Federal Aviation Administration

DOT/FAA/AM-24/02
Office of Aerospace Medicine
Washington, DC 20591

Gene Expression and Biomarker Utility in Postmortem Samples

Christopher J. Tracy
David C. Hutchings
Susan K. Munster
Vicky L. White
Scott J. Nicholson

Civil Aerospace Medical Institute
Federal Aviation Administration
6500 S. MacArthur Blvd
Oklahoma City, OK 73169

April 2024

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications website (https://www.faa.gov/go/oamtechreports) National Transportation Library's Repository & Open Science Access Portal (https://rosap.ntl.bts.gov/)

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-24/02	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Gene Expression and Biomarker Utility in Postmortem Samples		5. Report Date April 2024 6. Performing Organization Code AAM-612
7. Author(s) Christopher J. Tracy (https://orcid.org/0000-0002-4532-3327) David C. Hutchings (https://orcid.org/0009-0008-4575-8444) Susan K. Munster (https://orcid.org/0000-0002-3223-0076) Vicky L. White (https://orcid.org/0009-0007-1164-5233) Scott J. Nicholson (https://orcid.org/0000-0002-2201-744X)		8. Performing Organization Report No. DOT/FAA/AM-24/02
9. Performing Organization Name and Address Civil Aerospace Medical Institute Federal Aviation Administration 6500 S. MacArthur Blvd Oklahoma City, OK 73169		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered Technical Report 14. Sponsoring Agency Code
15. Supplementary Notes Technical report DOI: https://doi.org/10.21949/1529631		

16. Abstract

This study assesses the suitability of aviation accident autopsy samples as RNA sequencing input and for detecting gene expression differences between THC-positive and THC-negative samples. Postmortem brain, lung, muscle, and blood samples were collected from 57 aviation accident fatalities and comparisons were made between each tissue based on the presence or absence of THC or its primary or secondary metabolites (28 positive, 29 negative). RNA was extracted, and global transcriptional analysis was performed using total RNA-Seq. Twenty-two genes in lung (18 of which had Entrez annotation data) and four in muscle showed significant differential gene expression between THC-positive and THC-negative samples. It is possible that the observed expression patterns between the THC-positive and THC-negative groups were induced by smoking and not by THC, as many of the observed genes were also reported in the literature to change in response to smoking. Cannabis use is often accompanied by other substance use, and evidence of co-use was observed among some subjects. Therefore, we cannot state conclusively that the observed differences between the THC-positive and THC-negative groups were due solely to THC consumption, only that significant differences exist when the subject groups are so segregated. Regardless, this study is the first of its kind in reporting RNA-Seq data from postmortem tissues collected from aviation accident victims and that said data are of sufficient quality to derive significant differences between subject groups.

17. Key Words Gene expression, biomarkers, aviation accident investigation, RNA, cannabis, delta-9-THC	18. Distribution Statement Document is available to the public through the National Transportation Library: https://rosap.ntl.bts.gov/	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 33

Acknowledgements

We would like to thank Douglas Caldwell, Jensen Smillie, and Roxane Ritter of the FAA/CAMI Quality Assurance team and Sunday Hickerson of the FAA/CAMI Forensic Toxicology team for their assistance in collecting and identifying samples to be used in this study.

Table of Contents

Tables and Figures

List of Abbreviations

CAMI	Civil Aerospace Medical Institute
DEG	Differentially Expressed Gene
FAA	Federal Aviation Administration
FDR	False Discovery Rate
HGSC	Baylor College of Medicine Human Genome Sequencing Center
NTSB	National Transportation Safety Board
PCA	Principal Component Analysis
GEN	FAA/CAMI Functional Genomics Team
PMI	Postmortem Interval
QA	FAA/CAMI Quality Assurance Team
qPCR	Quantitative Reverse-Transcriptase Polymerase Chain Reaction
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA-Seq	RNA-Sequencing
Delta-9-THC	Delta-9-Tetrahydrocannabinol
TOX	FAA/CAMI Forensic Toxicology Team

Abstract

This study assesses the suitability of aviation accident autopsy samples as RNA sequencing input and for detecting gene expression differences between THC-positive and THC-negative samples. Postmortem brain, lung, muscle, and blood samples were collected from 57 aviation accident fatalities and comparisons were made between each tissue based on the presence or absence of THC or its primary or secondary metabolites (28 positive, 29 negative). RNA was extracted, and global transcriptional analysis was performed using total RNA-Seq. Twenty-two genes in lung (18 of which had Entrez annotation data) and four in muscle showed significant differential gene expression between THC-positive and THC-negative samples. It is possible that the observed expression patterns between the THC-positive and THC-negative groups were induced by smoking and not by THC, as many of the observed genes were also reported in the literature to change in response to smoking. Cannabis use is often accompanied by other substance use, and evidence of co-use was observed among some subjects. Therefore we cannot state conclusively that the observed differences between the THC-positive and THC-negative groups were due solely to THC consumption, only that significant differences exist when the subject groups are so segregated. Regardless, this study is the first of its kind in reporting RNA-Seq data from postmortem tissues collected from aviation accident victims and that said data are of sufficient quality to derive significant differences between subject groups.

1. Introduction

The National Transportation Safety Board (NTSB) is tasked by 49 U.S.C § 1111(g) (2021) to "investigate and report on accidents involving...aviation," the United States Department of Transportation order 1100.1C (3)(C)(2)(c)(2021) directs the Federal Aviation Administration's (FAA) Office of Aviation Safety to "investigate […] aircraft accidents and incidents [and] support […] NTSB accident and incident investigations," and 49 U.S.C. § 44507(a)(1-5) (2021) authorizes the FAA Civil Aerospace Medical Institute (CAMI) to "conduct civil aeromedical research." Under these directives, the FAA/CAMI Functional Genomics (GEN) team performs research to identify molecular biomarkers for use as novel tools for aviation accident investigation. In contrast, the FAA/CAMI Forensic Toxicology (TOX) team performs toxicological analysis on specimens collected postmortem from aviation accident victims to contribute to the determination of accident causation. Of particular interest is the use of delta-9-tetrahydrocannabinol (delta-9-THC), the main psychoactive compound found in cannabis, by pilots, the use of which "make[s] [a pilot] unqualified to hold an FAA-issued medical certificate" (FAA, 2016).

The 2020 National Survey on Drug Use and Health reports that, for individuals aged 12 or older, past-year cannabis use increased from 11.0% in 2002 to 17.5% in 2019 (Substance Abuse and Mental Health Services Administration, 2021). For individuals aged 18 to 25, past-year cannabis use increased from 29.8% in 2002 to 35.4% in 2019, and for individuals aged 26 or older, past-year cannabis use increased from 7.0% to 15.2% within the same time period. An analysis of the FAA toxicological accident records database showed that, for the 10-year period from 2007 to 2016, 3.4% of analyzed aviation accident fatalities tested positive for either THC or its main secondary metabolite in at least one tissue or bodily fluid (such as urine, vitreous humor, blood, etc.) (Norris et al., 2018). Furthermore, this number was consistent with the previous reporting period from 1996 to 2007. The impact of THC consumption on transportation is not limited solely to aviation; the National Roadside Surveys, conducted by the National Highway Transportation Safety Administration for 2007 (the first year data was collected on THC prevalence) and 2013-2014 (the final year of data) showed an increase in surveyed drivers testing positive for THC, rising from 8.6% in 2007 to 12.6% in 2013-2014 (Compton, 2017).

Genes that change in expression following THC consumption may be useful as molecular biomarkers, and if correlated with measures of THC-related cognitive impairment, may assist in the determination of aviation accident causation. To this end, the GEN team has previously studied ribonucleic acid (RNA) extracted from postmortem blood and tissue samples collected from NTSB-investigated general aviation fatalities at the time of autopsy (Burian et al., 2017). That study found RNA extracted from aviation accident victims to be highly variable and total RNA yields for some samples to be much greater than expected, suggesting microbial contamination. RNA Integrity Number (RIN) values, a 1-10 scale measure of RNA sample quality, were significantly lower in samples collected from victims who experienced moderate to high levels of trauma versus low-trauma victims. Additionally, the presence of prokaryotic rRNA

peaks on the RIN traces was positively correlated with postmortem interval (PMI), while eukaryotic rRNA were negatively correlated with PMI, suggesting the incidence of microbial contamination has a direct relationship with the amount of lag between time of death and time of autopsy. Relative gene expression patterns were determined using a panel of 35 genes for available tissues and blood from postmortem sources and blood from living individuals; significant correlations between living and postmortem sources were seen in brain, lung, skeletal muscle, and blood, suggesting that postmortem expression patterns in those tissues are most similar to those of live tissue. However, other studies in animal models of brain and in human blood demonstrate that postmortem changes in gene expression, on a gene-by-gene and transcriptome-wide basis using quantitative reverse-transcriptase polymerase chain reaction (qPCR) and RNA sequencing (RNA-Seq), can occur following death (Halawa et al., 2021, Antiga et al., 2021). While gene expression can be measured on autopsy-derived and postmortem specimens using qPCR methods, the usefulness of such specimens for high-throughput RNA-seq based gene expression analysis was unclear.

This study was conducted to determine if aviation accident autopsy samples can be used reliably for RNA-seq and if differences in gene expression are detectable between samples grouped by the presence of absence of THC. Each of those questions were answered affirmatively, although the gene expression differences noted may have been attributable to smoking and not to THC itself. This work also describes methods and procedures suitable for conducting gene expression analysis on low-quality or partially degraded specimens, what results might be expected from those analyses, and may provide guidance to those considering such efforts.

2. Methods

Institutional Review Board and Waver of Consent

All research was conducted with the approval of the FAA Institutional Review Board. A waiver of consent exists for specimens collected from fatal aviation accident victims, as a human subject, as defined by 49 C.F.R. §11.102(e)(1)(i), is “a living individual about whom an investigator conducting research obtains [...] biospecimens through [...] interaction with the individual, and [...] analyzes [...] the biospecimens.” As research samples for this study were only collected from deceased individuals, individual or familial consent was not required for sample collection. All personally identifiable information was removed from these samples, and RNA-Seq data are deposited in the limited-access Database of Genotypes and Phenotypes.

Sample Collection

The TOX team assessed forensic tissue samples for the presence of delta-9-THC or its primary or secondary metabolites, 11-hydroxy-delta-9-THC and 11-nor-9-carboxy-delta-9-THC. Samples that tested positive for any of these compounds were categorized as “THC-positive” and THC-negative control samples were then flagged for sampling and inclusion in this study by the FAA/CAMI Quality Assurance (QA) team. Tissue samples were allowed to thaw at room

temperature for approximately 1 hour before tissue sampling by GEN team personnel. Brain, lung, and muscle tissues, including psoas and other muscles collected during autopsies, were obtained using punches measuring 3 mm and 6 mm in diameter. The selection of brain regions was based on the discretion of individual medical examiners, and tissue collection from each subject was contingent upon availability. Samples were collected using Biopunches (Ted Pella, Inc., 15111-30, 15111-60), and placed into tubes containing RNA*later* Stabilization Solution (ThermoFisher Scientific, AM7021). Blood was originally collected at the time of autopsy in Vacutainer Sodium Fluoride/Potassium Oxalate 100 mg/20 mg tubes (BD Biosciences, 367001) for toxicological analysis. When available, two aliquots of approximately 1 mL of this preserved blood were collected from each subject. Blood samples were then stored in a -80°C freezer.

RNA Extraction

Tissue samples were homogenized using a TissueLyser II bead mill (QIAGEN, 85300) using the Adapter Set 2 x 24 and 5-mm stainless steel beads (QIAGEN, 69989) in QIAzol Lysis Reagent (brain) or Buffer RLT (lung and muscle). RNA was isolated from the homogenized tissue samples using either the RNeasy Lipid Tissue Mini kit (QIAGEN, 74804) for brain samples or the RNeasy Fibrous Tissue Mini kit (QIAGEN, 74704) for lung and muscle samples. On-column DNase I treatment was performed for all tissue extractions using the RNase-Free DNase Set (QIAGEN, 79254). RNA was isolated from whole blood samples using the Quick-RNA Whole Blood kit and treated with on-column DNase I (Zymo Research, R1201). RNA was analyzed for purity and yield using a NanoDrop 2000c spectrophotometer (ThermoFisher Scientific, ND-2000c) and a Qubit 3.0 fluorometer (ThermoFisher Scientific, Q33216) using the Broad Range Assay kit (ThermoFisher Scientific, Q10210). RINs were calculated using a 4200 TapeStation system (Agilent, G2991BA) and RNA ScreenTape tapes and reagents (Agilent, 5067-5576, 5067-5578, 5067-5577).

