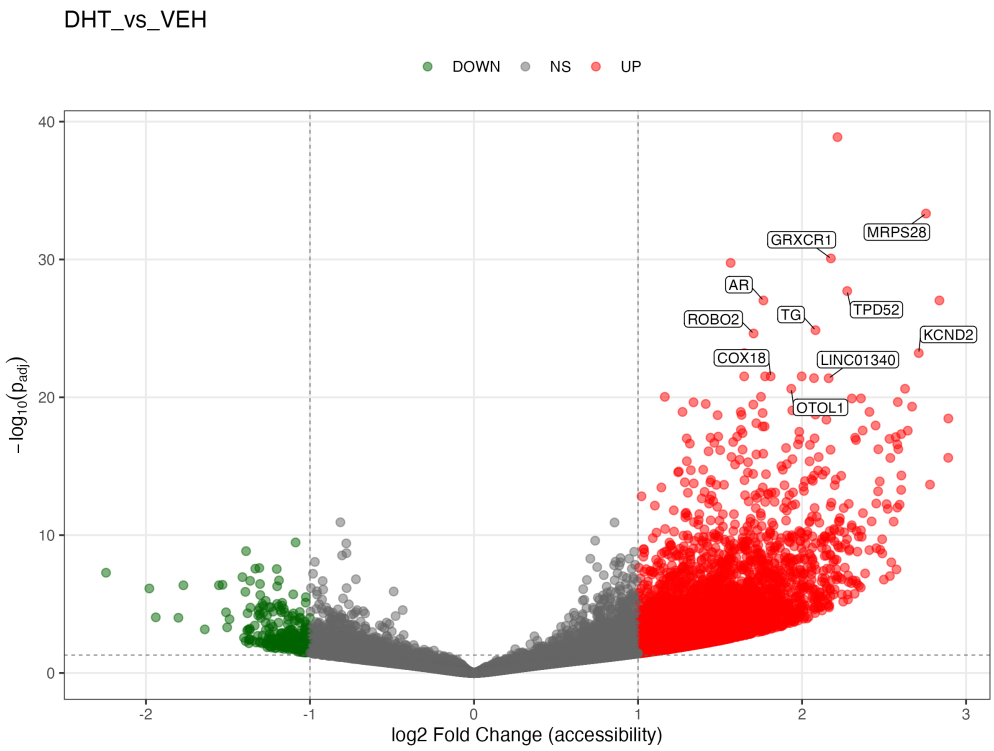

Figure 7: Volcano plot of differential chromatin accessibility across the consensus peaks. Each point is one peak; the x-axis is the log2 fold-change in accessibility between conditions and the y-axis is significance (−log10 adjusted p-value). Points are coloured by call (red = gained accessibility, dark green = lost, grey = not significant); dashed lines mark the |log2FC| and FDR thresholds, and the most significant peaks are labelled by nearest gene.

What this means for your data: Accessibility shifts strongly toward opening in DHT: 5381 peaks gained accessibility and 256 peaks lost accessibility (padj < 0.05, $|\log2FC| > 1$) in DHT vs VEH. The most significant gained regions lie near MRPS28, GRXCR1, TPD52, AR, TG.

7.3 Differential Heatmap

How to read this figure: Each row is one significant differentially accessible peak; each column is a sample. Cell colour is the peak's row-scaled accessibility (z-score across samples) on DiffBind's green scale: **pale green = below the peak's cross-sample mean, dark green = above**. Crisp condition-wise blocks (one shade per group) indicate a strong, reproducible programme; salt-and-pepper patterns indicate a noisier or more sample-specific signal.

