UNIVERSITÄT BERN

Graduate School for Cellular and Biomedical Sciences

University of Bern

Analysis of bronchoalveolar lavage transcriptome profiles of asthmatic horses by single-cell mRNA sequencing

PhD Thesis submitted by

Sophie Elena Sage

for the degree of

Doctor of Veterinary Medicine and Philosophy (DVM, PhD)

Supervisor
Prof. Dr. med. vet. Vinzenz Gerber
Swiss Institute of Equine Medicine
Department of Clinical Veterinary Medicine
Vetsuisse Faculty of the University of Bern
Co-advisor
Prof. Dr. rer. nat. Tosso Leeb
Institute of Genetics
Vetsuisse Faculty of the University of Bern

b
UNIVERSITÄT
BERN

Accepted by the Faculty of Medicine, the Faculty of Science and the Vetsuisse Faculty of the University of Bern at the request of the Graduate School for Cellular and Biomedical Sciences

Bern, Dean of the Faculty of Medicine

Bern, Dean of the Faculty of Science

Bern, Dean of the Vetsuisse Faculty Bern

Table of contents

Acknowledgments 2
Abbreviations 3
Abstract 5
Introduction 6
1.1 Severe equine asthma 6
1.1.1 Importance of equine asthma 6
1.1.2 Immunopathogenesis – current knowledge 10
1.1.3 Origin of the knowledge gap 13
1.2 Single-cell mRNA sequencing 16
1.2.1 A practical guide 16
1.2.2 Strengths 25
1.2.3 Challenges 31
Hypotheses and Aims of Thesis 41
Results 43
2.1 Study 1: Single-cell gene expression analysis of cryopreserved equine bronchoalveolar cells 43
2.2 Study 2: Long read transcriptome of equine bronchoalveolar cells 62
2.3 Study 3: Single-cell profiling of bronchoalveolar cells reveals a Th17 signature in neutrophilic severe equine asthma 76
Discussion 103
Outlook 112
References 114
Originally signed Declaration of Originality 126

Acknowledgments

I would like to express my profound gratitude to my supervisor Prof. Dr. Vinzenz Gerber for entrusting me with this project and offering me the opportunity to join the ISME team. The realization of my thesis was made possible by his research vision and unwavering support.

I wish to express my deepest appreciation to my co-supervisor, Prof. Dr. rer. nat. Tosso Leeb, for graciously providing his expertise and invaluable insights throughout the entirety of my thesis.

This endeavor would not have been possible without Dr. Vidhya Jagannathan. I am sincerely grateful for her kind support throughout my challenging journey in learning bioinformatics. I am also deeply indebted to Dr. Pamela Nicholson for her assistance with the sequencing experiments. I extend my sincere appreciation to the laboratory technicians at the Next Generation Sequencing Platform for their patient guidance during benchwork.

Thanks should also go to Dr. Laureen M. Peters for her technical expertise in cytology and her valuable intellectual contributions. Furthermore, I want to acknowledge the diligent work of the clinical laboratory technicians who analyzed our numerous samples.

Special thanks are due to my colleagues Dr. med. vet. Michelle Wyler and Dr. med. vet. Nicole Altermatt for their incredible enthusiasm in sampling nearly 100 horses in the middle of a worldwide pandemic. I would also like to acknowledge the horse owners for their trust.

During the last four years, I had the privilege of working with incredible veterinarians, technicians, barn staff, and office staff. Thank you for making each day at work so enjoyable.

I could not have undertaken this journey without the support of my family. Thank you for standing by my side through thick and thin. I would not be where I am today without you.

Lastly, a heartfelt thank you to my friends scattered across France, Switzerland and the USA for their encouragement and for the much-needed moments of decompression.

This work was supported by the Swiss National Science Foundation (Grant No. 31003A-162548/1) and the Internal Research Fund of the Swiss Institute of Equine Medicine, Bern, Switzerland (ISMEquine Research No. 33-890).

Abbreviations

AM	alveolar macrophage
BALF	bronchoalveolar lavage fluid
cDNA	complementary DNA
CITE-seq	cellular indexing of transcriptomes and epitopes by sequencing
DEG	differentially expressed gene
DGE	differential gene expression
EA	equine asthma
FACS	fluorescence activated cell sorting
FAANG	functional annotation of animal genome
GEM	gel beads-in-emulsion
HCA	human cell atlas
HPA	human protein atlas
HPC	high-performance computing
IFNγ	interferon gamma
IgE	Immunoglobulin E
IL	interleukin
ILC	immune lymphoid cell
MEA	mild-moderate equine asthma
NET	neutrophil extracellular trap
NK	natural killer cell
NKT	natural killer T cell
PBMC	peripheral blood mononuclear cell

PCR	polymerase chain reaction
PM	particulate matter
RNA-seq	mRNA sequencing
RT-PCR	reverse transcription polymerase chain reaction
ROS	reactive oxygen species
RT	reverse transcription
SEA	severe equine asthma
scRNA-seq	single-cell mRNA sequencing
SMRT	single molecule real-time
t-SNE	t-distributed stochastic neighbour embedding
Th	T helper
UMAP	uniform manifold approximation and projection
UTR	untranslated region
ZMW	zero-mode wavelength

Abstract

Severe equine asthma (SEA) is a common respiratory condition of horses, whose underlying immune mechanisms remain to be elucidated. In this thesis project, we took advantage of the recently developed single-cell mRNA (scRNA-seq) technology to investigate the immunological landscape of equine bronchoalveolar lavage fluid (BALF) cells in both health and disease. Initially, we conducted a pilot experiment involving three horses to demonstrate the feasibility of scRNA-seq on cryopreserved equine BALF samples. Although the experiment was successful, the proportion of reads aligning to the annotated equine reference transcriptome was suboptimal. To address this, we generated a custom equine BALF transcriptome using long-read sequencing, aiming to improve the quality of 3'-UTR annotation and document BALF-specific isoforms. While we identified several novel isoforms, the read mapping percentage did not improve when aligning our scRNA-seq transcripts to the custom transcriptome. By extending the 3'-UTRs of the existing reference annotation, we achieved a satisfactory read mapping percentage, enabling subsequent qualitative downstream analysis. Our scRNA-seq dataset encompassed six major cell populations: monocytes-macrophages, neutrophils, T cells, B cells and dendritic cells. Within the monocyte-macrophage and T cell groups, we identified previously uncharacterized cell subtypes. Encouraged by these findings, we applied our optimized experimental protocol and analysis pipeline to study SEA. ScRNA-seq analysis of cryopreserved BALF cells from 6 asthmatic and 5 healthy controls revealed the same major cell populations as observed in the pilot study. In addition to T cells and monocytes-macrophages, we characterized several cell subtypes within the B cell, dendritic cell and neutrophil populations. Differential gene expression analysis revealed a strong T helper (Th)17 signature in SEA, primarily driven by monocytes-macrophages and T cells. Notably, BALF from SEA horses was enriched in B cells, with a lower proportion of activated plasma cells. Neutrophils in the SEA group displayed increased migratory capacity and a heightened propensity to form neutrophil extracellular traps (NETs). An intriguing finding in both scRNA-seq experiments was the detection of a dual monocyte-lymphocyte population, potentially representing genuine cellular complexes engaged in an immunological synapse. In summary, this thesis project represents pioneering work employing scRNA-seq in the field of equine pulmonology. Our findings support a predominant Th17 immune pathway in SEA, necessitating further investigation to improve diagnostic tools and therapeutic management of severely asthmatic horses.

Introduction

Asthma is a chronic inflammatory airway disease affecting 5 – 10 % of the global human population (1). Only two domestic animal species are known to be affected by naturally-occurring asthma: cats and horses. Equine asthma is common in Northern hemisphere countries with a temperate climate. Two main forms of the disease are described: mild-moderate (MEA) and severe equine asthma (SEA). Despite the high prevalence and clinical importance of asthma in humans and horses, its pathophysiological mechanisms are yet to be elucidated. Over the past decade, substantial advances in respiratory research have been made thanks to single-cell RNA sequencing (scRNA-seq). With this novel technique, genetic expression can be studied at the level of a single cell, offering unmatched resolution. ScRNA-seq studies remain scarce in asthma research, in part because human pulmonary cells are not readily accessible. The aim of this PhD thesis was to apply scRNA-seq to equine bronchoalveolar cells to characterize the immunological landscape of SEA. In the introduction, we first outline our current understanding of the pathophysiology of SEA in order to identify the knowledge gaps. We then provide an overview of the intricacies and advantages of the scRNA-seq technology, emphasizing its potential to uncover the underlying immune mechanisms of SEA.