RNA-Seq cDNA Synthesis, Library Preparation, and Sequencing

Samples with sufficient concentration for sequencing (in general, at least 50-80 ng/µL when possible, although several blood samples had concentrations in the 30-40 ng/µL range, and one blood sample had a concentration of 13.8 ng/µL), were diluted in RNase-free water: if the sample concentration was >80 ng/µL, it was diluted to 80 ng/µL; if the sample concentration was <80 ng/µL but >50 ng/µL, it was diluted to 50 ng/µL; if the sample concentration was <50 ng/µL, it was not diluted. Samples then were shipped on dry ice to the Baylor College of Medicine Human Genome Sequencing Center (HGSC) for library preparation and sequencing. At the HGSC, RNA quality was assessed using a Fragment Analyzer 5300 (Agilent, M5311AA) and RNA Kit (15NT) reagents (Agilent, DNA-471-0500). Samples were randomly re-arrayed to minimize potential batch effects and spiked with ERCC synthetic RNA (ThermoFisher Scientific, 4456740). Libraries were prepared using the TruSeq Stranded Total RNA with Ribo-Zero Globin kit (Illumina Inc., 20020612) following the manufacturer's protocol (Illumina Inc., RS-122-9007DOC, Part # 15031048 Rev. E, October 2013). Ribosomal RNA/Globin-depleted RNA samples were purified using Agencourt RNAClean XP beads (Beckman Coulter, A63987)

and then used for first and second-strand cDNA synthesis. The cDNA was A-tailed and ligated with the TruSeq UD Indexes V2 (Illumina Inc., 20042113), PCR-amplified using the Illumina Primer Cocktail Mix, and then purified using AMPure XP beads (Beckman Coulter, A63882). Libraries were quantified using the Fragment Analyzer 5300, pooled in equimolar ratios, and then the pools were qPCR quantified. Libraries were sequenced on the Illumina NovaSeq 6000 platform using 2x150 bp paired-end reads to generate a target of 100 million reads per sample.

RNA-Seq Analysis Pipeline

Full alignment and quality control pipeline code, including individual software parameters and flags, is described in the Supplementary Material (supplemental_file_1). The pipeline was performed on a workstation running Red Hat Enterprise Linux 8. After each step of the pipeline, MultiQC (v1.14; Ewels et al., 2016) was used to generate html summary reports to collate the results generated by each program. Quality assessment of the raw gzipped fastq files was performed using FASTQC (v0.12.1; Andrews, 2010). The raw reads were trimmed of adapters and quality filtered using CutAdapt (v4.3; Martin, 2011). Quality assessment via FASTQC of the trimmed and filtered reads was performed a second time to assess the trimming and filtering. The trimmed and filtered reads were aligned to the ENCODE human reference genome primary assembly (GRCh38.p13, Release 43, indexed with a 149 bp sjdb overhang) and primary assembly comprehensive gene annotation (in GTF format), using STAR (v2.7.10.b; Dobin et al., 2013), generating Aligned.sortedByCoord.out.bam files and paired Unmapped.out.mate files consisting of all unmapped reads. BAM index files were generated using samtools (v1.17; Li et al., 2009). Post-alignment quality control on the Aligned.sortedByCoord.out.bam files was performed using Qualimap (v2.2.2d; García-Alcalde et al., 2012). Feature count matrices were generated using the featureCounts function of SubRead (v2.0.4; Liao et al., 2014). GNU Parallel was used to combine individual feature count matrices into a single matrix (Tange, 2018). Microbial contaminant investigation of the unmapped reads was performed using a pipeline consisting of Kraken (v1.1.1; Wood & Salzberg, 2014), Kraken2 (v2.1.2; Wood et al., 2019), Bracken (v2.8; Lu et al., 2017), and Krona (v2.8.1; Ondov et al., 2011) using the PlusPF index collection (retrieved 2023-03-21). Kraken2 output files were collated by species using the kraken-multiple-taxa.py script (Papudeshi, 2022).

Differential Gene Expression Analysis

Full analysis code is described in the Supplementary Material (supplemental_file_2). Differential gene expression analysis was performed using R Statistical Software (v4.3.0; R Core Team, 2023) in conjunction with RStudio (build 446; Posit Team, 2023). The analysis code is adapted from, in part, training material from the Harvard Chan Bioinformatics Core Differential Gene Expression workshop (Mistry et al., 2021). The following packages in R were called directly to perform these analyses and prepare figures and tables for publication: DESeq2 (v.1.40.1; Love et al., 2014), tidyverse (v2.0.0; Wickham et al., 2019), GeneStructureTools (v1.20.0; Signal, 2023), limma (v3.56.1; Ritchie et al., 2015), ggplot2 (v3.4.2; Wickham, 2016), pheatmap (v1.0.12; Kolde, 2019); apeglm (v1.22.1; Zhu et al., 2019), ggrepel (v0.9.3;

Slowikowski, 2023), DEGreport (v1.36.0; Pantano, 2023), AnnotationHub (v3.8.0; Morgan & Shepherd, 2023), ensembldb (v2.21.0; Rainer et al., 2019), annotables (v0.2.0; Turner, 2023), pals (v1.7; Wright, 2021), patchwork (v1.1.2; Pedersen, 2022), ggpubr (v0.6.0; Kassambara, 2023), scales (v1.2.1; Wickham & Seidel, 2022), ggiraphExtra (v0.3.0; Moon, 2020), flextable (v0.9.2; Gohel & Skintzos, 2023a), ggiraph (v0.8.7; Gohel & Skintzos, 2023b), officer (v0.6.2; Gohel, 2023), and magrittr (v2.0.3; Bache & Wickham, 2022).

The dataset was sorted by tissue and then each tissue subset was analyzed using the DESeq2 package with the design parameter "design=~ THC". Log fold change shrinkage was performed using the apeglm package and the coefficient "THC_positive_vs_negative." A false discovery rate (FDR)-adjusted *P*-value significance threshold of 0.1 and a $\log_2$ fold change threshold of 0.58, corresponding to a fold change of approximately 1.5, were used as filters. Significant differentially expressed genes were annotated with Entrez ID information.

Microbial Contaminant Analysis

Full analysis code is described in the Supplementary Material (supplemental_file_9). Analysis was performed using R Statistical Software in conjunction with RStudio. To examine the extent of microbial contamination present within the postmortem samples used in this study, microbial contaminant analysis was performed on the paired Unmapped.out.mate files generated as part of the RNA sequencing alignment pipeline (i.e., reads that were not mapped to the human genome index during read alignment). Sequence reads homologous to *Homo sapiens* were filtered out (i.e., reads that were otherwise excluded during alignment for mapping to too many loci, for being too short, or for other mapping-related reasons), then summed across the remaining species for each sample and then across each sample for each tissue. The total number of reads (microbial contamination plus uniquely mapped reads) for each tissue was calculated and used to determine the percent microbial contamination for each of the four tissues. The following packages in R were called directly to perform this analysis and prepare figures for publication: tidyverse, ggplot2, scales, pals, ggrepel, and patchwork.

3. Results

Aviation Accident Postmortem Samples

RNA was extracted from postmortem brain, lung, muscle, and blood samples from 57 fatal general aviation accident victims, depending on tissue and blood availability. Not every tissue was available/collected from every subject at the time of autopsy, and in some cases, individual samples were so degraded that it was impossible to isolate RNA. Therefore, the actual number of samples extracted differs from tissue to tissue: 53 brain samples (28 THC-negative and 25 THC-positive), 52 lung samples (27 THC-negative and 25 THC-positive), 53 muscle samples (26 THC-negative and 27 THC-positive), and 43 blood samples (22 THC-negative and 21 THC-positive).

Following RNA extraction, RNA quality and quantity was assessed (Table 1). Mean sample concentration was highly variable, with brain samples having the highest RNA

concentrations, followed by muscle, blood, and lung (Fig. 1A). The first standard deviation values for mean concentration for all samples were quite large, but especially for the blood samples (Table 1 “Concentration” and Fig. 1A). Mean RIN values varied across tissues, but uniformly indicated moderate to high RNA degradation. Blood samples demonstrated the lowest values, indicating the most degradation and muscle the highest, and each sample type displayed a high degree of inter-subject variability (Fig. 1B). The mean 260/280 nm and 260/230 nm absorbance ratios, measures of nucleic acid purity and level of organic compound contamination respectively, were approximately 2.0 and between 2.0 and 2.2, respectively, indicating that RNA extraction methods used in this study yielded pure RNA without significant contamination by organic compounds (Fig. 1C and 1D). In most comparisons, blood-extracted RNA demonstrated the lowest and most variable metric in each measurement. While the RNA extracted from these samples was more degraded than is generally desirable for RNA-Seq, sample concentrations and purity were within the range expected of spectrophotometrically pure RNA, and samples were used for RNA-Seq.

Table 1. RNA extraction metrics from postmortem tissue samples. Mean values ± 1 standard deviation for the RNA extraction metrics sample concentration (ng/μL), RIN value, 260/280 nm absorbance ratio, and 260/230 nm absorbance ratio for blood (N=43), brain (N=53), lung (N=52), and muscle (N=53) samples.

Note. RIN = RNA Integrity Number.

Tissue	Concentration	RIN	A260/280 Ratio	A260/230 Ratio
Brain	506.8±457.3	4.2±1.0	2.07±0.03	2.16±0.14
Lung	164.4±154.6	3.9±1.4	2.09±0.03	2.13±0.12
Muscle	319.5±178.8	5.0±1.5	2.08±0.02	2.13±0.10
Blood	242.3±346.6	3.1±1.2	1.99±0.09	2.10±0.22