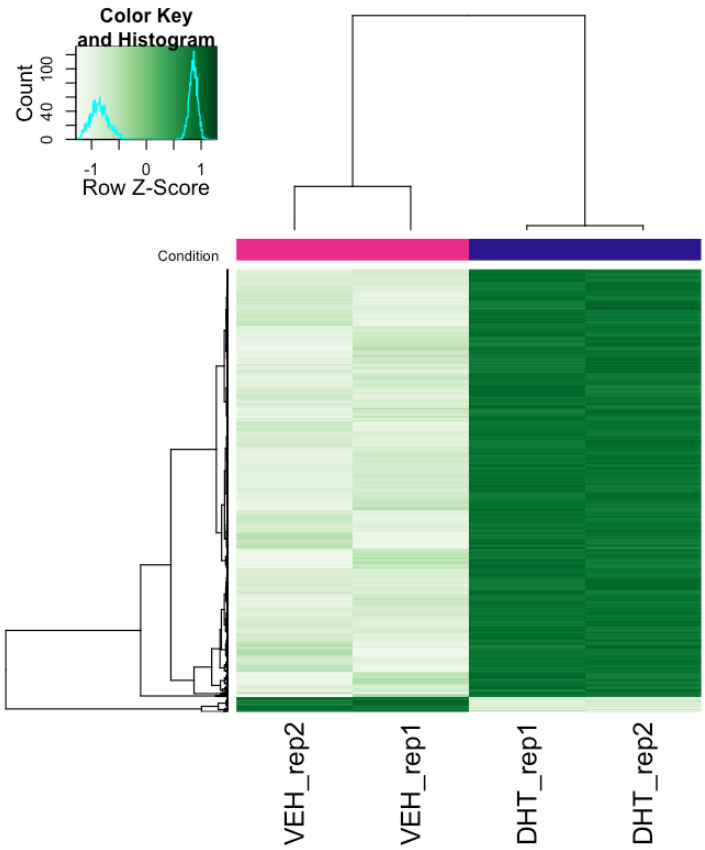

Figure 8: Heatmap of the significant differentially accessible peaks (FDR < 0.05). Rows are peaks and columns are samples, both hierarchically clustered; colour encodes row-scaled accessibility (z-score across samples) on DiffBind's sequential green palette (pale = below the row mean, dark = above).

What this means for your data: The 5637 differentially accessible peaks split the samples into two clean condition blocks: DHT and VEH samples each share one accessibility pattern, confirming the change is reproducible across replicates rather than driven by a single sample.

8 TF Footprinting (TOBIAS)

Transcription-factor footprints were profiled with **TOBIAS** (ATACorrect Tn5 bias correction, ScoreBigwig, BINDetect) against the JASPAR2024 vertebrate motif set. This is a **per-site, base-resolution** view: it reads the shape of the Tn5 cut profile to detect bases protected by a bound factor, then BINDetect quantifies, per motif, how much that footprint signal **differs between conditions**. It answers "is this factor actively bound, and does its binding change?" — and is complemented by the set-level chromVAR activity in the next section.

How to read this figure: Each bar is one TF motif. Bar length is the differential footprint score; a positive score means the TF's binding sites are more protected (more bound) in the first condition. TFs at the extremes are the candidate drivers of the accessibility differences seen above. The full ranked table ships in `results_workbook.xlsx`.