1.1 Severe equine asthma

1.1.1 Importance of equine asthma

a. Overview of equine asthma

Asthma is a chronic, non-infectious, inflammatory disease of the lungs. In horses, two main disease categories are described: MEA (formerly inflammatory airway disease IAD) and SEA (formerly recurrent airway obstruction RAO or heaves), mostly observed in horses 7 years or older. Based on a survey in the United Kingdom, SEA is thought to affect 10 - 17% of the equine population in countries with temperate climate (2). Although MEA is widely recognized as a prevalent condition among horses, affecting as many as 80% of racehorses (3), there remains a lack of agreement regarding its exact prevalence (4). Both forms of the disease are characterized by chronic lower airway inflammation, bronchial hyperreactivity, partially reversible airway obstruction and increased tracheal mucus production. Clinical signs consist in a variable combination of cough, nasal discharge, exercise intolerance, and abnormal

breathing effort. The distinction between MEA and SEA is clinical, with severely asthmatic horses displaying increased breathing effort at rest (5). This presents with nasal flare and increased abdominal effort at expiration, eventually resulting in abdominal muscle overdevelopment (known as “heave line”). It is still debated whether MEA and SEA represent two different pathological entities or whether they are part of a continuum. Indeed, some MEA-affected horses never develop SEA (4). One argument in favor of two distinct pathophysiological entities was the presence of airway remodeling only in SEA. However, a recent study identified signs of bronchial remodeling in horses with MEA (6), fueling the hypothesis that MEA may represent an early stage of SEA. Since this thesis project specifically centers on the investigation of SEA, we will now shift our focus exclusively to this form of EA.

Diagnosis

In clinical practice, the diagnosis of SEA is typically based on the history, the clinical examination, and the response to therapy. The Horse Owner Assessed Respiratory Signs Index (HOARSI) questionnaire can be used to identify cases based on their respiratory sign history (7,8). Diagnosis confirmation is typically achieved through cytological examination of the bronchoalveolar lavage fluid (BALF). According to an expert consensus (4), SEA is characterized by a significant airway neutrophilia, with more than 25% neutrophils. However, the stringency of this cytological criterion has been called into question. For instance, paucigranulocytic SEA, where BALF cytology is normal, is occasionally encountered. This may reflect suboptimal sample quality due to bronchoconstriction and may not be representative of the lower airway compartment (9). Furthermore, it is now observed that some horses exhibit all the clinical features of SEA without the typical severe BALF neutrophilia, potentially due to improved husbandry practices and subsequent reduced dust exposure (10,11). Another characteristic of EA is an increased tracheal mucus score observed during endoscopy (12). Spirometry is paramount to the diagnosis of human asthma (1). However, it is not feasible to instruct a horse to forcefully exhale into an instrument. Several alternative methods exist to evaluate equine lung function, such as esophageal balloon catheterization, flow plethysmography, forced oscillatory mechanics or flow interruption techniques (9,13). However, lung function testing devices are not widely accessible and require expertise for both implementation and interpretation of results. Consequently, researchers in the field now advocate for a comprehensive approach that combines history, clinical and pathological findings, and/or lung function testing to classify horses as asthmatic or healthy (14–16).

Treatment

The mainstay of treatment is antigen avoidance through dust exposure reduction. Ideally, horses should be maintained in an environment that closely mimics their natural habitat, allowing them continuous pasture turnout. As this is rarely feasible, several measures must be implemented to reduce dust exposure in the stable, such as enhancing stall ventilation, avoiding storage of hay above the horses, or sprinkling the riding arena prior to use. A pivotal measure involves substituting dry hay with alternatives like soaked or steamed hay, haylage, or hay pellets. Moreover, straw bedding should be replaced by low dust alternatives (4). Pharmacological interventions are only palliative and involve the administration of corticosteroids and bronchodilators, either orally or via inhalation (4). Severe equine asthma is not a curable disease; nevertheless, allergen avoidance and corticosteroid therapy can partially reverse airway remodeling (17).

b. Impact of equine asthma

Severe equine asthma has a significant impact on horse welfare. It is a lifelong and often progressive condition. Alongside the persistent cough, horses may experience acute episodes of dyspnea. As the disease progresses, the increased effort required for breathing leads to heightened energy demands, exhaustion and subsequent weight loss (18). Unfortunately, disease management can sometimes negatively affect the overall well-being of the horse. When alternative hay options are limited, owners may resort to feeding hay pellets, which contradicts the natural feeding behavior of horses, which typically involves grazing for more than 12 hours a day. This can cause significant stress and lead to the development of stereotypies (19). Pharmacological treatment also carries certain risks. Although there is currently no conclusive evidence linking corticosteroid therapy to laminitis, anecdotal reports suggest a potential association (20). Furthermore, prolonged use of β2 agonists for bronchodilation may cause cardiovascular remodeling (21). The progressive nature of the clinical signs, coupled with the time and cost required for dust reduction measures, sometimes leads to early euthanasia.

Studies from the 80s demonstrated that lower airway inflammation was a common source of wastage in the equine racing industry (22,23). More recently, MEA has been shown to significantly reduce racing speed (3). Equine asthma can also hinder the participation of sport horses in competitions, given that many pharmacological treatments are classified as controlled substances. The high prevalence of EA and its established effect on performance suggest a

substantial impact on the equine industry, even though the exact economic burden has not been quantified.

c. A model of human asthma

In 2019, asthma affected an estimated 262 million people worldwide, with an alarming rate of 1,000 daily asthma-related deaths (1). While improving access to medical care and reducing exposure to risk factors are crucial for alleviating the disease burden, continuous research efforts are necessary to improve disease detection, characterization, and the development of more effective therapies.

Available animal models to study human asthma include mice, rats, guinea pigs, cats, dogs, sheep, horses, and nonhuman primates. Among them, only cats and horses are capable of developing asthma naturally (24). Murine models are extensively employed in asthma research due to their accessibility, cost-effectiveness, high level of standardization and ease of gene manipulation. Murine-specific reagents are readily accessible, and existing research data using murine models facilitate comparisons and the building upon previous findings. However, the induction of lower airway inflammation in mice necessitates antigen sensitization protocols, raising concerns about its relevance to naturally occurring disease in humans. Additionally, mice and human lungs differ significantly in terms of anatomy and physiology, and the requirement for terminal studies to collect pulmonary samples raise ethical considerations (24).

The translation of murine results to humans yielded disappointing outcomes so far (25). Therefore, it is advocated that findings from murine models should be initially investigated in naturally occurring disease models to assess treatment efficacy in real-life situations before application to humans (24). This approach allows for the study of asthma in all its heterogeneity and complexity. Among potential animal models, horses show promise due to their similarities to humans in terms of lung physiology and anatomy. Specifically, horses with SEA exhibit striking similarity to humans with severe neutrophilic asthma (24). Additionally, large volumes of biological samples can be collected from horses, and airway samples such as BALF or bronchial biopsies can be obtained from standing horses with light sedation, avoiding the need for general anesthesia or terminal studies. Furthermore, the long lifespan of horses, sometimes exceeding 25 years, enables the study of disease progression, including airway remodeling. Nonetheless, the use of horses as an experimental model is constrained by space requirements for housing and associated costs. Another potential limitation is the lack of standardization, as horses have diverse genetic backgrounds and are exposed to various aerosols. Equine-specific

reagents are limited in availability, and there is a smaller body of equine studies, resulting in less extensive knowledge about this species compared to rodent models (24).

In summary, SEA is a prevalent disease with a significant impact on animal welfare. The diagnosis of SEA is challenging, relying on the evaluation of multiple criteria. Treatment options are primarily palliative, often accompanied by side effects and the potential for positive doping tests. Therefore, the discovery of new biomarkers and therapeutic targets is of great importance. Furthermore, gaining a better understanding of the pathophysiology of SEA could provide valuable insights into the mechanisms of some forms of human asthma.