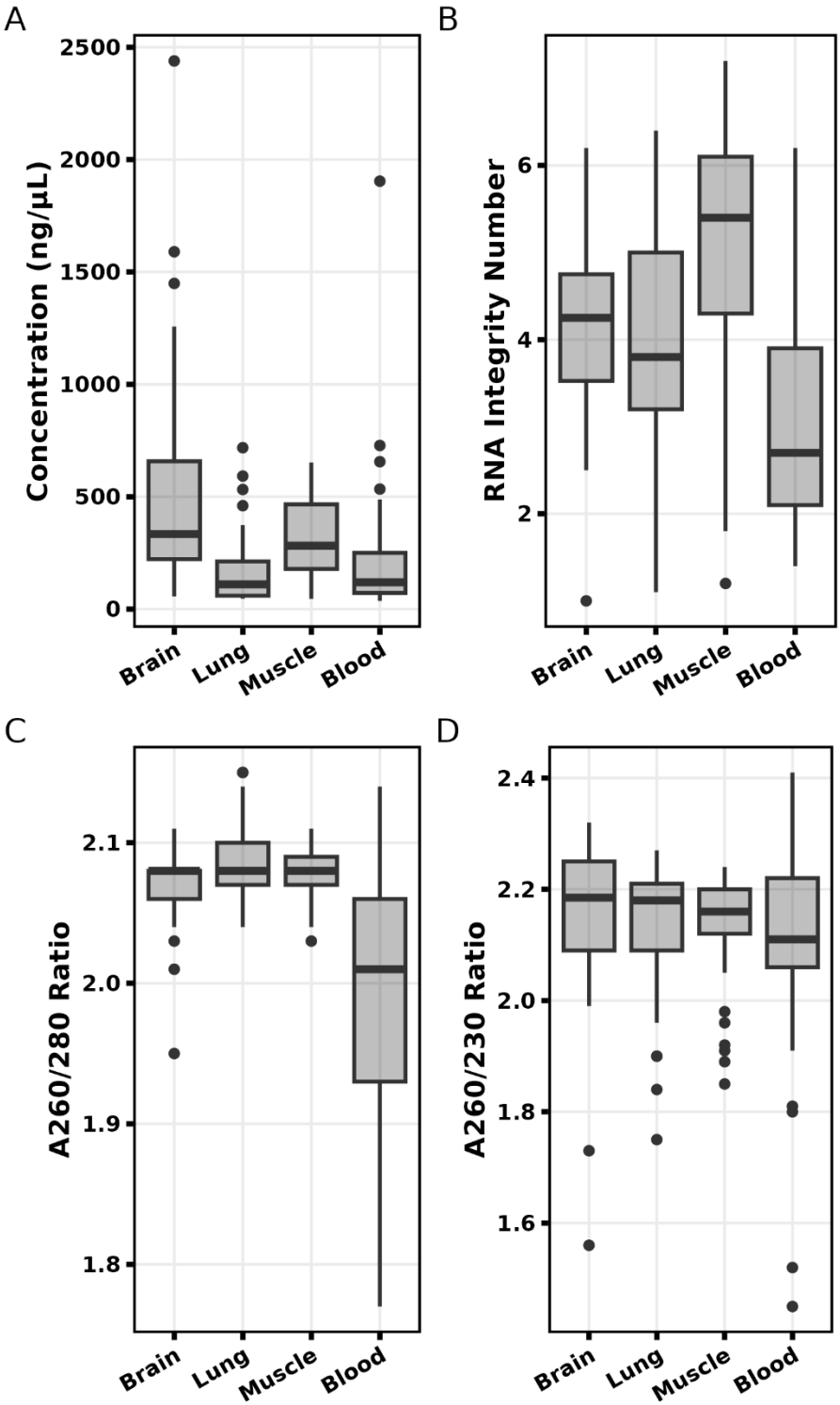

Figure 1. RNA extraction metrics. Boxplots of extracted RNA sample concentration (A), RIN values (B), 260/280 nm absorbance ratio (C), and 260/230 nm absorbance ratio (D) for each of the four tissues.

Note. RIN = RNA Integrity Number.

Differential Gene Expression Analysis

Total RNA sequencing data (2x150 paired-end) were generated from the RNA extracted from each sample. Principal component analysis (PCA) performed on rlog transformed counts within sample types indicated separation largely by a single component for brain, muscle, and blood, with no clustering based on THC presence or absence. The brain and lung population contained a single THC-positive sample that strongly segregates along PC2, while the remaining THC-positive and THC-negative samples cluster across PC1 (Fig. 2A and 2B). Similarly, muscle and blood samples do not show segregation on the basis of THC status, with samples clustering along PC1 and PC2 irrespective of THC (Fig. 2C and 2D).

Differential gene expression analysis was performed to determine if significant gene expression changes were associated with the presence of THC. The comparison regarded THC-negative samples as negative values; therefore, positive log2FoldChange measures indicate higher expression in THC-positive samples (upregulation), and negative values signify lower expression in THC-positive samples (downregulation). No significant differentially expressed genes (DEGs) were observed in brain or blood samples for the THC-positive vs THC-negative comparison (Fig. 3A and 3D). A total of 22 DEGs were observed in lung samples, with 20 genes showing upregulation and two genes showing downregulation (Fig. 3B). In muscle, four DEGs were observed, with three genes showing upregulation and one gene showing downregulation (Fig. 3C). These 22 and four significant DEGs were plotted by $\log_{10}$ normalized counts (Fig. 4), demonstrating imperfect separation of gene expression measurements between THC-positive and THC-negative samples even in differentially expressed genes. The DEGs were also annotated by Entrez gene ID, reducing the number of DEGs with available annotations to 18 for lung (Table 2) and remaining at four for muscle (Table 3).

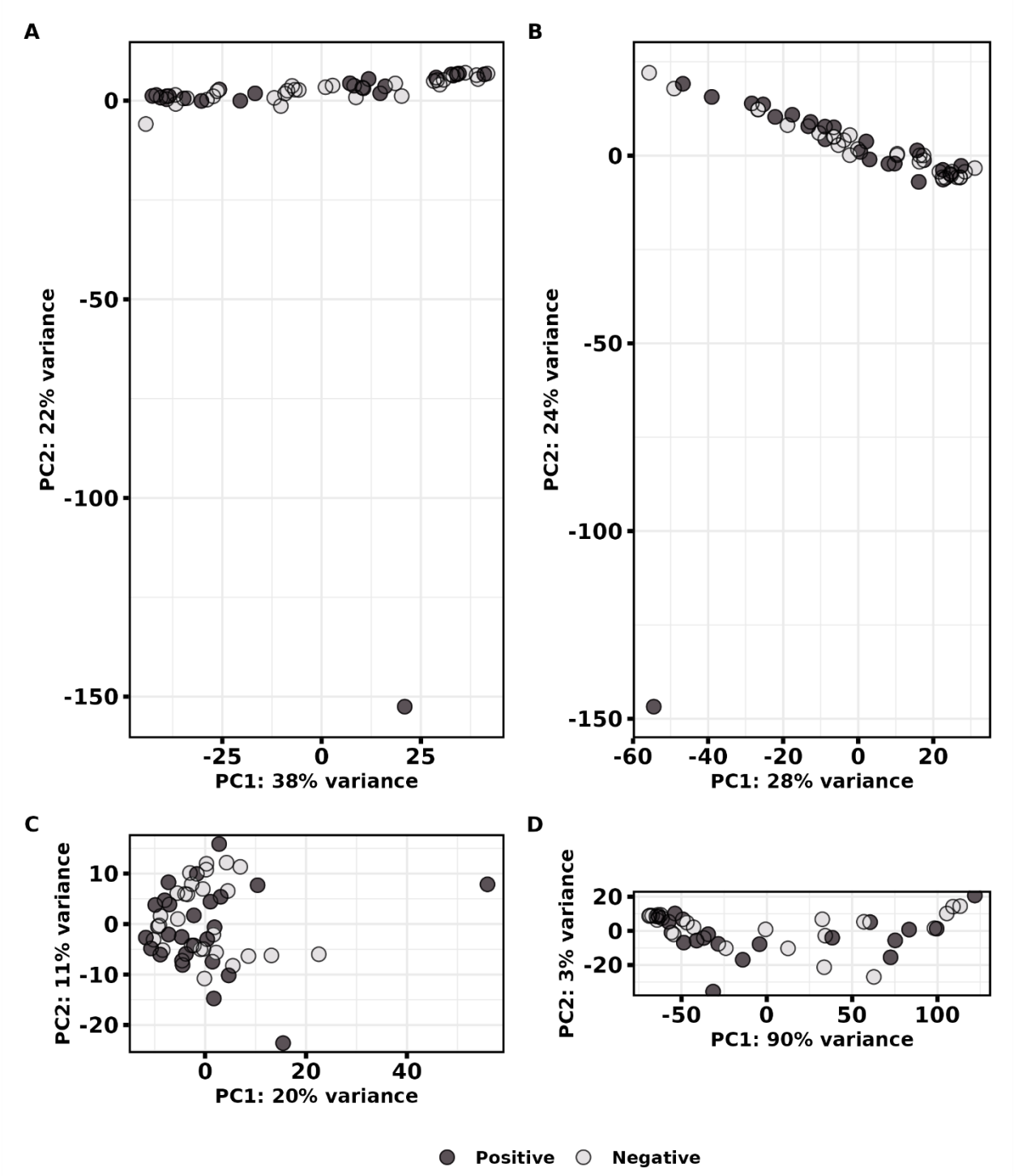

Figure 2. Principal component analysis. PCA was performed using DESeq2's plotPCA function on normalized, regularized log transformed counts. Samples do not noticeably cluster by THC status for any of the four tissues: brain (A), lung (B), muscle (C), or blood (D).

Note. THC = Tetrahydrocannabinol.

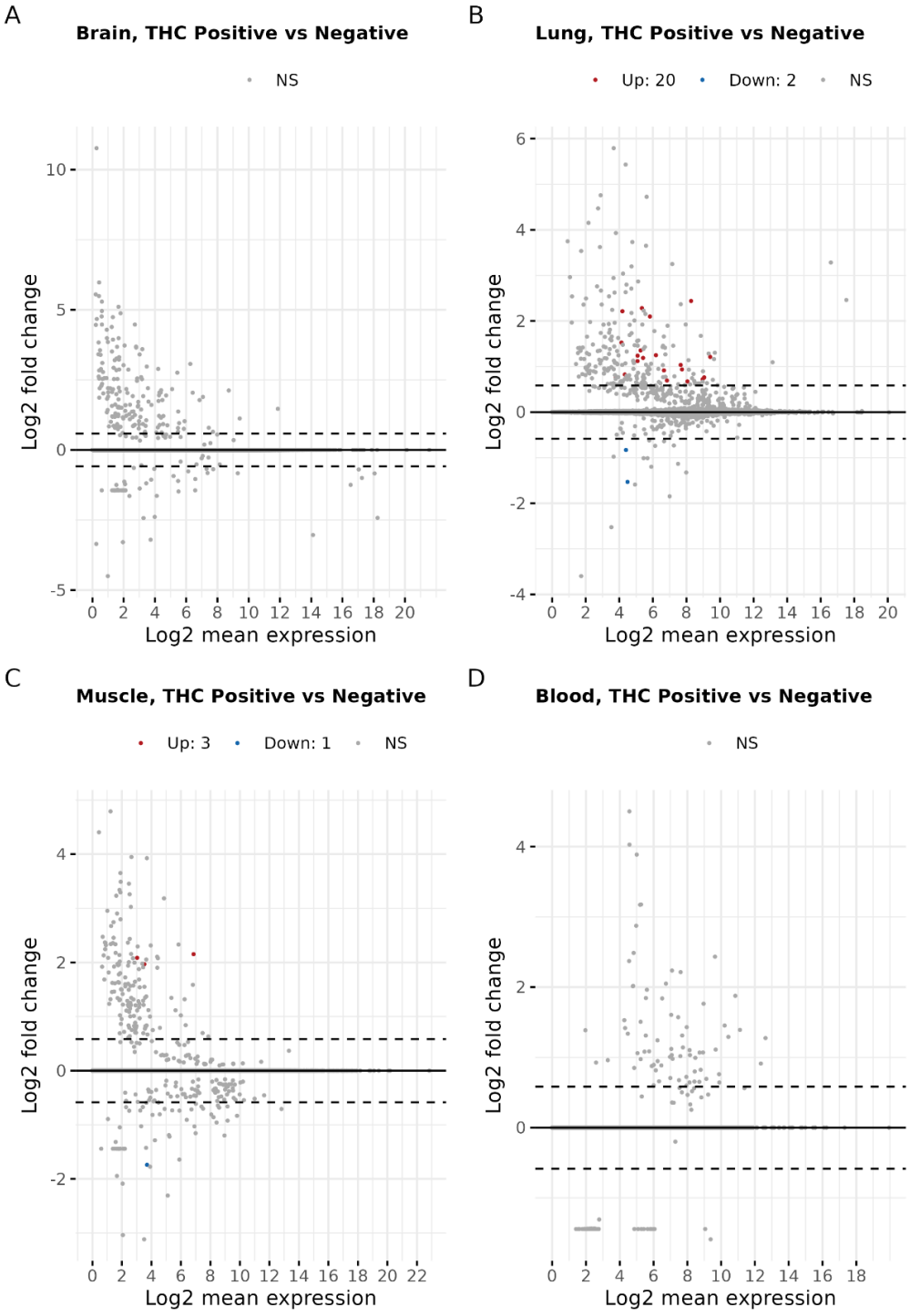

Figure 3. MA plots of significant THC-positive vs THC-negative DEGs. MA plots for THC-positive vs THC-negative comparisons for brain (A), lung (B), muscle (C), and blood (D). The horizontal dashed lines represent the positive and negative 0.58 $\log_2$ fold change thresholds, and the red dots and blue dots represent significant differentially expressed up-regulated and down-regulated genes, respectively, with an FDR-adjusted *P* value threshold of 0.05. No significant DEGs were identified in brain and blood.

Note. THC = tetrahydrocannabinol; DEG = differentially expressed gene; FDR = false discovery rate.