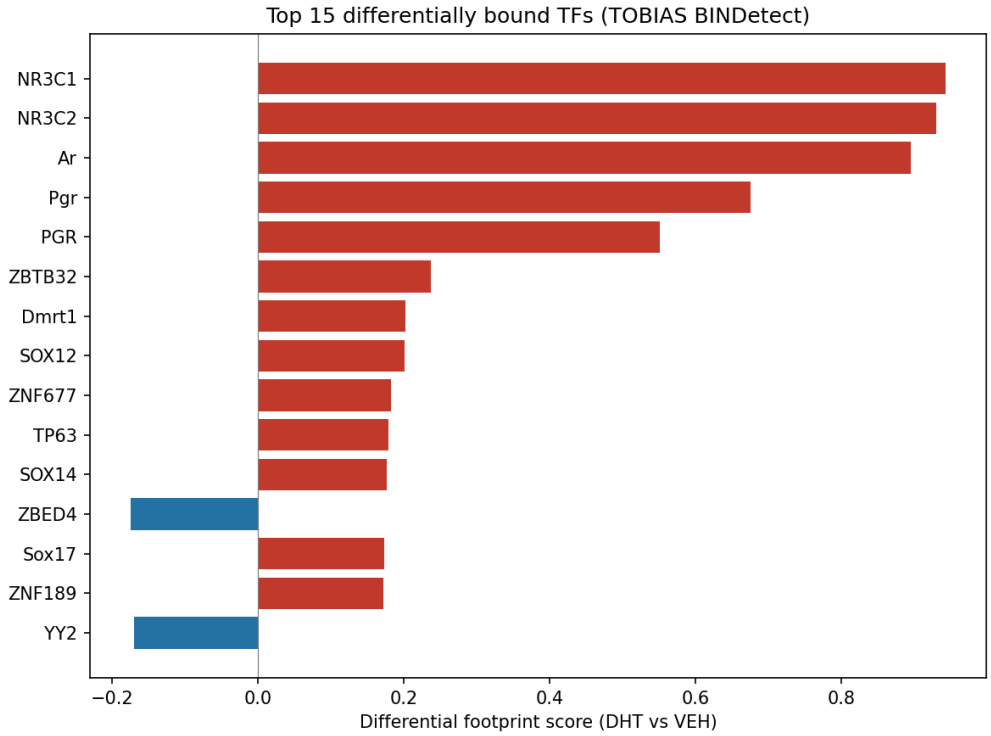

Figure 9: Top transcription factors ranked by differential footprint score (TOBIAS BINDetect). Each bar is one motif; bar length is the change in footprint (binding) strength between conditions, with positive values (and colour) indicating stronger footprints in the first condition.

What this means for your data: TOBIAS BINDetect compared TF footprints for DHT vs VEH; the top 15 TFs by absolute differential binding score are shown (averaged over 4 cross-condition replicate pairs); 813 TFs at $p < 0.05$. Positive scores indicate stronger footprints in DHT.

9 TF Motif Accessibility (chromVAR)

Where TOBIAS (previous section) works **per-site at base resolution** and asks whether a factor's footprint changes between conditions, **chromVAR** works at the **set level**: for each TF motif it aggregates accessibility across *all* peaks containing that motif into one bias-corrected deviation Z-score per sample. It answers "is this factor's motif programme globally more open in some samples?" The two views are complementary — footprinting localises binding, chromVAR summarises programme-level activity.

How to read this figure: Rows are TF motifs (the most variable across samples), columns are samples. Colour is the bias-corrected deviation Z-score: red = more accessible than expected, blue = less. TFs whose rows split cleanly by condition are the motif programmes that track the biology. The full deviation and variability tables ship in `results_workbook.xlsx`.

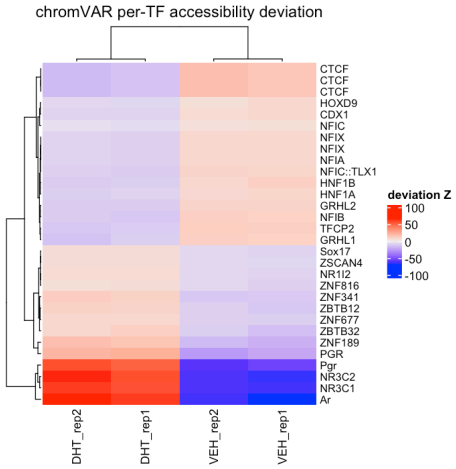

Figure 10: Heatmap of chromVAR transcription-factor accessibility deviations. Rows are the most variable TF motifs and columns are samples (both hierarchically clustered); colour encodes the bias-corrected accessibility-deviation Z-score (red = more accessible than expected, blue = less).

What this means for your data: chromVAR scored 879 TF motifs across 4 samples; the 30 most variable motifs are shown. Rows that split by condition mark the TF programmes tracking the biology.

10 Pathway Enrichment

Genes with promoter-proximal accessible peaks were tested for over-representation in GO:BP, KEGG, and Reactome.