1.1.2 Immunopathogenesis – current knowledge

Equine asthma is triggered by an exaggerated immune response to respirable particulates found in the environment and particularly in hay dust. These entails organic and inorganic elements such as endotoxins, fungal elements, mites or ammonia (4). Equine asthma is a disease of domestication associated with hay feeding and stabling. A less common form of the disease, summer pasture EA, occurs when horses react to pollen. Equine asthma can be broadly classified as an allergy, adhering to the definition of a disease resulting from the immune system's reaction to an otherwise harmless antigen. However, it is uncertain whether EA represents an allergy *sensu stricto*, namely a type I hypersensitivity with Th2-mediated production of IgE.

The convergence of the innate and adaptive immune responses gives rise to three major types of cell-mediated effector immunity, known as Th1, Th2 and Th17-mediated diseases. Because these immune pathways do not rely exclusively on T helper (Th) cells, it has been proposed to label them type 1, type 2, and type 3 responses instead (26). Unfortunately, this terminology has not been widely adopted, and the terms have been used interchangeably. Type 1 immunity involves the activation of specific immune cells like immune lymphoid cells 1 (ILC1s), natural killer (NK) cells, TC1 cytotoxic T cells, and $CD4^+$ Th1 cells. These cells produce interferon γ (IFNγ) and work together to protect against intracellular microbes by activating mononuclear phagocytes. Type 2 immunity, on the other hand, relies on the activation of ILC2s, TC2 cells, and Th2 cells, which produce interleukin (IL) 4, IL5, and IL13. These substances trigger the activation of mast cells, basophils, eosinophils, and the production of IgE antibodies, providing defense against helminths and venoms. Type 3 immunity is mediated by ILC3s, TC17 cells, and Th17 cells, which produce IL17, IL22, or both. These

substances activate mononuclear phagocytes, recruit neutrophils, and induce epithelial antimicrobial responses, offering protection against extracellular bacteria and fungi. Exacerbated type 1 and type 3 immunity can contribute to autoimmune diseases, while uncontrolled type 2 responses may lead to allergic diseases (*sensu stricto*) (26). To further elevate the level of complexity, it is important to acknowledge the high plasticity of T cells. Their phenotype is not fixed and can undergo changes. For instance, Th17 cells have the ability to transition into Th1 cells. As a result, overlapping inflammatory states may emerge, characterized by transient immunological signatures (27).

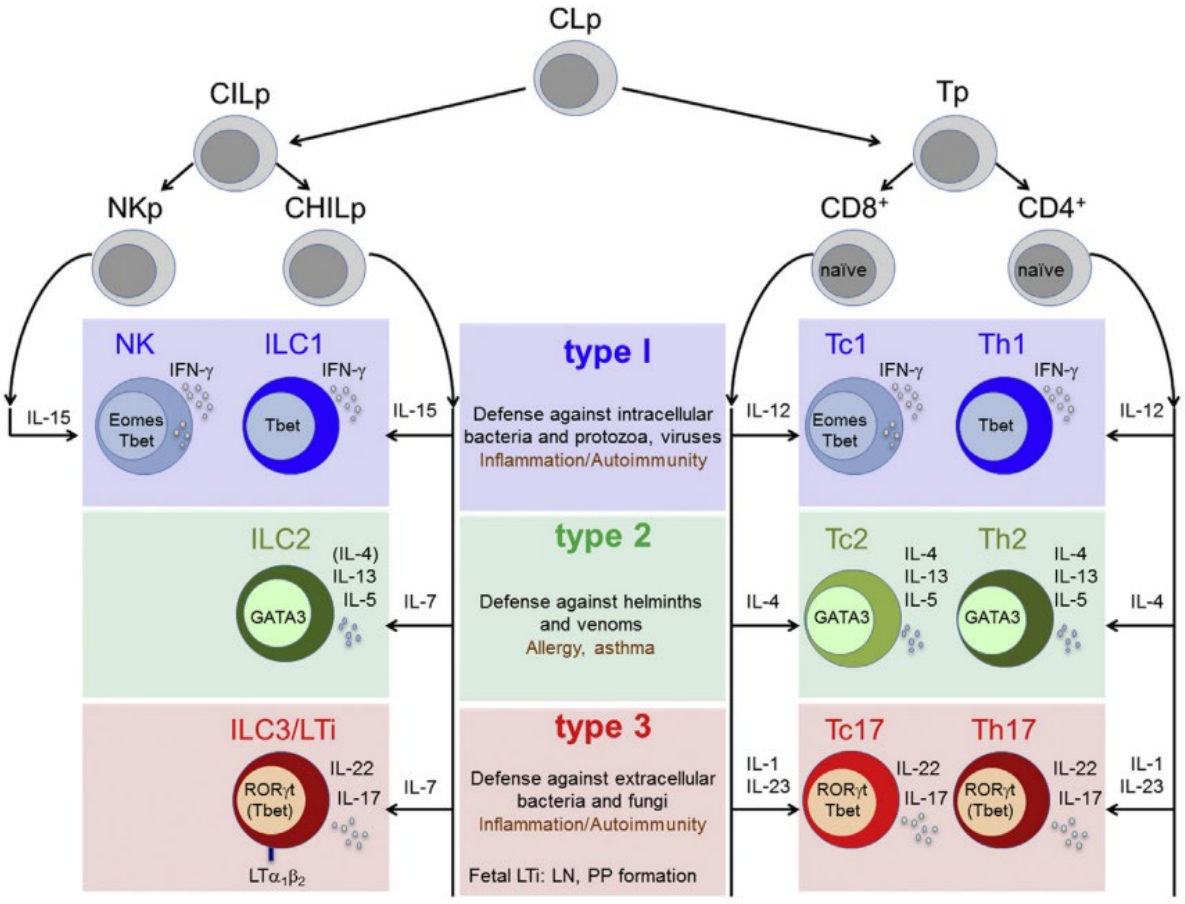


FIGURE 1. The 3 major types of innate and adaptive cell-mediated effector immunity. Type 1 immunity is composed of T-bet$^+$ IFNγ–producing CD4$^+$Th1 cells and ILC1s and T-bet$^+$Eomes$^+$CD8$^+$ TC1 and NK cells. Type 2 immunity is composed of GATA3$^+$CD4$^+$ Th2 cells, CD8$^+$ TC2 cells, and ILC2s, which produce IL4, IL5, and IL13. Type 3 immunity is composed of RORγt (RORC)$^+$CD4$^+$Th17 cells, CD8$^+$ TC17 cells, and ILC3s, producing IL17, IL22, or both. *CILp, Common innate lymphoid precursor; CLp, common lymphoid precursor; LN, lymph node; LTi, lymphoid tissue inducer; PP, Peyer patch; Tp, T-cell progenitor*. From Annunziato et al. 2015 (26)).

Studies investigating the predominant immune pathway in equine asthma have yielded conflicting results. The disease has been attributed alternatively to a type 2, type 1, type 3, or mixed immune response based on mRNA or cytokine analysis of peripheral blood or BALF from affected horses (Table 1) (28). Traditionally, equine asthma has been considered a type 2 disease, and this hypothesis is supported by the observation that affected horses may also suffer from other hypersensitivities such as insect bite hypersensitivity or urticaria (29,30).

Recent studies have increasingly implicated the Th17 pathway in equine asthma. Elevated levels of *IL17* mRNA have been detected in the BALF of horses with equine asthma following antigen exposure (11). Analyses of mediastinal lymph nodes have provided evidence of a predominant Th17 response in SEA (31). Dysregulation of miRNA in the serum of asthmatic horses has further supported the existence of a mixed Th2/Th17 response (32). Additionally, a comprehensive miRNA-mRNA study in equine lung tissues suggested a predominant Th17 pathway, along with some indications of a parallel Th2-type response (33). An increased expression of the Th17-associated CXCL13 chemokine has been found in stimulated peripheral blood mononuclear cells (PBMCs) from SEA horses. A recent flow cytometry study on equine BALF cells clearly identified a local type 3 response in SEA, with an increased frequency of $CD4^+IL17A^+$ lymphocytes. The IL17 family, particularly IL17A and IL17F, play a crucial role in the type 3 response, contributing to the recruitment and activation of neutrophils (26).