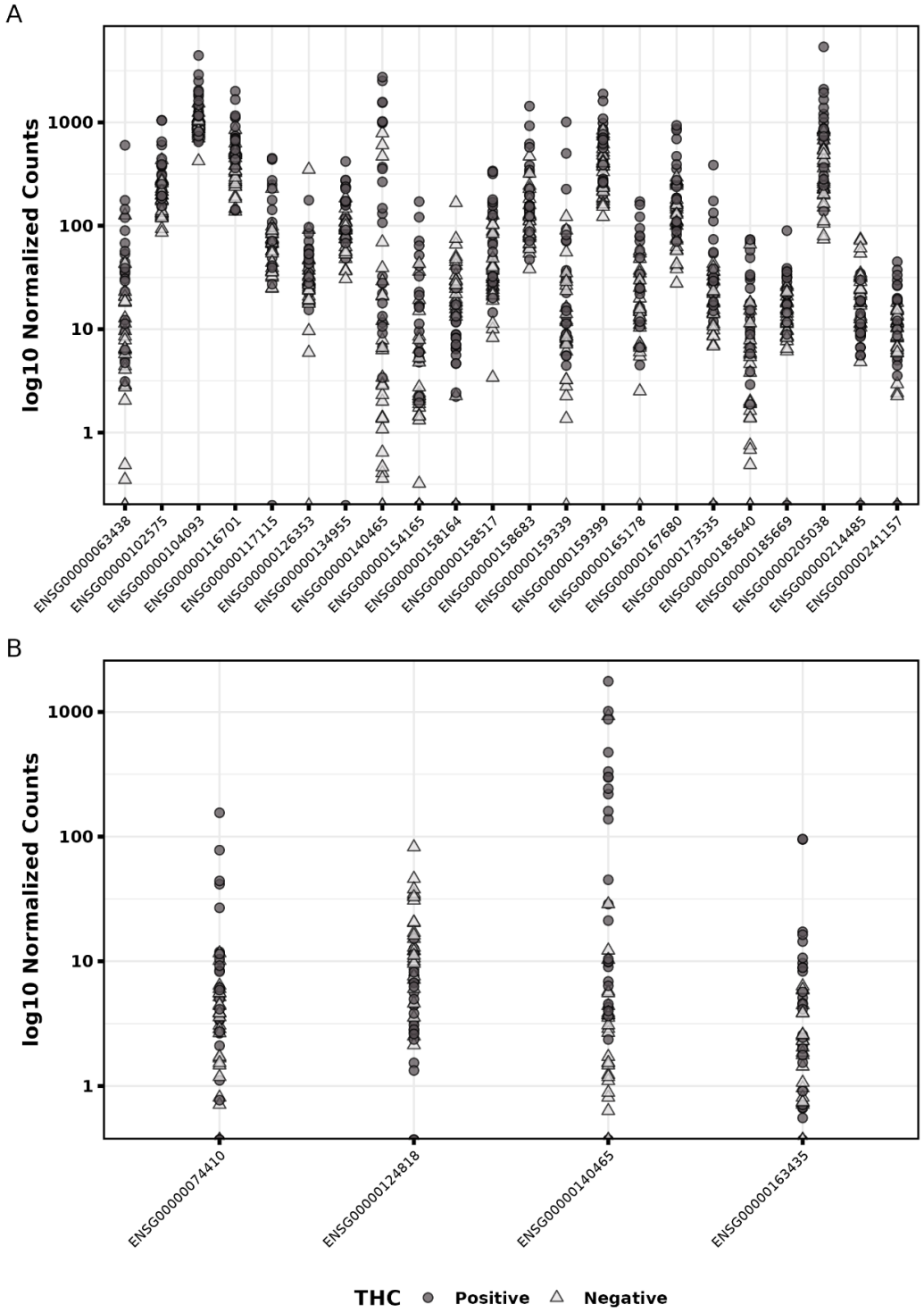

Figure 4. Significant THC-positive vs THC-negative DEGs. The normalized counts of the significant DEGs arranged from left to right by Ensembl ID number: lung (A) and muscle (B).

Note. THC = tetrahydrocannabinol; DEG = differentially expressed gene.

Table 2. Significant lung DEGs. Entrez annotated differentially expressed genes identified in THC-positive vs THC-negative lung samples using an FDR-adjusted *P* value (padj) significance threshold of 0.1 and a $\log_2$ fold change (log2FoldChange) threshold of 0.58.

Note. THC = tetrahydrocannabinol; FDR = false discovery rate.

Gene	Entrez	Symbol	baseMean	log2FoldChange	lfcSE	pvalue	padj	Chr	Start	End	Strand	Description
ENSG00000063438	57491	AHRR	39.680	2.281	0.485	4.47e-08	9.72e-04	5	321,714	438,291	1	aryl hydrocarbon receptor repressor
ENSG00000158164	11013	TMSB15A	21.441	-1.531	0.320	1.08e-07	1.18e-03	X	102,513,682	102,516,739	-1	thymosin beta 15A
ENSG00000154165	2838	GPR15	17.183	2.215	0.547	9.34e-07	6.78e-03	3	98,531,978	98,534,681	1	G protein-coupled receptor 15
ENSG00000159339	23569	PADI4	55.639	2.098	0.542	2.59e-06	1.12e-02	1	17,308,195	17,364,004	1	peptidyl arginine deiminase 4
ENSG00000173535	8794	TNFRSF10C	37.443	1.353	0.340	2.76e-06	1.12e-02	8	23,102,921	23,117,445	1	TNF receptor superfamily member 10c
ENSG00000185669	333929	SNAI3	19.078	0.825	0.213	3.59e-06	1.12e-02	16	88,677,688	88,686,507	-1	snail family transcriptional repressor 3
ENSG00000205038	93035	PKHD1L1	681.956	1.212	0.310	3.60e-06	1.12e-02	8	109,362,461	109,537,207	1	PKHD1 like 1
ENSG00000158517	653361	NCF1	71.617	1.249	0.355	1.48e-05	4.03e-02	7	74,774,011	74,789,315	1	neutrophil cytosolic factor 1
ENSG00000165178	654817	NCF1C	33.028	1.238	0.365	2.08e-05	4.52e-02	7	75,156,639	75,172,044	-1	neutrophil cytosolic factor 1C pseudogene
ENSG00000116701	4688	NCF2	532.308	0.762	0.226	2.93e-05	5.11e-02	1	183,554,461	183,590,905	-1	neutrophil cytosolic factor 2
ENSG00000158683	168507	PKD1L1	200.743	1.035	0.314	3.05e-05	5.11e-02	7	47,740,202	47,948,466	-1	polycystin 1 like 1, transient receptor potential channel interacting
ENSG00000185640	338785	KRT79	16.521	1.525	0.484	3.87e-05	6.01e-02	12	52,821,408	52,834,311	-1	keratin 79
ENSG00000134955	219855	SLC37A2	113.411	0.694	0.215	4.73e-05	6.86e-02	11	125,063,302	125,090,516	1	solute carrier family 37 member 2
ENSG00000102575	54	ACP5	265.403	0.674	0.224	9.04e-05	9.62e-02	19	11,574,653	11,579,993	-1	acid phosphatase 5, tartrate resistant
ENSG00000117115	11240	PADI2	100.079	0.916	0.310	1.03e-04	9.62e-02	1	17,066,761	17,119,451	-1	peptidyl arginine deiminase 2
ENSG00000140465	1543	CYP1A1	308.918	2.439	0.797	7.14e-05	9.62e-02	15	74,719,542	74,725,536	-1	cytochrome P450 family 1 subfamily A member 1
ENSG00000159399	3099	HK2	499.815	0.722	0.238	8.89e-05	9.62e-02	2	74,834,127	74,893,359	1	hexokinase 2
ENSG00000167680	10501	SEMA6B	212.123	0.938	0.312	8.34e-05	9.62e-02	19	4,542,593	4,581,776	-1	semaphorin 6B

Table 3. Significant muscle DEGs. Entrez annotated differentially expressed genes identified in THC-positive vs THC-negative muscle samples using an FDR-adjusted *P* value (padj) significance threshold of 0.1 and a $\log_2$ fold change (log2FoldChange) threshold of 0.58.

Note. THC = tetrahydrocannabinol; FDR = false discovery rate.

Gene	Entrez	Symbol	baseMean	log2FoldChange	lfcSE	pvalue	padj	Chr	Start	End	Strand	Description
ENSG00000140465	1543	CYP1A1	115.075	2.153	0.644	2.50e-20	1.40e-15	15	74,719,542	74,725,536	-1	cytochrome P450 family 1 subfamily A member 1
ENSG00000124818	221391	OPN5	12.025	-1.740	0.329	4.50e-09	1.26e-04	6	47,781,982	47,832,780	1	opsin 5
ENSG00000074410	771	CA12	10.452	1.967	0.506	2.97e-06	4.14e-02	15	63,321,378	63,381,846	-1	carbonic anhydrase 12
ENSG00000163435	1999	ELF3	7.176	2.085	0.533	2.58e-06	4.14e-02	1	202,007,945	202,017,183	1	E74 like ETS transcription factor 3

Microbial Contaminant Analysis

The brain samples had the lowest percent microbial contamination at 2.18% of total reads originating from exogenous contamination (Fig. 5A), followed by muscle at 5.52% (Fig. 7A), lung at 9.55% (Fig. 6A), and blood at 13.23% (Fig. 8A). The top five most prevalent microbial species in each sample were determined. The blood samples showed the greatest microbial diversity with 21 microbial species with over 1,000,000 matching reads in an individual sample across all blood samples (Fig. 8B), followed by the lung samples with 15 species (Fig. 6B), the muscle samples with 14 species (Fig. 7B), and the brain samples with 10 species (Fig. 5B). Finally, the number of times a given species appeared in the top five most prevalent microbial species per sample across tissues was determined. Across brain, lung, and muscle, *Bacillus subtilis* was the most frequently prevalent species, with *Methanocaldococcus sp. FS406-22* most frequently prevalent in blood (Fig. 5C, 6C, 7C, and 8C), corresponding to the use of the ERCC spike-in prior to RNA-Seq library construction, which derives a proportion of its spike-in fragments from *Bacillus subtilis* and *Methanocaldococcus jannaschii* genomes (Jiang et al., 2011).