How to read these plots: Each dot / bar is a biological pathway — a curated set of genes that act together (e.g. "Wnt signalling" or "axon guidance"). The horizontal axis is the enrichment statistic: bigger values mean a larger fraction of the pathway's genes overlap the open-chromatin gene list than expected by chance. Dot **size** is the number of overlap genes (bigger = more robust). We screen three databases — GO:BP (broad biology), KEGG (canonical signalling/metabolism) and Reactome (curated, expert-vetted pathways) — because each catches different things.

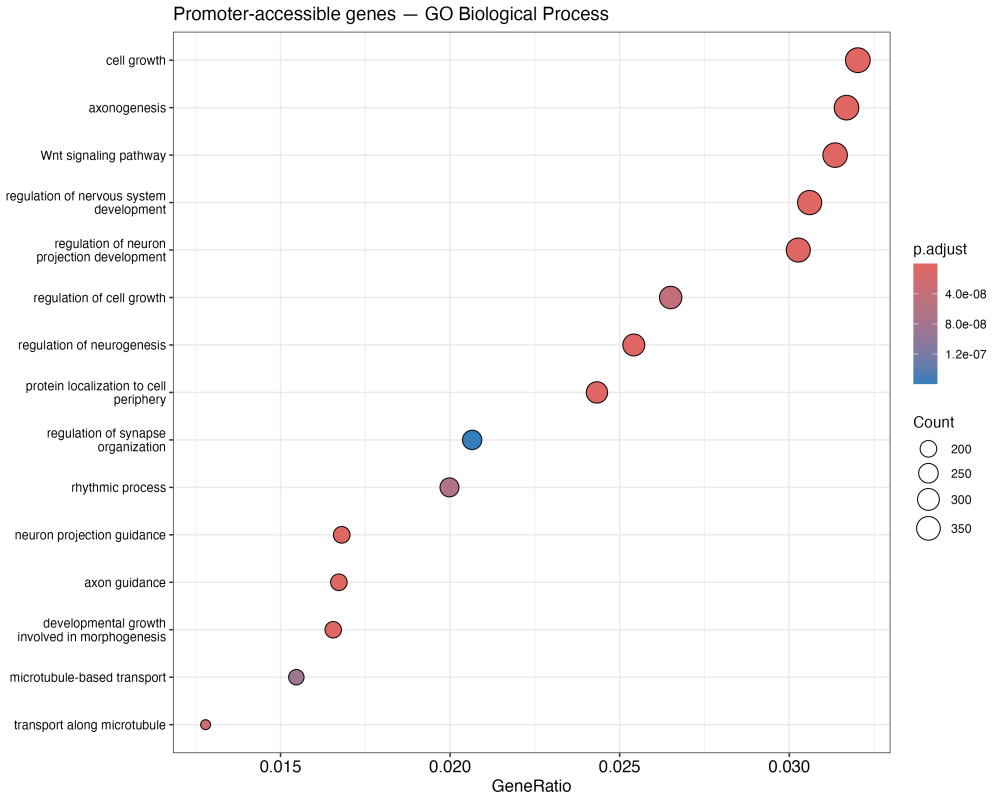

Figure 11: Dot plot of the most enriched pathways among genes with promoter-proximal accessible peaks (clusterProfiler over-representation test). The x-axis is the gene ratio (fraction of the pathway's genes recovered); dot size is the number of overlapping genes and dot colour is the BH-adjusted p-value.