TABLE 1. Cytokines reported in SEA-affected horses according to T helper subtype (32–41). From Simões et al. 2022 (28)

Th2	Th17	Th1/Th2	Th1/Th17	Th2/Th17	Undefined
↑ IL-4 [1(r)]	↑ CXCL13 [2(r)]	↑ IL-4 [1(r)]	↑ IL-1β [1(r)]	↓ miR-197 [2(r)]	↓ IFN-γ [1(r)]
↑ IL-5 [1(r)]		↑ IFN-γ [1(r)]	↑ IL-8 [1(r); 3(r); 3(p)]	↑ miR-744 [2(r)]	↓ IL-4 [1(r)]
↓ IFN-γ [1(r)]			↑ IFN-γ [1(r)]	↓ miR-26a [4(r)]	↓ IL-5 [1(r)]
			↑ TNF-α [1(r)]	↑ miR-31 [4(r)]	↓ IL-13 [1(r)]
			↑ IL-17 [1(r)]	↓ TNF-α [4(r)]	
				↑ IL-4R [4(r)]	

[1]—BALF recovered cells; [2]—Peripheral blood; [3]—bronchial epithelial biopsy; [4]—Lung tissue (post-mortem). r—RNA detected; p—protein detected. ↑—increased expression; ↓—decreased expression. IL—interleukin; IL-4R—interleukin 4 receptor; IFN-γ—gamma-interferon; TNF-α—tumor necrosis factor-α; miR—microRNA.

Airway neutrophilia is a hallmark of SEA. Neutrophils play an important role in eliminating pathogens from the lungs through various mechanisms, including phagocytosis, production of cytokines, chemokines, proteases, reactive oxygen species (ROS), and neutrophil extracellular traps (NETs) formation. Their prompt elimination through apoptosis is essential

to prevent excessive inflammation and tissue damage. NETs, a specific form of apoptosis, involves the release of cellular DNA along with antimicrobial peptides and proteases. While their primary function is pathogen elimination, they can also induce tissue damage and sustain chronic inflammation. NET formation is enhanced in the BALF from horses with SEA (42,43). Therefore, neutrophils could participate to lower airway inflammation in SEA through increased recruitment, altered function and/or persistence in the lungs. The precise role of neutrophils in the immunopathogenesis of equine asthma is currently unclear, and it remains to be determined if they are active participants or simply a consequence of the local inflammatory response.

The reason for the development of SEA in some horses and not others, despite being exposed to the same environment, remains incompletely understood. Individual susceptibility appears to be influenced, at least in part, by genetic factors. Genetic predisposition has been observed in Swiss Warmblood horses with SEA, and the mode of inheritance may vary among families within the same breed (44). Equine PBMCs have shown genetic heterogeneity in their response to allergens, further supporting the role of genetic susceptibility in the disease (45).

Despite several decades of research on EA, our comprehension of its immunopathogenesis remains limited. In the following section, we will explore the possible causes behind the knowledge gap to determine the optimal strategy for addressing and closing this gap.

1.1.3 Origin of the knowledge gap

The discrepant findings observed in EA studies conducted over the past decades reflect the intricate nature of the disease, influenced by variations in study design, study population, and limitations of experimental techniques.

Disease heterogeneity

Asthma is a highly complex disease encompassing various triggers, risk factors, clinical presentations, pathological characteristics and treatment responses. In humans, asthma has been classified into different phenotypes, based on clinical manifestation and response to therapy (27). Similarly, EA has been divided into two main forms: MEA and SEA. However, there is growing recognition of the need to consider the disease in terms of endotypes, which account better for the heterogeneity encountered in clinical practice. An endotype describes distinct

pathophysiologic mechanisms at the cellular and molecular level (27), which may or may not align with the clinical phenotype. In human asthma, two broad categories have been identified: type 2 (or T2-high) and non-type 2 (or T2-low) asthma (27), with each category comprising several endotypes (27). Type 2 asthma, often referred to as allergic asthma, is characterized by airway eosinophilia. In contrast, non-type 2 asthma can be either neutrophilic or paucigranulocytic, similar to SEA. Some subsets of human non-type 2 asthma exhibit similarities to SEA (5), such as very late-onset asthma, which is associated with aging, and organic-dust induced asthma of agricultural workers exposed to high dust levels (46). The inclusion of different EA endotypes within the same study has likely contributed to the inconsistency of results observed thus far. To enhance result interpretation and improve clinical applicability, research efforts should focus on studying one endotype at a time. Currently, we propose using the existing categorization of MEA and SEA, further subclassifying them based on the predominant cell type in the BALF. These endotypes could include neutrophilic SEA, paucigranulocytic SEA, neutrophilic MEA, mastocytic MEA, eosinophilic MEA, or mixed-cell-type MEA. As knowledge advances, additional endotypes or more refined subtypes may be identified.

Study population design

The diversity of the study population is a potential factor contributing to the challenges in establishing conclusive immunological pathways for SEA. There is considerable variation in the inclusion and exclusion criteria used to select asthmatic horses across different studies. Horses may be in different phases of the disease, ranging from acute exacerbation to remission. Moreover, the inclusion of horses from genetically diverse backgrounds, often from different breeds, adds complexity. Maintaining a herd of asthmatic horses for research purposes necessitates significant expenses and adequate space, often leading to the recruitment of privately-owned horses. This comes with limited control over various factors, such as aerosol exposure, feeding and bedding type, exercise or medications. Collecting BALF may not always be well-accepted, particularly by trainers, due to perceived invasiveness or concerns about its impact on performance (9). Obtaining BALF samples from control horses without respiratory symptoms is even more challenging. Additionally, a common limitation of equine studies is their small sample size, which limits statistical power and interpretation of results.

The study design itself can significantly influence the outcomes. Equine asthma can be investigated without any intervention, or horses may undergo antigen challenges with different types or doses of antigens. The timing between sample collections may vary, and experiments

can be performed ex vivo or in vitro, using stimulated or unstimulated cells. Again, the type and dose of stimulation may not be consistent across studies.

The choice of materials and methods employed in the research also has a substantial impact on the findings. While blood is easily accessible in horses, it primarily reflects systemic inflammatory responses rather than the local pulmonary environment. Tracheal wash, BALF, bronchial biopsies, or lung biopsies each contain unique cell types and cannot be directly compared. Fluid samples like tracheal wash or BALF present particular challenges during the analysis due to the difficulty in accounting for dilution effects. Furthermore, the experimental techniques used to analyze these samples have inherent limitations. Cytological examination of BALF, although widely used, is inherently subjective and allows for the differentiation of only five distinct cell populations (macrophages, lymphocytes, neutrophils, mast cells, and eosinophils) (47). Antibody-based techniques like immunohistochemistry and flow cytometry can differentiate additional subpopulations within lymphocytes, but their applicability in horses is limited by the availability of validated antibodies (48,49). Investigating the regulation of specific genes can be done by analyzing individual messenger RNA transcripts in BALF (43,50–52), primarily through reverse transcription polymerase chain reaction (RT-PCR). However, this approach is hypothesis-driven, with limited throughput and potential investigator bias. In contrast, global transcriptomics is a high-throughput technique that provides an unbiased approach to studying gene expression. Recent application of transcriptomics to equine PBMCs revealed impaired cell cycle regulation in SEA and upregulation of the Th17-associated gene *CXCL13*, a potential therapeutic target (34). Nevertheless, when performing bulk RNA sequencing (RNA-seq) of mixed cell populations, critical differences between individual cells are obscured. This is an important consideration when performing RNA-seq of BALF, especially due to the substantial impact of EA on the cellular composition of the lower respiratory tract.

In contrast, single-cell mRNA sequencing (scRNA-seq) provides the ability to analyze different cell populations in a sample and capture their individual transcriptomes simultaneously. This method eliminates the need for equine-specific reagents and reduces the reliance on prior knowledge of marker genes for cell type identification. In the subsequent sections, we will present an overview of this innovative technique, including its principles of operation, strengths, and limitations.

1.2 Single-cell mRNA sequencing

Since its introduction in 2009 (53), the application of scRNA-seq has gained widespread popularity owing to the greater accessibility of commercial platforms, the reduced sequencing cost, and the development of dedicated software tools that simplify data analysis. Presently, scRNA-seq has become a readily accessible technique for investigating physiology and diseases in both humans and animal models.