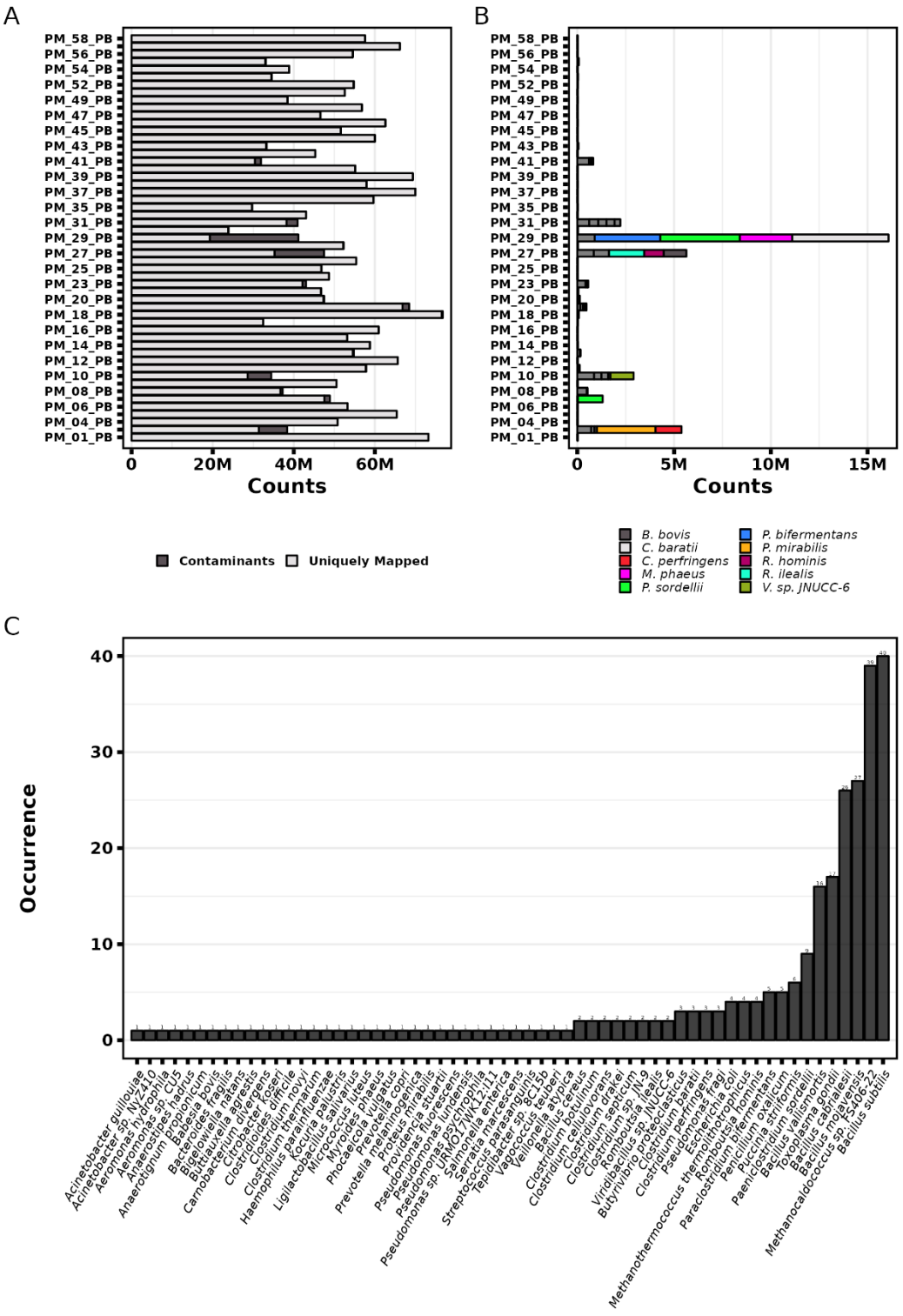

Figure 5. Brain microbial contaminant analysis. (A) Stacked column plot of uniquely mapped human RNA-Seq reads and reads associated with microbial contaminants per sample. (B) Stacked column plot of the top five most prevalent microbial species in each sample. The species with read counts >1,000,000 in individual samples are colored according to the legend. (C) Column plot representing the number of samples for which each species appeared in the top five most prevalent microbial species for each sample.

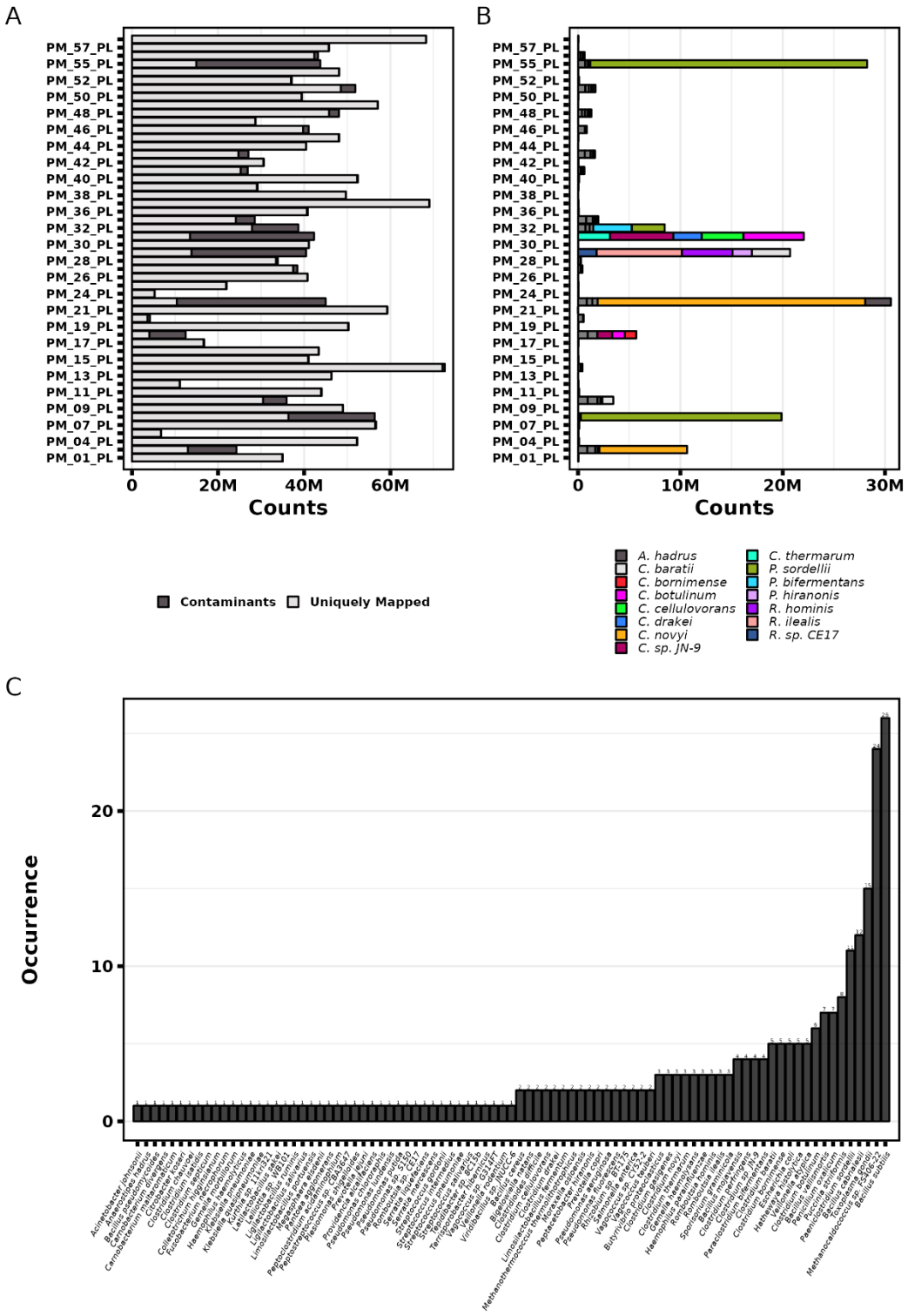

Figure 6. Lung microbial contaminant analysis. (A) Stacked column plot of uniquely mapped human RNA-Seq reads and reads associated with microbial contaminants per sample. (B) Stacked column plot of the top five most prevalent microbial species in each sample. The species with read counts >1,000,000 in individual samples are colored according to the legend. (C) Column plot representing the number of samples for which each species appeared in the top five most prevalent microbial species for each sample.

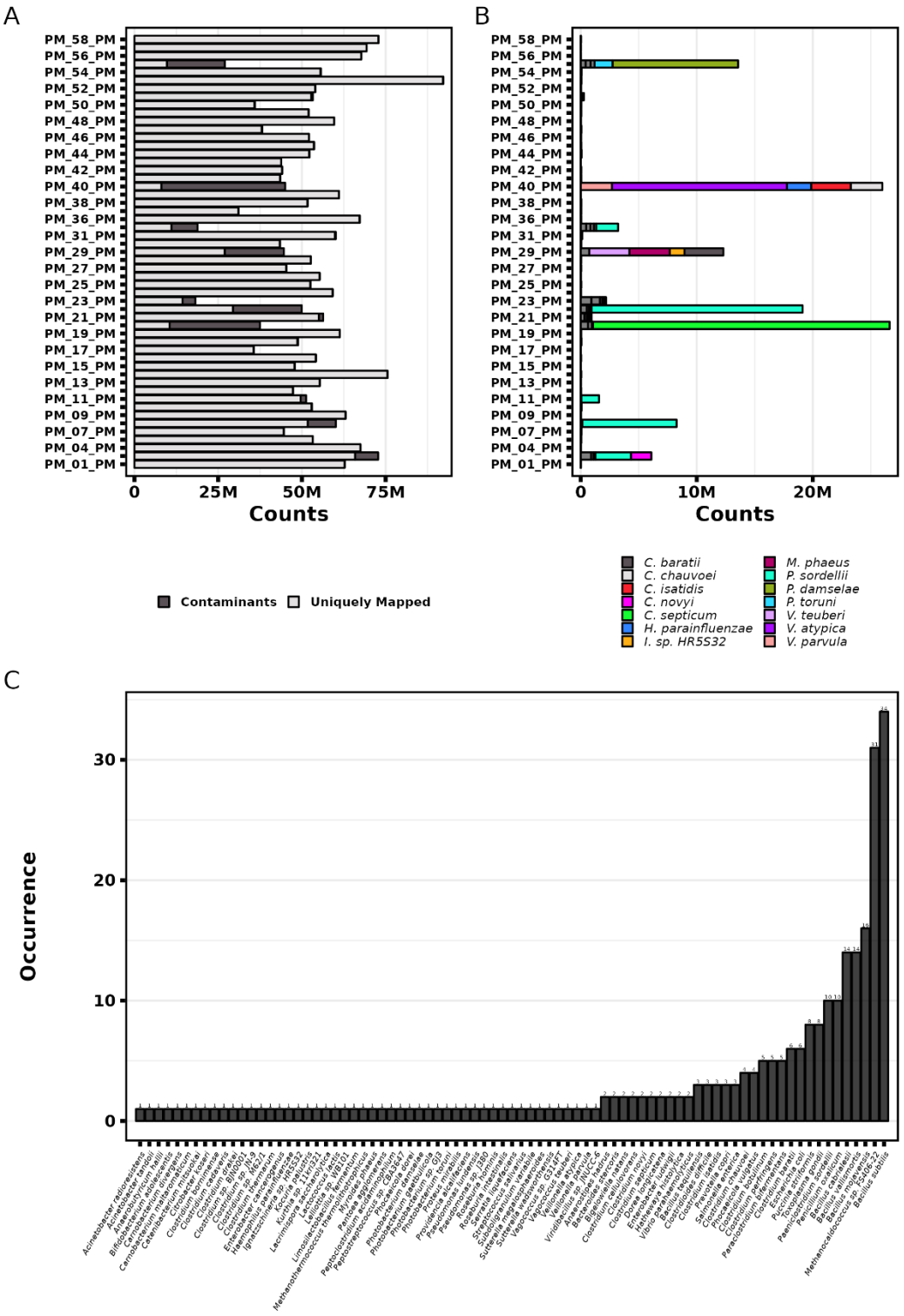

Figure 7. Muscle microbial contaminant analysis. (A) Stacked column plot of uniquely mapped human RNA-Seq reads and reads associated with microbial contaminants per sample. (B) Stacked column plot of the top five most prevalent microbial species in each sample. The species with read counts >1,000,000 in individual samples are colored according to the legend. (C) Column plot representing the number of samples for which each species appeared in the top five most prevalent microbial species for each sample.