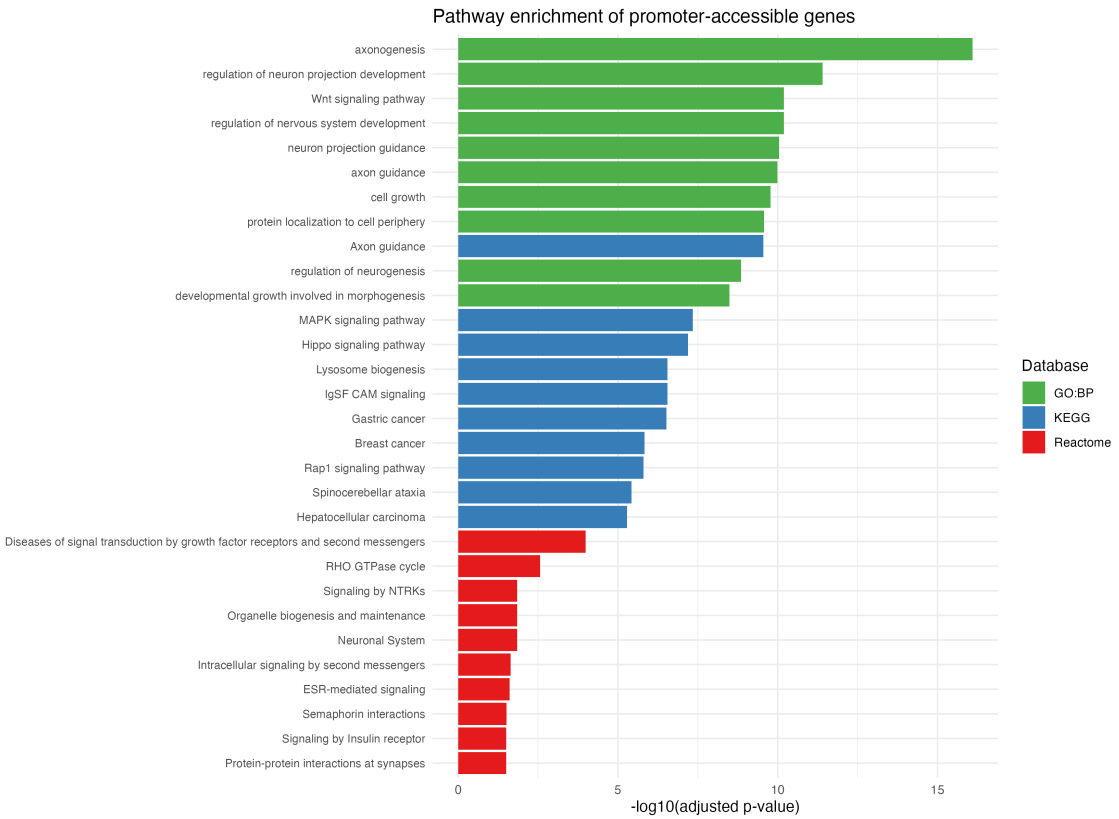

Figure 12: Bar plot of the top enriched terms across GO:BP, KEGG and Reactome. Bar length is enrichment significance (−log10 BH-adjusted p-value); terms are grouped by source database.

What this means for your data: Top pathways enriched in 16598 promoter-accessible genes: axonogenesis; regulation of neuron projection development; regulation of nervous system development.

11 Conclusions

Three independent views of the data converge on a single conclusion: a single dose of DHT switches on the androgen receptor regulatory programme in VCaP cells, and that programme is written directly into the open chromatin. The differential accessibility result is strongly asymmetric, with more than twenty times as many regions gained as lost in DHT, which is what a ligand activated nuclear receptor does when it opens its target enhancers. chromVAR then names the receptor: of 879 motif programmes, AR is the most variable across the experiment, and its activity flips cleanly from closed in vehicle to open in DHT. TOBIAS footprinting, working at base resolution on the Tn5 cut profile rather than on motif sets, independently ranks the androgen and steroid receptor family at the top of differentially bound factors. One interpretive caveat is worth stating plainly. The androgen receptor (AR) shares an almost identical DNA response element with the glucocorticoid, mineralocorticoid and progesterone receptors (NR3C1, NR3C2 and PGR). Motif and footprint methods read the DNA sequence under the open region, so these receptors are not fully separable and tend to co rank whenever the steroid receptor programme is active. In VCaP, an androgen driven prostate cancer line in which AR is the dominant and biologically relevant receptor, this family level signal is the AR programme; the co ranking of NR3C1 and NR3C2 reflects shared motif grammar, not independent glucocorticoid or mineralocorticoid activity. The recovery of FOXA1 and HOXB13, the pioneer and co factors that license AR binding specifically in prostate tissue, supports this reading. For follow up, the regions that both gained accessibility and carry a strengthened AR footprint in DHT, listed in the accompanying workbook, are the high confidence candidate androgen responsive regulatory elements. These are the natural starting point for linking the chromatin response to target gene expression or for designing perturbation experiments.