1.2.1 A practical guide

The methods developed for scRNA-seq are built upon a shared underlying principle, which involves tagging transcripts with a cellular identifier prior to sequencing. There are two main approaches: plate-based techniques and microfluidic-based techniques. In plate-based methods, cells are captured on multi-well plates or microfuge tubes before fluorescence-activated cell sorting (FACS). However, plate-based methods suffer from a relatively low cell throughput compared to microfluidics. Microfluidic-based techniques involve capturing single cells within microfluidic droplets. These techniques can operate even with a low starting cell number, are cost-efficient, and enable simultaneous analysis of gene expression profiles in a highly parallel manner (54).

In this research project (studies 1 and 3), we used the single-cell Chromium Controller from 10X Genomics, which is currently the most commonly used microfluidics-based platform. Our focus will therefore be on the protocol specific to the 10X Genomics Chromium Single Cell Solution, providing an illustrative example of a typical scRNA-seq experimental workflow.

a. Experimental workflow overview

Figure 2 gives an overview of the experimental workflow using the 10X Genomics Chromium Single Cell Gene Expression Solution. A single-cell suspension is prepared, either from a liquid biological sample or from tissue. The solution is loaded onto the Chromium chip, which is placed in the Chromium Controller for cDNA library creation. In this step, each cell is isolated in a droplet containing a specific barcode and the necessary reagents for cell membrane lysis, reverse transcription, and cDNA amplification. The resulting cell-specific cDNA library is then indexed to identify the source experiment. Multiple libraries can be pooled and sequenced simultaneously. After demultiplexing, data undergo quality control before being aligned to a reference genome. The cell-specific transcripts are further processed to improve

signal-to-noise ratio before data visualization and interpretation. Each step will be subsequently detailed (55).

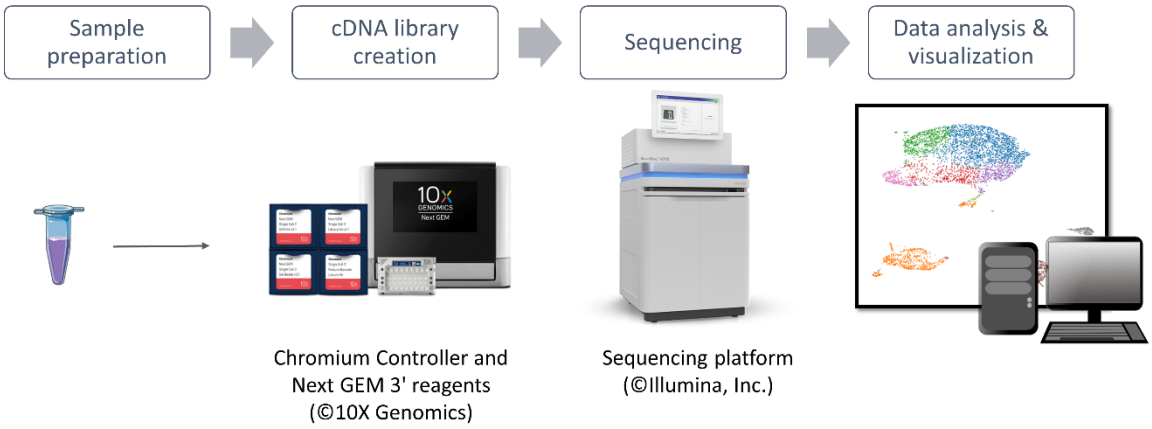


FIGURE 2: Standard experimental workflow for single-cell mRNA sequencing using the 10X Genomics Chromium Single Cell Gene Expression Solution

b. Sample preparation

The objective of sample preparation is to generate a clean suspension of viable and healthy single cells. This can be achieved by processing liquid biological samples like blood or BALF, or by extracting cells from tissue. Pure cell lines or mixed-cell samples can be used. When isolating cells from tissue, mechanical or enzymatic dissociation methods are employed. In experiments focusing on a specific cell population, cell sorting techniques such as FACS-sorting can be applied (55).

Sterile handling and the use of nuclease-free reagents and consumables are recommended. Sufficient volumes should be used for cell washing and resuspension to prevent aggregation and clumping. Cell suspensions should be filtered using appropriately sized cell strainers, with pore sizes larger than the cell diameter, to remove clumps and debris. Gentle pipetting is essential to preserve cell viability. Centrifugation steps should also be kept to a minimum. To minimize cellular aggregates, dead cells, and contaminants such as reverse-transcription (RT) inhibitors, protocol optimization may be necessary. This entails adjusting the number of wash steps, the composition of the resuspension media, or the centrifugation parameters. The final single-cell preparation should ideally have a viability greater than 90% and a concentration between 0.7-1.2 x 10^6 cells/mL to optimize cell encapsulation. Cell

counting can be performed manually or with an automated counter. Some cell counters use electrical impedance to simultaneously assess cell membrane integrity. Alternatively, cells may be stained and visualized with microscopy or flow cytometry to determine their viability. Minimizing processing time is important to preserve cell viability and limit RNA degradation (56).

ScRNA-seq can be performed on either fresh or preserved cells. Popular preservation methods are cryopreservation and methanol fixation. Preservation protocols are generally tailored to specific samples or cell types. Therefore, thorough testing and optimization of the protocol are necessary to ensure the preservation maintains the desired cellular composition and viability after storage.

c. Construction of cell-specific cDNA libraries

The 10X Genomics technology is based on microfluidic droplet chemistry with co-encapsulation of a cell and a barcoding bead in a droplet. The sequence for cDNA library preparation is depicted in figure 3.

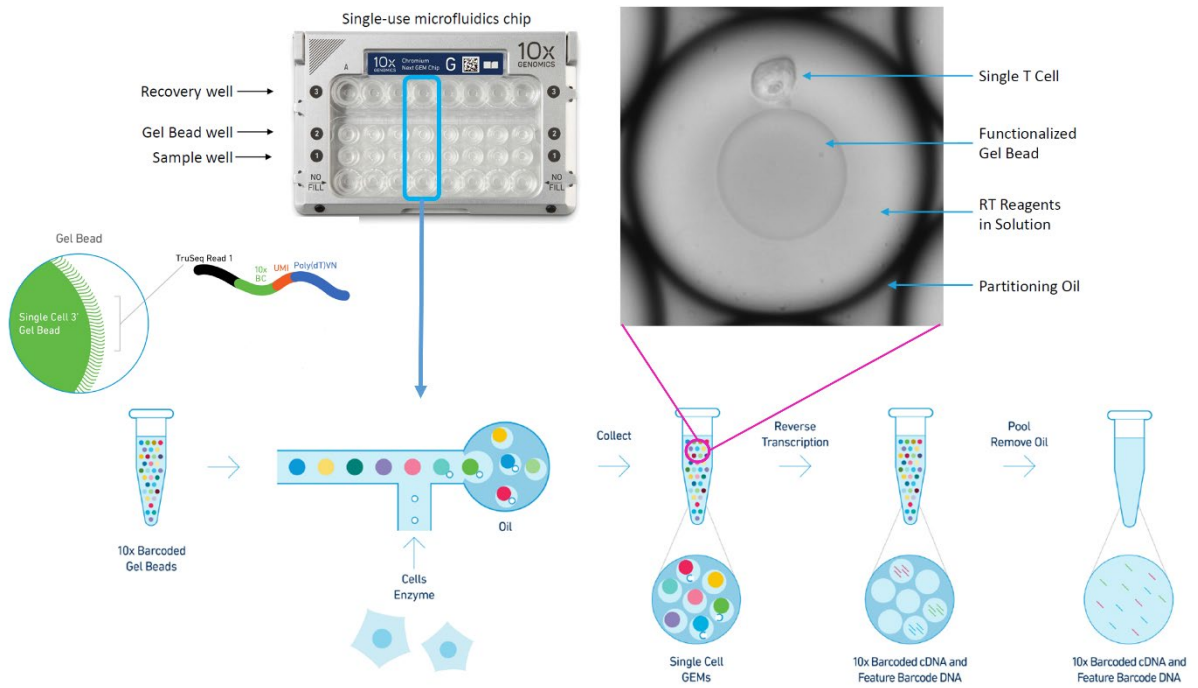


FIGURE 3: Workflow overview for cDNA library preparation with 10X Genomics Chromium Single Cell Gene Expression Solution. *GEM, Gel Beads-in-emulsion; RT, Reverse Transcription* (adapted from © 10X Genomics, Inc. 2019)

On the Chromium chip, one lane of wells is dedicated to the single-cell preparation. The two remaining lanes of wells contain the barcoded beads and an oil suspension, respectively. The barcoded beads consist of synthetic beads coated with oligonucleotides comprising the following regions (57):

- A linker region connecting the oligonucleotide to the bead.
- A primer region (TruSeq Read1) for subsequent cDNA amplification.
- A cell barcode: a nucleotide sequence that is common to all oligonucleotides on a specific bead but unique to each bead. This barcode traces back the cellular origin of the transcript.
- A unique molecular identifier (UMI): a nucleotide sequence that is unique to each oligonucleotide on the bead. The UMI informs about the amplification level, ensuring that amplicons of the same read are only counted once.
- A poly-d(T) region, which captures the polyadenylated RNAs released within the droplet. This enables the capture of polyadenylated transcripts, including most eukaryotic mRNAs, as well as certain long noncoding RNAs and antisense transcripts.