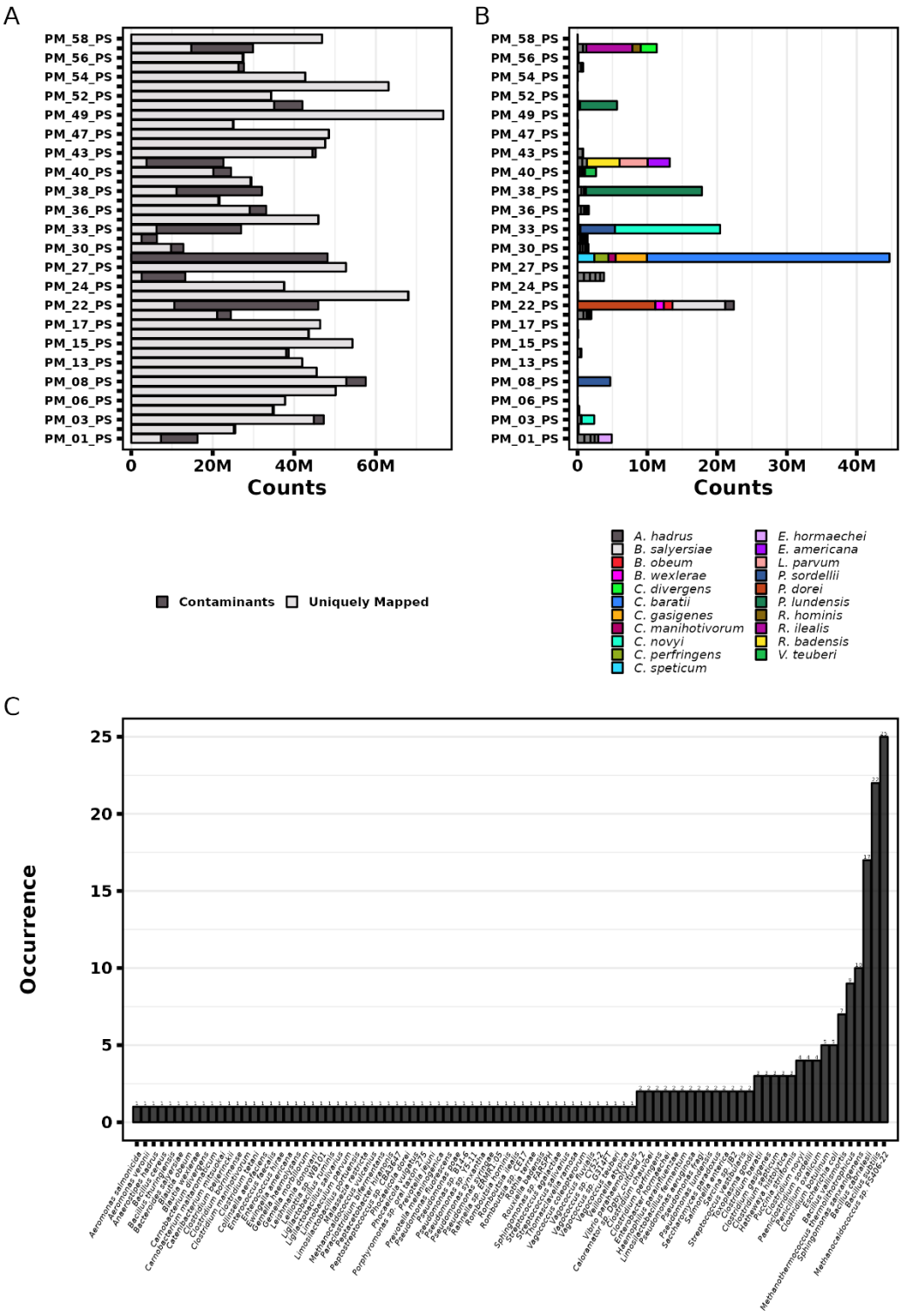

Figure 8. Blood microbial contaminant analysis. (A) Stacked column plot of uniquely mapped human RNA-Seq reads and reads associated with microbial contaminants per sample. (B) Stacked column plot of the top five most prevalent microbial species in each sample. The species with read counts >1,000,000 in individual samples are colored according to the legend. (C) Column plot representing the number of samples for which each species appeared in the top five most prevalent microbial species for each sample.

4. Discussion

RNA Quality

The RNA extracted from postmortem brain, lung, muscle, and blood samples collected from fatal general aviation accident victims is, in general, of poor quality in comparison to RNA extracted from samples collected using best practices. Under ideal circumstances, tissue samples are collected immediately after death, flash frozen in liquid nitrogen, stored on dry ice, or placed into one of several commercial DNA/RNA stabilization/preservation reagents, and then immediately placed into an ultra-low (-80°C) temperature freezer for storage until RNA is extracted (Auer et al., 2014; QIAGEN, 2023). In some cases, stabilization reagents allow for samples to be kept for a period at room temperature. Otherwise, at all times during sample collection, handling, and up until extraction, great care is taken to prevent the tissue samples from thawing by working quickly and transporting intact tissue samples on dry ice whenever outside of an ultra-low temperature freezer; deviations from these practices may result in RNA degradation and a marked decrease in RIN. In contrast, tissue samples collected for analysis in this study were exposed to adverse conditions at all stages during the collection process. The subject population experienced, by definition, violent and/or catastrophic injury leading to or following death, and in some cases experienced post-crash fire, or remained for hours or days at ambient temperature and weather conditions. Remains were transported to and sampled at a Medical Examiner or Coroner's office from one to several days following refrigeration. Following autopsy, tissues may or may not have been frozen and sent via parcel service from across the United States with or without ice packs in insulated Styrofoam shipping containers to the FAA/CAMI facility in Oklahoma City, OK. The QA team then inventoried and accessioned the specimens and stored them in a freezer (-20°C). After toxicological evaluation by the TOX team (turnaround time for which is up to 30 days for cannabinoid-negative samples and up to 60 days for cannabinoid-positive samples, during which time the tissues remained in storage at -20°C), the GEN team collected samples as they were available for this study. Tissues were thawed at room temperature, biopsied and then placed into DNA/RNA stabilization reagent before finally being stored in ultra-low (-80°C) temperature freezers.

Storing mammalian whole blood samples at 4°C for as little as 16 hours has been seen to significantly decrease the integrity of RNA such that it is no longer considered high-quality (a RIN value of >8) but is still usable in gene expression array analyses (a RIN value between 6 and 7) and after as few as three days at 4°C is considered degraded (a RIN value of <3) (Song & Zhou, 2020). The blood samples analyzed in this study were originally collected in Vacutainer Sodium Fluoride/Potassium Oxalate tubes for use in toxicological analysis, not with the intention of preserving and stabilizing RNA. Such tubes exist, for example the BD PAXgene Blood RNA Tube, and had these tubes been utilized for blood collection, the quality of the extracted RNA would likely have been higher. Decreasing RIN values of RNA extracted from human postmortem tissues has been correlated with increasing PMI in a clinical setting, but the extent of the degradation and the period over which it occurs is tissue-dependent (Walker et al., 2016).

Therefore, it is unsurprising that the RNA extracted from tissue samples in the present study was of low quality. However, surprisingly, the samples were not entirely degraded (corresponding to a RIN value of 1, or for which a RIN value could not be determined), with a mean RIN value across all samples of 4.13±1.44. Any downstream analysis using RNA extracted from aviation accident victims should be done with the foreknowledge that samples will be degraded; however, the samples in this study were useful for sequencing, albeit with a large degree of expression variability.

Putative Smoking and/or THC Related Biomarkers

All of the genes identified as significantly differentially expressed in this study in lung tissue (Table 2) have previously been identified in the literature as associated in some capacity with tobacco and cigarette smoking in lung specimens and blood, respiratory conditions, or certain lung-associated cancers and tumors. *AHRR* is a known epigenetic biomarker of tobacco smoking, with the cg05575921 CpG residue showing significant differential methylation in DNA extracted from whole blood (Zeilinger et al., 2013; Andersen et al., 2015). Persistent downregulation of *TMSB15A* has been implicated in chronic obstructive pulmonary disease progression (Samaha et al., 2021). Increased *GPR15* expression and methylation at the *GPR15* ch19859270 CpG occurs in helper T cells in response to both tobacco and cannabis smoking (Andersen et al., 2020; Bauer, 2021; Andersen et al., 2021). Increased, but not significant, *PADI4* expression has been observed in bronchoalveolar lavage cells in smokers relative to non-smokers (Makrygiannakis et al., 2008). The *TNFRSF10C* promoter is methylated in primary lung adenocarcinomas from nonsmokers compared to current and former smokers (Tessema et al., 2009). Increased *SNAI3* expression has been observed in lung adenocarcinomas, squamous cell carcinomas, and large cell NE carcinomas from smokers (Prieto et al., 2017). *PKHD1L1* has been identified as a putative prognostic biomarker between non-smoking and smoking-associated lung adenocarcinomas (Zhou et al., 2019). Some evidence exists that suggests *NCF1* may be regulated by polycyclic aromatic hydrocarbons, such as those found in cigarette smoke, while *NCF1C* is a *NCF1* pseudogene with no known function (Pinel-Marie et al., 2009; Zhong et al., 2018). *NCF2* has been implicated in several relevant cancers, including esophageal squamous cell carcinoma and non-small cell lung cancer, and its expression may be upregulated by smoking (Qin et al., 2020; Yang et al., 2021; Zhou et al., 2021). A single nucleotide polymorphism within *PKD1L1* has been identified within a genome-wide association study for overlap of asthma and obstructive pulmonary disease in African Americans (Hardin et al., 2014). *KRT79* was identified amongst a panel of genes whose expression was downregulated in peripheral blood following smoking cessation but was seen to be upregulated in THC-positive lung tissues in this study (Ungaro et al., 2016). In bronchoalveolar lavage cells, a differentially methylated position within *SLC37A2* has been identified in smokers versus non-smokers (Ringh et al., 2019). *ACP5* expression has been seen to be induced by cigarette smoke in lung specimens of self-reported smokers (Morissette et al., 2014). Significantly increased expression of *PADI2* has been observed in the bronchoalveolar lavage cells of smokers (Makrygiannakis et al., 2008). *CYP1A1* has long been known to be induced in response to cigarette smoke, and the encoded

protein of *CYP1A1*, a cytochrome P450 superfamily member, is integral to the metabolic activation of polycyclic aromatic hydrocarbons (Nebert, 1991; Slattery et al., 2004). *HK2* and its gene product hexokinase 2 regulate AHR (aryl hydrocarbon receptor) signaling, the transcription of which is further affected by halogenated aromatic hydrocarbon compounds, polycyclic aromatic hydrocarbons, and tryptophan derivatives; *HK2* over-expression has been linked to tumor growth (Watzky et al., 2022). *SEMA6B* expression has previously been identified as having a dose-response relationship with daily smoked cigarettes and total smoked pack years (Cuppen et al., 2018).

In muscle tissue, several novel genes were identified as being differentially expressed (Table 3), alongside *CYP1A1*, which was also found to be differentially expressed in lung tissue. *OPN5* does not appear to have been previously associated with smoking, cannabis, or THC. Similarly, there is no evidence linking *CA12* with either cannabis or THC. There is little research linking *CA12* expression with smoking, aside from *CA12* being associated with several cancers, including esophageal squamous cell carcinoma and oral squamous cell carcinoma, for both of which smoking is considered a risk factor (Cheong et al., 2009; Xing & Liu, 2017). *ELF3* has been found to be differentially expressed in squamous cell cancers and adenocarcinomas of smokers, and its overexpression has also been identified in non-small cell lung cancer, regardless of smoker status (Woenckhaus et al., 2006; Wang et al., 2018).

Postmortem Contaminant in Accident Samples

Given the nature of the postmortem accident samples, the extent of microbial contamination present in the samples is unsurprising. The relative lack of contamination present in the brain samples is understandable, given the position of the brain relative to the other three tissues. Short of physical penetration of the skull resulting in exposure of the brain to exterior elements or some other particularly adverse conditions such as a long postmortem interval allowing for putrefaction to begin, it is expected that the brain would be relatively free of microbial contamination. Similarly, the muscle samples used in this study were primarily collected from the psoas muscle, a paraspinal muscle located in the lumbar region of the trunk. Again, short of severe injury to this region leading to its exposure to the outside air or becoming punctured by debris during the accident, it is expected that this muscle should remain intact and, therefore, relatively uncontaminated prior to putrefaction. Healthy human lung has a pre-existing microbiome antemortem, with microbes from *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Fusobacteria*, and *Actinobacteria* present under normal circumstances (Moffatt & Cookson, 2017). Further, the lungs have been thought to decompose more quickly relative to muscles, which are amongst the last tissues to decompose, alongside tendons and bones (Javan et al., 2019; Dash & Das, 2020). Finally, postmortem blood showed the highest degree of contamination, yet multiple samples had very low microbial counts, which is in keeping with an artifactual disparity in blood collection sites. Depending on the state of the decedent, a peripheral blood sample from the femoral vein is preferred; however, in cases of severe trauma, the

subclavian or jugular veins may be used. If peripheral blood is unavailable, cardiac blood can be collected, though it is at risk of putrefaction.

Additionally, cavity blood may also be sampled, but is generally contaminated by the rupture of the gut, urine, or decomposing bodily fluids (Beresford, 2023). While the extent of microbial contamination from subject to subject and from tissue to tissue varies, no sample from any of the subjects for any of the four tissues collected for this study was free of contaminants. Given the nature of these samples and the realities associated with aviation accidents and autopsy, microbial contamination, and, in some cases, extensive contamination, of postmortem samples is unavoidable and should be expected for any studies utilizing these samples in the future.