---

12 Appendix — Top Peak Annotations

How to read this table: Each row is one accessible peak. **seqnames / start / end** locate the peak on the genome; **annotation_simple** tells you what genomic feature it overlaps (promoter, intron, distal intergenic, …); **SYMBOL** is the nearest annotated gene; **distanceToTSS** is the distance in base pairs from the peak centre to that gene's TSS (positive = downstream, negative = upstream). Use this leaderboard to spot the peak-by-peak hits worth eyeballing first; the full ranked annotation table ships in `results_workbook.xlsx`.

> **Full peak annotation:** the complete table (all peaks, all annotation columns) is delivered in `results_workbook.xlsx`, sheet *Peak annotation*. Only the curated top-25 is shown above to keep the PDF readable.

Report generated by OmicsDesk Sample Report project — 2026-06-09 12:16:31.897604 Pipeline: fastp + Bowtie2 + MACS3 + ChIPseeker + deepTools + DiffBind + chromVAR + TOBIAS Deliverables: PDF report + Excel workbook (`results_workbook.xlsx`) + MultiQC HTML For questions contact: hello@omicsdesk.com

Table 4: Top 25 genes by peak signal (one row per gene, the strongest associated peak shown). The complete per-peak annotation table for all samples is in 'results_workbook.xlsx', sheet 'Peak annotation'.

seqnames	start	end	annotation_simple	SYMBOL	distanceToTSS	sample
chr9	39791249	39791733	Intron	GLIDR	18395	DHT_rep1
chr2	49973930	49974445	5' UTR	NRXN1	0	DHT_rep1
chr3	104206986	104207353	Intergenic (distal)	MIR548AB	-682870	DHT_rep1
chr2	58080980	58081366	Intron	VRK2	0	DHT_rep1
chr3	20081172	20081553	Intron	KAT2B	40726	DHT_rep2
chr5	41193663	41194193	Intron	C6	67245	DHT_rep2
chr9	66900565	66901117	5' UTR	ZNF658	44139	DHT_rep2
chr18	65876312	65876652	Intron	CDH7	126060	DHT_rep2
chr3	27369147	27369665	Promoter	NEK10	0	DHT_rep1
chr3	158570951	158571425	Promoter	MLF1-DT	0	DHT_rep2
chr6	103236871	103237335	Intergenic (distal)	GRIK2	1155248	DHT_rep1
chr5	82396885	82397193	Intergenic (distal)	ATP6AP1L	117423	DHT_rep1
chr8	47167256	47167621	Intron	SPIDR	93256	VEH_rep1
chr9	32136740	32137179	Intergenic (distal)	ACO1	-247424	DHT_rep2
chr5	110726491	110727359	5' UTR	SLC25A46	-10777	VEH_rep2
chr13	88948447	88948915	Intergenic (distal)	LINC00440	-232409	DHT_rep2
chr6	95592965	95593508	Intron	MANEA	15480	DHT_rep1
chr8	93754688	93755200	Promoter	TMEM67	0	VEH_rep2
chr13	52252817	52253067	Intron	MRPS31P5	38490	DHT_rep1
chr12	45655301	45655489	Intergenic (distal)	LINC00938	62556	DHT_rep1
chr8	50115465	50116030	Intron	SNTG1	205676	DHT_rep2
chr7	143836514	143837475	Exon	TCAF1	64723	DHT_rep1
chr21	13980094	13980553	Promoter	ANKRD20A11P	0	DHT_rep2
chr2	209368979	209369231	Intron	MAP2	-54816	DHT_rep2
chr8	91069972	91070647	Promoter	OTUD6B	0	VEH_rep1