After being filled with the single-cell suspension, the Chromium chip is loaded onto the Chromium Controller, where cells are individually encapsulated into water-in-oil droplets containing a barcoded bead. The encapsulation efficiency, ranging from 50% to 65% according to the manufacturer, depends on the cellular concentration of the input solution. Low cell concentration results in a decreased number of encapsulated cells and lower final reads and cell counts for analysis, while higher concentrations correlate with an increased frequency of doublet or multiplet formation. Multiplets represent the co-encapsulation of multiple cells within a single droplet. The final cell concentration is optimized to maintain multiplet rate under 5% (57). Following encapsulation, the cell membrane is lysed, and the poly(A) tail of the released mRNA is captured by the poly-d(T) region of the bead's oligonucleotides. Reverse transcription generates the first strand of cDNA, followed by the synthesis of the second strand. Each droplet now virtually contains a cDNA library with a unique cell barcode. Because each cell contains a small amount of mRNA, cDNA must be amplified prior to sequencing. Droplets are broken down before cDNA amplification with PCR. The amplified cDNA then undergoes enzymatic fragmentation before the ligation of sample indexes and sequencing adapters in

preparation for the sequencing step. The resulting cDNA library comprises the following elements, depicted in figure 4:

- TruSeq Read 1 (28 nucleotides): sequence containing the cell barcode and the UMI. This read provides information about the cellular origin and amplification level of the read.
- TruSeq Read 2 (91 nucleotides): read of interest, which will be aligned to a reference genome or transcriptome.
- Sample index: random oligonucleotide sequence specific to the experiment, allowing for multiplexing of multiple cDNA libraries in a single sequencing run.
- P5: forward primer for Illumina sequencing
- P7: reverse primer for Illumina sequencing

For our research project, we utilized the 3'-end short read sequencing approach, where only the 3' extremity of the captured mRNA is subjected to RT. Alternative protocols enable sequencing of long reads or 5'-end short reads.

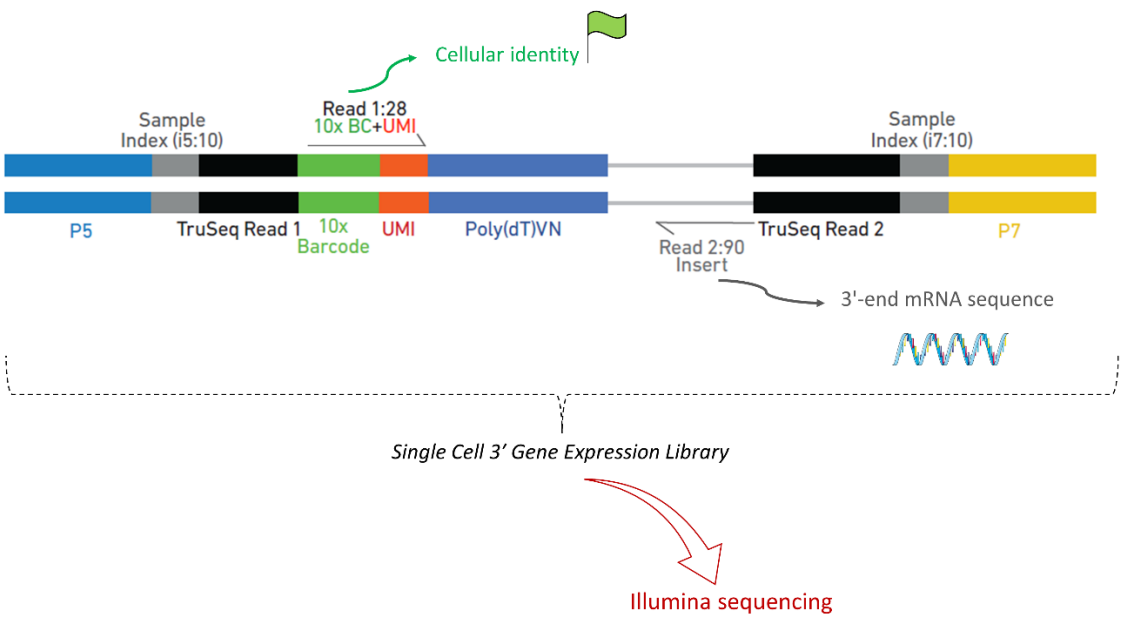


FIGURE 4: Representative 10X Genomics Chromium Single Cell 3' dual index library schematic (adapted from © 10X Genomics, Inc. 2019)

d. Illumina sequencing

The indexed cDNA library can be stored for later sequencing or immediately sequenced. In our research project, we used Illumina sequencing, a widely adopted technique. Illumina sequencing relies on the sequencing-by-synthesis approach, where the target strand is replicated using DNA polymerases and fluorescently-labeled nucleotides. Fluorescence signals are measured at each cycle to determine the incorporated nucleotide. The sequencing depth correlates with the number of reads per cell. To analyze a larger number of cells while minimizing sequencing costs, it is common to limit the number of cycles. However, it is important to ensure that the sequencing reaches a sufficient depth to accurately identify most of the transcripts present in the sample. For optimal results, 10X Genomics recommends using 150 cycle kits (57).

e. Data analysis

The range of computational tools for scRNA-seq data analysis continues to expand. Some commercial scRNA-seq platforms provide integrated software, such as the Loupe Browser from 10X Genomics, to aid researchers without bioinformatics expertise. However, these tools often operate as black boxes with limited transparency regarding the underlying algorithms. The preferred approach for scRNA-seq data analysis is through programming, typically utilizing the R or Python language. Because scRNA-seq data analysis still is in its infancy, there are no standardized pipelines. Each tool has its own advantages and limitations, and the selection should be guided by expert knowledge based on the specific application and research question (55). In this thesis project, we developed a custom pipeline (Figure 5) using a curated set of R packages.

Data preprocessing (54,58)

The raw data is acquired through the sequencing of the pooled libraries. The reads are sorted into separate FASTQ files, based on their corresponding libraries. This sorting process generates three distinct types of FASTQ files for each library: one containing gene IDs (quantified genes), another containing cell barcode data (quantified cells), and a count matrix.

- **Mapping and alignment**

Reads are sorted based on their cell barcodes and aligned to a reference assembly with gene annotation.

- **Feature annotation and quantification**

A count matrix is generated, in which each column corresponds to a cell barcode and each row represents a gene. The entries in the matrix indicate the number of reads that contains the corresponding cell barcode and map to the specific gene. Reads carrying the same UMI are collapsed into a single read.

- **Quality control and filtering**

Cells are subjected to filtering based on their RNA count (feature count) and the proportion of mitochondrial reads. The determination of filtering thresholds usually involves visual examination of the data. Cells with an exceptionally high RNA count are likely to be doublets or multiplets, while a low RNA count may indicate lysed cells or the encapsulation of ambient RNA. A high fraction of mitochondrial reads is indicative of dead or damaged cells, in which the leakage of cytosolic RNA leaves only mitochondrial transcripts confined within intact double-membrane mitochondria.