5. Conclusions

In this study, THC-positive and THC-negative control brain, lung, muscle, and blood samples were collected from fatal general aviation accident victims, RNA was extracted and sequenced using RNA-Seq, and then differential gene expression and microbial contamination analyzed. RNA quality across all tissues, as assessed by RIN value, was generally poor, falling considerably below the field-accepted standard for RNA-Seq (RIN>7.0). While this or even more severe degradation could be anticipated in light of the less-than-ideal sample preservation approach, this study supports prior work by the GEN team that postmortem aviation accident samples can yield viable molecular insights. Several of the samples collected for use in this study were putrefied at the time of sampling, as noted by the GEN team, and yet RNA was extracted from all. In light of this, we expect that RNA extracted from most aviation accident samples will be suitable for RNA-Seq.

Many samples were also heavily contaminated with microbial RNA. While the contamination itself is concerning from a sample-handling perspective, the alignment process of RNA-Seq filters these microbial RNA reads from the uniquely mapped human RNA read counts that are used for downstream differential gene expression analysis. Therefore, the main practical concern related to this contamination is that the microbial RNA subtracts from the total number of useful reads for a given sample. Also, if a given gene of interest is highly conserved across taxa, it is possible that sequence similarity could result in some mapping errors and inflate the counts of some genes by adding microbial to human transcripts. Unless a method of depleting microbial RNA prior to library preparation is developed or targeted analyses for human-specific expression are employed, postmortem samples should be sequenced at read depths sufficient to capture a global view of gene expression despite any microbial contamination (the present study targeted 100 million reads per sample). Despite their limitations, postmortem samples represent a unique resource for future studies, as long the proper caveats related to their handling are acknowledged and accounted for in experimental design.

No significant differentially expressed genes were identified for the THC-positive vs. THC-negative comparison in brain or blood samples. A number of such genes were identified,

however, in lung and muscle samples. Furthermore, the majority of these genes have been previously associated in the literature with smoking or respiratory system health in some capacity. These genes may represent putative biomarkers of THC consumption, though not of active impairment resulting from THC use, and it is unclear whether these genes will be useful in detecting THC use or simply smoking in general. Future studies are necessary to identify and correlate genetic biomarkers of THC consumption with cognitive impairment. This study lays the groundwork for how such biomarkers could be identified in practice to identify THC-related cognitive impairment as a contributing factor in fatal aviation accidents within the United States.

Author contributions: Conceptualization: SJN; Methodology: CJT, DCH, and SJN; Software: CJT; Investigation: CJT, DCH, SKM, VLW, and SJN; Formal Analysis: CJT; Visualization: CJT; Writing—original draft: CJT; Writing—review and editing: CJT, DCH, SKM, VLW, and SJN. **Funding**: This research was funded by the Federal Aviation Administration. **Conflicts of interest**: The authors declare that they have no competing interests. **Peer review**: Internal review was accomplished in accordance with AAM-600-0002-WI. **Data accessibility**: A list of scripts, R packages, R Markdown notebooks, data files, and metadata files needed for independent study replication are included as supplements to this publication.

Supplementary Files

Alignment

- supplemental_file_1_FAA_GEN_alignment_pipeline_gencode_v1.2_PM.sh; alignment and quality control pipeline code in the form of a bash script

Differential Gene Expression Analysis

- supplemental_file_2_gencode_PM_DESeq2_analysis_untrimmed.Rmd; R Markdown document containing the full DGEA code, supplemental files 3-8 are required
- supplemental_file_2_gencode_PM_DESeq2_analysis_untrimmed.nb.html; html accessible notebook of the above R Markdown document
- supplemental_file_3_metadata_untrimmed.csv; tissue and THC classification for all samples along with RNA extraction metrics
- supplemental_file_4_metadata_PB_untrimmed.csv; tissue and THC classification for PB samples only
- supplemental_file_5_metadata_PL_untrimmed.csv; tissue and THC classification for PL samples only
- supplemental_file_6_metadata_PM_untrimmed.csv; tissue and THC classification for PM samples only
- supplemental_file_7_metadata_PS_untrimmed.csv; tissue and THC classification for PS samples only

- supplemental_file_8_PM_gencode_featurecounts.Rmatrix.txt; matrix containing featurecounts data for all samples

Microbial Contaminant Analysis

- supplemental_file_9_PM_taxonomy_plots.Rmd; R Markdown document containing the full taxonomy analysis code, supplemental files 10-14 are required
- supplemental_file_9_PM_taxonomy_plots.nb.html; html accessible notebook of the above R Markdown document
- supplemental_file_10_gencode_PB_species_kraken_summary; species level kraken results summary for PB samples
- supplemental_file_11_gencode_PL_species_kraken_summary; species level kraken results summary for PL samples
- supplemental_file_12_gencode_PM_species_kraken_summary; species level kraken results summary for PM samples
- supplemental_file_13_gencode_PS_species_kraken_summary; species level kraken results summary for PS samples
- supplemental_file_14_star_alignment_plot.tsv; alignment read counts for all samples

6. References

Andersen, A. M., Lei, M. K., Beach, S. R. H., Philibert, R. A., Sinha, S., & Colgan, J. D. (2020). Cigarette and Cannabis Smoking Effects on GPR15+ Helper T Cell Levels in Peripheral Blood: Relationships with Epigenetic Biomarkers. *Genes, 11*(2), 149. https://doi.org/10.3390/genes11020149

Andersen, A. M., Lei, M. K., Beach, S. R. H., & Philibert, R. A. (2021). Inflammatory biomarker relationships with helper T cell GPR15 expression and cannabis and tobacco smoking. *Journal of Psychosomatic Research, 141*, 110326. https://doi.org/10.1016/j.jpsychores.2020.110326

Andersen, A. M., Dogan, M. V., Beach, S. R., & Philibert, R. A. (2015). Current and Future Prospects for Epigenetic Biomarkers of Substance Use Disorders. *Genes, 6*(4), 991–1022. https://doi.org/10.3390/genes6040991

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Antiga, L. G., Sibbens, L., Abakkouy, Y., Decorte, R., Van Den Bogaert, W., Van de Voorde, W., & Bekaert, B. (2021). Cell survival and DNA damage repair are promoted in the human blood thanatotranscriptome shortly after death. *Scientific Reports, 11*(1), 16585. https://doi.org/10.1038/s41598-021-96095-z

Auer, H., Mobley, J. A., Ayers, L. W., Bowen, J., Chuaqui, R. F., Johnson, L. A., Livolsi, V. A., Lubensky, I. A., McGarvey, D., Monovich, L. C., Moskaluk, C. A., Rumpel, C. A., Sexton, K. C., Washington, M. K., Wiles, K. R., Grizzle, W. E., & Ramirez, N. C. (2014). The effects of frozen tissue storage conditions on the integrity of RNA and protein. *Biotechnic & Histochemistry : Official Publication of the Biological Stain Commission, 89*(7), 518–528. https://doi.org/10.3109/10520295.2014.904927

Bache, S., & Wichkam, H. (2022). *magrittr: A Forward-Pipe Operator for R*. https://CRAN.R-project.org/package=magrittr

Bauer, M. (2021). The Role of GPR15 Function in Blood and Vasculature. *International Journal of Molecular Sciences, 22*(19), 10824. https://doi.org/10.3390/ijms221910824

Beresford, P. (2023). Guidance for Obtaining Post Mortem Samples for Toxicology Analysis. https://www.nbt.nhs.uk/sites/default/files/document/Obtaining%20Post%20Mortem%20Samples%20for%20Toxicological%20Analysis.pdf

Burian, D., Uyhelji, H. A., McCauley, A., Williams, D., Kupfer, D. M., White, V. L., White, C., Hutchings, D., Jung, R., & Smith, S. W. (2017). *Postmortem samples from aviation accident victims maintain tissue-specific mRNA expression profiles* (DOT/FAA/AM-17/16). Federal Aviation Administration. https://www.faa.gov/sites/faa.gov/files/data_research/research/med_humanfacs/oamtechreports/201716.pdf

Cheong, S. C., Chandramouli, G. V., Saleh, A., Zain, R. B., Lau, S. H., Sivakumaren, S., Pathmanathan, R., Prime, S. S., Teo, S. H., Patel, V., & Gutkind, J. S. (2009). Gene expression in human oral squamous cell carcinoma is influenced by risk factor exposure. *Oral Oncology, 45*(8), 712–719. https://doi.org/10.1016/j.oraloncology.2008.11.002

Chu, A., Chaiton, M., Kaufman, P., Goodwin, R. D., Lin, J., Hindocha, C., Goodman, S., & Hammond, D. (2023). Co-Use, Simultaneous Use, and Mixing of Cannabis and Tobacco: A Cross-National Comparison of Canada and the US by Cannabis Administration Type. *International Journal of Environmental Research and Public Health, 20*(5), 4206. https://doi.org/ https://doi.org/10.3390%2Fijerph20054206

Compton, R. P. (2017). *Marijuana-Impaired Driving – A Report to Congress* (DOT HS 812 440). National Highway Traffic Safety Administration. https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/812440-marijuana-impaired-driving-report-to-congress.pdf

Cuppen, B. V. J., Rossato, M., Fritsch-Stork, R. D. E., Concepcion, A. N., Linn-Rasker, S. P., Bijlsma, J. W. J., van Laar, J. M., Lafeber, F. P. J. G., Radstake, T. R., & all SRU investigators (2018). RNA sequencing to predict response to TNF-α inhibitors reveals possible mechanism for nonresponse in smokers. *Expert Review of Clinical Immunology, 14*(7), 623–633. https://doi.org/10.1080/1744666X.2018.1480937

Dash, H. R., & Das, S. (2020). Thanatomicrobiome and epinecrotic community signatures for estimation of postmortem time interval in human cadaver. *Applied Microbiology and Biotechnology, 104*(22), 9497–9512. https://doi.org/10.1007/s00253-020-10922-3

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England), 29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England), 32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Federal Aviation Administration. (2016). *"I tested positive for marijuana on a DOT/FAA drug test, but recreational use of marijuana is legal in my state. Will my test be cancelled?" Q&As for Safety-Sensitive Employees.* https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/drug_alcohol/policy/qa_sse/a3

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics (Oxford, England), 28*(20), 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

Gohel, D. (2023). *officer: Manipulation of Microsoft Word and PowerPoint Documents*. https://CRAN.R-project.org/package=officer

Gohel, D., & Skintzos, P. (2023a). *flextable: Functions for Tabular Reporting*. https://CRAN.R-project.org/package=flextable

Gohel, D., & Skintzos P. (2023b). *ggiraph: Make 'ggplot2' Graphics Interactive*. https://CRAN.R-project.org/package=ggiraph

Halawa, A. A., El-Adl, M. A., & Marghani, B. H. (2021). Postmortem heat stress upregulates thanatotranscriptome of genes encode inflammation, apoptosis and neuronal stress in brain of rats at short postmortem intervals. *Australian Journal of Forensic Sciences, 53*(3), 271-282. https://doi.org/10.1080/00450618.2019.1682669

Hardin, M., Cho, M., McDonald, M. L., Beaty, T., Ramsdell, J., Bhatt, S., van Beek, E. J., Make, B. J., Crapo, J. D., Silverman, E. K., & Hersh, C. P. (2014). The clinical and genetic features of COPD-asthma overlap syndrome. *The European Respiratory Journal, 44*(2), 341–350. https://doi.org/10.1183/09031936.00216013

Javan, G. T., Finley, S. J., Tuomisto, S., Hall, A., Benbow, M. E., & Mills, D. (2019). An interdisciplinary review of the thanatomicrobiome in human decomposition. *Forensic Science, Medicine, and Pathology, 15*(1), 75–83. https://doi.org/10.1007/s12024-018-0061-0

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., Gingeras, T. R., & Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research, 21*(9), 1543–1551. https://doi.org/10.1101/gr.121095.111

Kassambara, A. (2023). *ggpubr: 'ggplot2' Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr

Kolde, R. (2019). *pheatmap: Pretty Heatmaps*. https://CRAN.R-project.org/package=pheatmap

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England), 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England), 30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science, 3*, e104. https://doi.org/10.7717/peerj-cs.104

Makrygiannakis, D., Hermansson, M., Ulfgren, A. K., Nicholas, A. P., Zendman, A. J., Eklund, A., Grunewald, J., Skold, C. M., Klareskog, L., & Catrina, A. I. (2008). Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Annals of the Rheumatic Diseases, 67*(10), 1488–1492. https://doi.org/10.1136/ard.2007.075192

Martin, M. (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal, 17*(1): 10-12. https://doi.org/10.14806/ej.17.1.200

Mistry, M., Piper, M., Liu, J., & Khetani, R. (2021, May 24). hbctraining/DGE_workshop_salmon_online: Differential Gene Expression Workshop Lessons from HCBC (first release). *Zenodo*. https://doi.org/10.5281/zenodo.4783481

Moffatt, M. F., & Cookson, W. O. (2017). The lung microbiome in health and disease. *Clinical Medicine (London, England), 17*(6), 525–529. https://doi.org/10.7861/clinmedicine.17-6-525

Moon, K. (2020). *ggiraphExtra: Make Interactive 'ggplot'. Extension to 'ggplot2' and 'ggiraph'*. https://CRAN.R-project.org/package=ggiraphExtra

Morgan, M., & Shepherd, L. (2023). *AnnotationHub: Client to access AnnotationHub resources*. R package version 3.10.0, https://bioconductor.org/packages/AnnotationHub

Morissette, M. C., Lamontagne, M., Bérubé, J. C., Gaschler, G., Williams, A., Yauk, C., Couture, C., Laviolette, M., Hogg, J. C., Timens, W., Halappanavar, S., Stampfli, M. R., & Bossé, Y. (2014). Impact of cigarette smoke on the human and mouse lungs: a gene-expression comparison study. *PloS One, 9*(3), e92498. https://doi.org/10.1371/journal.pone.0092498

Nebert, D. W. (1991). Role of genetics and drug metabolism in human cancer risk. *Mutation Research, 247*(2), 267–281. https://doi.org/10.1016/0027-5107(91)90022-g

Norris, A., Cliburn, K., Kemp, P., & Skaggs, V. (2018). *Assessing Trends in Cannabinoid Concentrations Found in Specimens from Aviation Fatalities between 2007 and 2016* (DOT/FAA/AM-18/2). https://rosap.ntl.bts.gov/view/dot/57205

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics, 12*, 385. https://doi.org/10.1186/1471-2105-12-385

Pantano, L. (2023). *DEGreport: Report of DEG analysis*. https://doi.org/10.18129/B9.bioc.DEGreport

Papudeshi, B. (2022). npbhavya/Kraken2-output-manipulation. *Zenodo*. https://doi.org/10.5281/zenodo.6000803

Pedersen, T. (2022). *patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork

Pinel-Marie, M. L., Sparfel, L., Desmots, S., & Fardel, O. (2009). Aryl hydrocarbon receptor-dependent induction of the NADPH oxidase subunit NCF1/p47^{phox} expression leading to priming of human macrophage oxidative burst. *Free Radical Biology & Medicine, 47*(6), 825–834. https://doi.org/10.1016/j.freeradbiomed.2009.06.025

Posit Team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. http://www.posit.co/

Prieto, T., De Sá, V., Olivieri, E., Da Silva, E., Reis, R., Carraro, D., & Capelozzi, V. (2017). P1. 07-030 Gene Signature of EMT in Neuroendocrine Lung Carcinoma: A Comparative Analysis with Adenocarcinoma and Squamous Cell Carcinoma: Topic: Molecular Changes. *Journal of Thoracic Oncology, 12*(1), S714-S715. https://doi.org/10.1016/j.jtho.2016.11.942

Protection of Human Subjects, 49 C.F.R. § 11. (2017). https://www.ecfr.gov/current/title-49/part-11/section-11.102

QIAGEN, Inc. (2023). RNeasy Mini Handbook 04/2023. https://www.qiagen.com/us/resources/download.aspx?id=f646813a-efbb-4672-9ae3-e665b3045b2b&lang=en

Qin, K., Zheng, Z., He, Y., Gao, Y., Shi, H., Mo, S., Zhang, J., & Rong, J. (2020). High expression of neutrophil cytosolic factor 2 (NCF2) is associated with aggressive features and poor prognosis of esophageal squamous cell carcinoma. *International Journal of Clinical and Experimental Pathology, 13*(12), 3033–3043.

R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rainer, J., Gatto, L., & Weichenberger, C. X. (2019). ensembldb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics (Oxford, England), 35*(17), 3151–3153. https://doi.org/10.1093/bioinformatics/btz031

Regions and Centers, 49 U.S.C. § 44507. (2021). https://www.govinfo.gov/app/details/USCODE-2021-title49/USCODE-2021-title49-subtitleVII-partA-subpartiii-chap445-sec44507

Ringh, M. V., Hagemann-Jensen, M., Needhamsen, M., Kular, L., Breeze, C. E., Sjöholm, L. K., Slavec, L., Kullberg, S., Wahlström, J., Grunewald, J., Brynedal, B., Liu, Y., Almgren, M., Jagodic, M., Öckinger, J., & Ekström, T. J. (2019). Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in

bronchoalveolar lavage cells. *EBioMedicine, 46*, 290–304. https://doi.org/10.1016/j.ebiom.2019.07.006

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), e47. https://doi.org/10.1093/nar/gkv007

Samaha, E., Vierlinger, K., Weinhappel, W., Godnic-Cvar, J., Nöhammer, C., Koczan, D., Thiesen, H. J., Yanai, H., Fraifeld, V. E., & Ziesche, R. (2021). Expression Profiling Suggests Loss of Surface Integrity and Failure of Regenerative Repair as Major Driving Forces for Chronic Obstructive Pulmonary Disease Progression. *American Journal of Respiratory Cell and Molecular Biology, 64*(4), 441–452. https://doi.org/10.1165/rcmb.2020-0270OC

Signal, B. (2023). *GeneStructureTools: Tools for spliced gene structure manipulation and analysis*. https://doi.org/10.18129/B9.bioc.GeneStructureTools

Slattery, M. L., Samowtiz, W., Ma, K., Murtaugh, M., Sweeney, C., Levin, T. R., & Neuhausen, S. (2004). CYP1A1, cigarette smoking, and colon and rectal cancer. *American Journal of Epidemiology, 160*(9), 842–852. https://doi.org/10.1093/aje/kwh298

Slowikowski, K. (2023). *ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*. https://CRAN.R-project.org/package=ggrepel

Song, J., & Zhou, J. (2020). Effects of preservation duration at 4 °C on the quality of RNA in rabbit blood specimens. *PeerJ, 8*, e8940. https://doi.org/10.7717/peerj.8940

Substance Abuse and Mental Health Services Administration. (2021). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2020 National Survey on Drug Use and Health*. HHS Publication No. PEP21-07-01-003, NSDUH Series H-56 https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf

Tange, O. (2018). GNU Parallel 2018. In *GNU Parallel 2018* (p. 112). Ole Tange. https://doi.org/10.5281/zenodo.1146014

Tessema, M., Yu, Y. Y., Stidley, C. A., Machida, E. O., Schuebel, K. E., Baylin, S. B., & Belinsky, S. A. (2009). Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers. *Carcinogenesis, 30*(7), 1132–1138. https://doi.org/10.1093/carcin/bgp114

Turner, S. (2023). *annotables: Ensembl Annotation Tables*. https://github.com/stephenturner/annotables

United States. (2021). 49 U.S.C. § 1111 - General authority (Subtitle II, Chap. 11, Subchap. II, Sec. 1111). https://www.govinfo.gov/content/pkg/USCODE-2021-title49/html/USCODE-2021-title49-subtitleII-chap11-subchapII-sec1111.htm

U.S. Department of Transportation. (2021). *FAA Organization – Policies and Standards (Order No. 1100.1C)*. Federal Aviation Administration. https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.information/documentID/1040403

Ungaro, R. C., Mehandru, S., Pandey, G., Clemente, J., Faith, J., Lee, H., ... & Sands, B. (2016). A Pilot Study on the Effects of Smoking Cessation on Intestinal Gene Expression and the Microbiome: 725. *Official Journal of the American College of Gastroenterology| ACG, 111*, S326-S327.

Walker, D. G., Whetzel, A. M., Serrano, G., Sue, L. I., Lue, L. F., & Beach, T. G. (2016). Characterization of RNA isolated from eighteen different human tissues: results from a rapid human autopsy program. *Cell and Tissue Banking, 17*(3), 361–375. https://doi.org/10.1007/s10561-016-9555-8

Wang, H., Yu, Z., Huo, S., Chen, Z., Ou, Z., Mai, J., Ding, S., & Zhang, J. (2018). Overexpression of ELF3 facilitates cell growth and metastasis through PI3K/Akt and ERK signaling pathways in non-small cell lung cancer. *The International Journal of Biochemistry & Cell Biology, 94*, 98–106. https://doi.org/10.1016/j.biocel.2017.12.002

Watzky, M., Huard, S., Juricek, L., Dairou, J., Chauvet, C., Coumoul, X., Letessier, A., & Miotto, B. (2022). Hexokinase 2 is a transcriptional target and a positive modulator of AHR signaling. *Nucleic Acids Research, 50*(10), 5545–5564. https://doi.org/10.1093/nar/gkac360

Wickham, H. (2016). ggplot2: *Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

Wickham, H., & D. Seidel (2022). *scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Woenckhaus, M., Klein-Hitpass, L., Grepmeier, U., Merk, J., Pfeifer, M., Wild, P., Bettstetter, M., Wuensch, P., Blaszyk, H., Hartmann, A., Hofstaedter, F., & Dietmaier, W. (2006). Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *The Journal of Pathology, 210*(2), 192–204. https://doi.org/10.1002/path.2039

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology, 15*(3), R46. https://doi.org/10.1186/gb-2014-15-3-r46

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Wright, K. (2021). pals: *Color Palettes, Colormaps, and Tools to Evaluate Them*. https://CRAN.R-project.org/package=pals

Xing, J., & Liu, C. (2017). Identification of genes associated with histologic tumor grade of esophageal squamous cell carcinoma. *FEBS Open Bio, 7*(9), 1246–1257. https://doi.org/10.1002/2211-5463.12228

Yang, S., Tang, D., Zhao, Y. C., Liu, H., Luo, S., Stinchcombe, T. E., Glass, C., Su, L., Shen, S., Christiani, D. C., Wang, Q., & Wei, Q. (2021). Potentially functional variants of ERAP1, PSMF1 and NCF2 in the MHC-I-related pathway predict non-small cell lung cancer survival. *Cancer Immunology, Immunotherapy: CII, 70*(10), 2819–2833. https://doi.org/10.1007/s00262-021-02877-9

Zeilinger, S., Kühnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A., Strauch, K., Waldenberger, M., & Illig, T. (2013). Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS One, 8*(5), e63812. https://doi.org/10.1371/journal.pone.0063812

Zhong, J., Olsson, L. M., Urbonaviciute, V., Yang, M., Bäckdahl, L., & Holmdahl, R. (2018). Association of NOX2 subunits genetic variants with autoimmune diseases. *Free Radical Biology & Medicine, 125*, 72–80. https://doi.org/10.1016/j.freeradbiomed.2018.03.005

Zhou, D., Sun, Y., Jia, Y., Liu, D., Wang, J., Chen, X., Zhang, Y., & Ma, X. (2019). Bioinformatics and functional analyses of key genes in smoking-associated lung adenocarcinoma. *Oncology Letters, 18*(4), 3613–3622. https://doi.org/10.3892/ol.2019.10733

Zhou, S., Liu, S., Liu, X., & Zhuang, W. (2021). Bioinformatics Gene Analysis of Potential Biomarkers and Therapeutic Targets for Unstable Atherosclerotic Plaque-Related Stroke. *Journal of Molecular Neuroscience: MN, 71*(5), 1031–1045. https://doi.org/10.1007/s12031-020-01725-2

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics (Oxford, England), 35*(12), 2084–2092. https://doi.org/10.1093/bioinformatics/bty895