Downstream analysis (54,58)

- **Normalization**

Following the removal of dead cells and multiplets, the data undergoes normalization to address amplification and count depth biases. Due to variations introduced during the experimental workflow, such as differences in capture and reverse transcription efficiency, the RNA counts encompass both biological and technical variation. Normalization mitigates the technical variation while preserving biological variation. Following normalization, the gene expression levels should not correlate with the sequencing depth of the cell.

- **Data correction**

Data correction aims to address potential confounding factors such as batch effect, dropout or other biological effects (see “Challenges” section below). For example, genes associated with the cell cycle phase can be regressed out to account for the cell cycle effect. It is important to consider that not all factors need to be corrected, as doing so may eliminate important biological information. Instead, the decision of which factor to consider depends on the specific objectives of the downstream analysis. If a batch effect is suspected, linear methods can be applied for batch effect correction. Integration is typically performed when dealing with different

conditions to ensure that the cell types of one condition align with the corresponding cell types in the other condition (e.g. control neutrophils align with asthmatic neutrophils).

- **Feature selection**

Highly variable genes, which contribute most to the data variability, are selected. These genes are considered biologically meaningful, and their selection helps increase the signal-to-noise ratio and reduce computational complexity. The number of highly variable genes retained for downstream analysis typically ranges between 1,000 and 5,000, depending on the dataset and objectives.

- **Dimensionality reduction**

ScRNA-seq data is characterized by a high-dimensional space defined by the number of cells and genes, often referred to as the 'curse of dimensionality'. While high-dimensional data theoretically contains more information, in practice, most of the data is redundant and adds noise, obscuring the relevant signals. Dimensionality reduction techniques are employed to reduce noise, facilitate computational analysis, and aid data visualization. These algorithms aim to capture the underlying structure of the data in as few dimensions as possible. Principal component analysis is the most commonly used method for dimensionality reduction.

- **Clustering**

After dimensionality reduction, cells are clustered based on the similarity of their gene expression patterns. Various algorithms can be used for cell clustering, with popular choices being *k*-means and the Louvain algorithms. The number of clusters obtained depends on the chosen clustering resolution, which is a subjective decision based on the researcher's requirements. For example, in some cases, it may be desirable to isolate all T cell subtypes, while in others, it may be appropriate to keep them grouped together (59).

- **Visualization**

The clustered data is visualized using machine-learning algorithms such as t-distributed stochastic neighbour embedding (t-SNE) or uniform manifold approximation and projection (UMAP). While t-SNE was commonly used in the early stages of scRNA-seq, UMAP is now preferred due to its computational efficiency and ability to preserve the global structure of the clusters.

- **Annotation** (60)

Cell cluster annotation can be achieved through manual or automated approaches, both relying on external sources of information.

Expert knowledge is paramount to manual annotation. One strategy involves visualizing expression values of canonical marker genes derived from the literature. Another strategy is to examine the marker genes extracted from the dataset. Marker genes are identified using differential expression testing between clusters. Common statistical tests such as the Wilcoxon rank-sum test or the t-test are used to rank genes by their difference in expression. The top-ranked genes are considered marker genes. Interpretation of the marker genes usually focuses on upregulated genes in the cluster of interest, although downregulated genes may sometimes prove valuable. Marker genes can be compared to reference sets, originating either from cell type-specific mRNA expression from bulk RNA-seq of FACS-sorted populations, or from previous scRNA-seq experiments.

Automated methods compare the expression pattern of the scRNA-seq dataset to a reference set. The general principle is to identify a gene expression pattern in a single cell or cell cluster that matches a characteristic gene expression signature of a known cell type or state. The cell or cluster is then assigned the corresponding label with an associated confidence score.

- **Differential Gene Expression (DGE) analysis**

Differential gene expression analysis aims to identify distinct expression profiles between two conditions, typically health and disease. Differential gene expression analysis estimates which genes are statistically more expressed in one group compared to the other within each cell type. Although the principle is similar to bulk DGE analysis, algorithms for single-cell data are specifically designed to handle artifacts inherent to scRNA-seq. Individual searches for the differentially expressed genes (DEGs) can be conducted in literature or databases that provide functional information, such as the Human Protein Atlas (61). Functional enrichment analysis can also be employed to identify the predominant pathways associated with the DEGs.

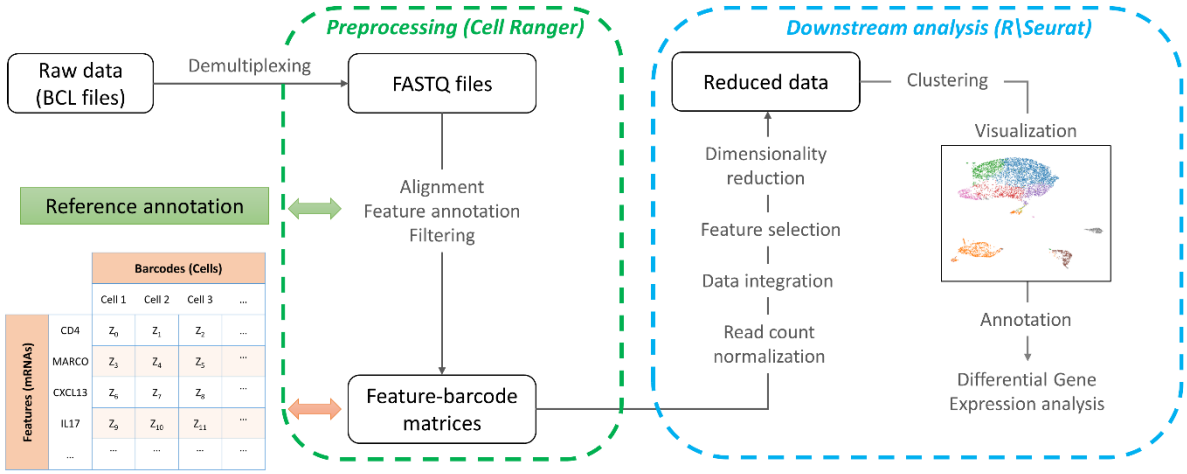


FIGURE 5: Data analysis workflow for equine bronchoalveolar cells from asthmatic and control horses (Study 3), illustrating the standard scRNA-seq data analysis sequence.

There are many additional approaches for exploring scRNA-seq data that cannot be exhaustively listed here. For example, trajectory inference, also known as pseudotime analysis, explores dynamic cellular processes such as cellular differentiation. Another captivating approach is the analysis of receptor-ligand interactions, which may identify potential therapeutic targets. The strengths of scRNA-seq analyses will be further discussed in the following section.

1.2.2 Strengths

a. A high-resolution analysis

A new definition of cell identity (62)

ScRNA-seq offers an unprecedented level of detail in analyzing biological samples. Unlike bulk transcriptomics, which averages signals from cells with unique phenotypes, scRNA-seq allows us to trace the cellular origin of each RNA molecule in a biological sample, redefining our understanding of what constitutes a cell. The definition of cell type relies on the combination of various characteristics, including lineage, differentiation stage, location in the body or organ, specific state (e.g., activation status), or current cell cycle phase. Originally, cells were characterized based on morphology, function, and expression of marker proteins, in particular cell surface proteins. Cells were perceived as distinct and well-defined entities due to their classification based on the presence or absence of specific cell surface markers (for

example, Treg cells are defined as $CD4^+FOXP3^+$ cells). However, scRNA-seq experiments have revealed that cell types are not rigid entities; they are transient in both time and space. Cells can undergo different differentiation pathways, become activated, or migrate to different parts of the body where their functions change. Consequently, cells exist in various states influenced by factors such as the genome, environmental stimuli, or spatial context. By employing scRNA-seq, it becomes feasible to elucidate the intricate complexity of cell types, portraying them as constituents of a continuous spectrum rather than discrete entities.

Preservation of the single-cell information (63)

ScRNA-seq technologies enable the profiling of gene expression in each individual cell within a mixed cell sample. This allows for the identification and accurate description of different cell populations and their specific transcriptomes. This aspect is particularly significant when studying SEA, as the increase in inflammatory cells in BALF can introduce bias in global expression comparisons between groups. Having access to the single-cell transcriptome provides a deeper understanding and facilitates the interpretation of expression changes. Averaging transcriptomic information in mixed cell samples has been shown to blur information, potentially leading to false interpretations (known as Simpson's paradox, see figure 6A). Studying the cytokine production capacity in specific FACS-sorted lymphocyte populations within BALF has proved to be superior to analyzing the overall mixed lymphocyte population when it comes to illustrating the impact of EA (16). Additionally, single-cell transcriptomics enable the differentiation between changes in cell composition and changes in gene expression. For instance, if IL17 mRNA levels are increased in a blood sample, bulk RNA analysis cannot determine whether it stems from upregulated IL17 gene expression or from an expansion of the IL17-expressing cell population (Figure 6B). Time series studies, commonly conducted to study disease progression or tissue development, yield a mixture of cells at different differentiation or activation states at each sampling point. Using bulk techniques, the time-averaged expression of a gene may mislead the interpretation of its regulation during development. In contrast, scRNA-seq provides a more accurate representation of the gene's dynamics (Figure 6C).

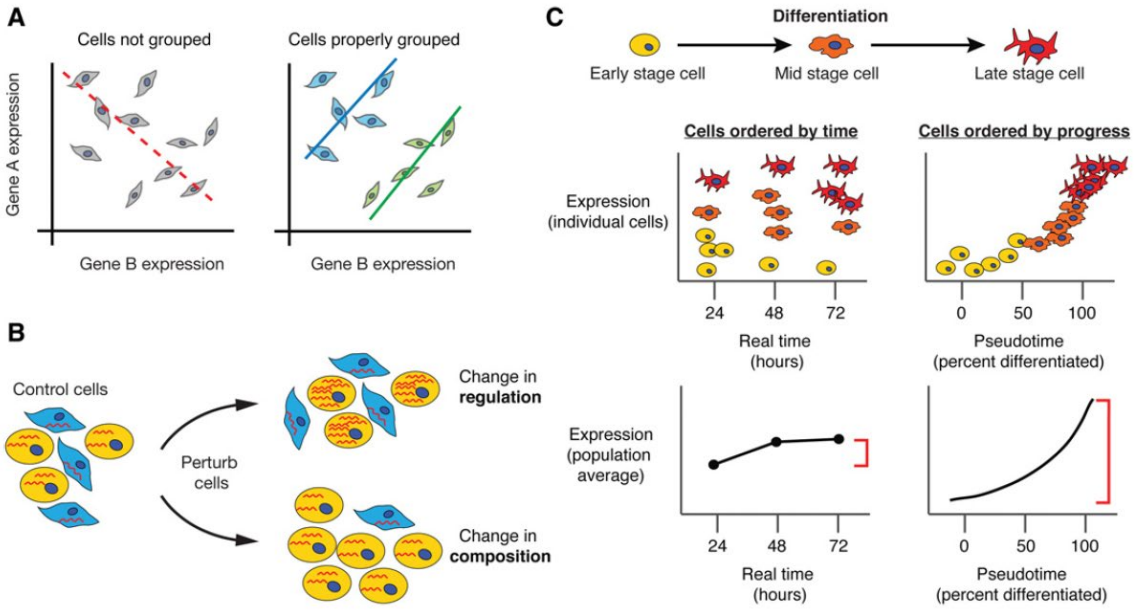


FIGURE 6. Single-cell measurements preserve crucial information that is lost by bulk genomics assays. (A) Simpson's Paradox describes the misleading effects that arise when averaging signals from multiple individuals. (B) Bulk measurements cannot distinguish changes due to gene regulation from those that arise due to shifts in the ratio of different cell types in a mixed sample. (C) Time series experiments are affected by averaging when cells proceed through a biological process in an unsynchronized manner. A single time point may contain cells from different stages in the process, obscuring the dynamics of relevant genes. Reordering the cells in "pseudotime" according to biological progress eliminates averaging and recovers the true signal in expression. From Trapnell et al. 2015 (63)

b. An unbiased approach

Before scRNA-seq became available, researchers could achieve single-cell resolution by sorting the cells they are interested in. Typically, cells are FACS-sorted based on their size and the expression of specific surface proteins. However, this approach is experimenter-biased and relies on preselected antibody panels. It hinders the discovery of new cell types or interesting cell states that may exhibit reduced or absent expression of typical surface proteins.

In contrast, scRNA-seq offers a hypothesis-free approach that enables exploratory experiments and the generation of novel hypotheses. By bypassing the dependence on surface markers, scRNA-seq uncovers previously unknown cell types and states. Notably, scRNA-seq is not constrained by species specificity, relying solely on the detection of poly(A)-tailed RNA. This aspect is particularly relevant in equine research, where the range of validated antibodies is limited (49).

c. Collaboration and data sharing

Numerous journals now mandate researchers to make their data openly available. For scRNA-seq, this entails depositing raw data or gene-cell count matrices in dedicated online repositories, along with posting associated computational pipelines in public databases. This requirement is crucial for upholding scientific integrity in data science, as it ensures reproducibility of results through the utilization of provided raw data and code (64).

Making scRNA-seq datasets freely accessible is of paramount importance due to the recognition that the value of a single dataset can be significantly enhanced when combined with other datasets (56). To facilitate data sharing, efforts are underway to consolidate scRNA-seq results and create species- and tissue-specific single-cell atlases. Noteworthy initiatives focused on compiling comprehensive data on organ and tissue structure and function include Tabula Muris and the Human Cell Atlas (HCA). Tabula Muris offers valuable insights into respiratory system cells for disease modeling studies, providing data from 100,000 cells across 20 mouse organs. The HCA project encompasses multiple organs and tissues from healthy humans, featuring a dedicated lung network that integrates various assay modalities to explore cellular function. During the COVID-19 pandemic, the HCA lung network swiftly analyzed combined datasets to investigate ACE2 expression and identify clinical factors associated with viral entry in different tissues, demonstrating its ability to generate high-quality data within a short timeframe (64). The development of a similar atlas specific to equine lung cells is highly desirable. Such an atlas could be constructed through scRNA-seq analysis of BALF, tracheal wash samples, bronchial biopsies, and lung tissue biopsies, providing a comprehensive molecular landscape of equine pulmonary cells.

d. Recent advances attributed to scRNA-seq

The diverse and expanding applications of scRNA-seq technology are revolutionizing our understanding of cellular heterogeneity and rare cell populations within complex samples. It has proven particularly effective in investigating unique cell types such as T cells with unique receptors (TCR), brain neurons, and cells from early-stage embryos (55). Moreover, scRNA-seq has reshaped our knowledge of cell lineage, developmental processes, aging, and various diseases, including cancer (56). Notably, it has significantly contributed to oncology research by accurately identifying clones, detecting gene expression changes, and providing insights into the tumor microenvironment and inflammatory cell composition. Recent studies in lung adenocarcinoma and breast cancer have demonstrated its potential in predicting chemotherapy

response, monitoring treatment progress, and detecting minimal residual disease, thus advancing precision medicine and personalized cancer treatment (65).

In the field of respiratory research, scRNA-seq analysis of the human lung has successfully identified 41 out of 45 previously known cell types, while also revealing 14 previously unidentified cell populations (66). Furthermore, scRNA-seq has led to the discovery of the pulmonary ionocyte and of a profibrotic macrophage population involved in pulmonary fibrosis. Such insights have significant implications for understanding disease mechanisms and may have practical implications for gene therapy approaches targeting cystic fibrosis, as ionocytes are a prominent source of the CFTR protein in the lungs (64).

Although scRNA-seq has only recently been introduced in asthma research, it has already produced noteworthy results, which are listed in figure 7. For instance, an analysis of respiratory samples from asthmatic patients and healthy controls using scRNA-seq revealed altered differentiation of epithelial cells in asthma, along with increased numbers of goblet cells, mucous ciliated cells, and Th2 cells (67). This study also indicated heightened cellular interactions involving Th2 cells, suggesting that altered communication between immune and structural cells may underlie airway inflammation in asthma.

By elucidating the unique expression patterns of inflammatory genes at the individual-cell level, scRNA-seq provides a valuable framework for comprehending the specific roles of immune cells and structural cells, as well as the significance of their interactions in asthma development. Similarly, scRNA-seq holds promise for advancing our understanding of SEA. Nonetheless, it is important to acknowledge that this technique, due to the extensive volume of complex data it generates, also presents certain limitations, which will be addressed in the subsequent section.

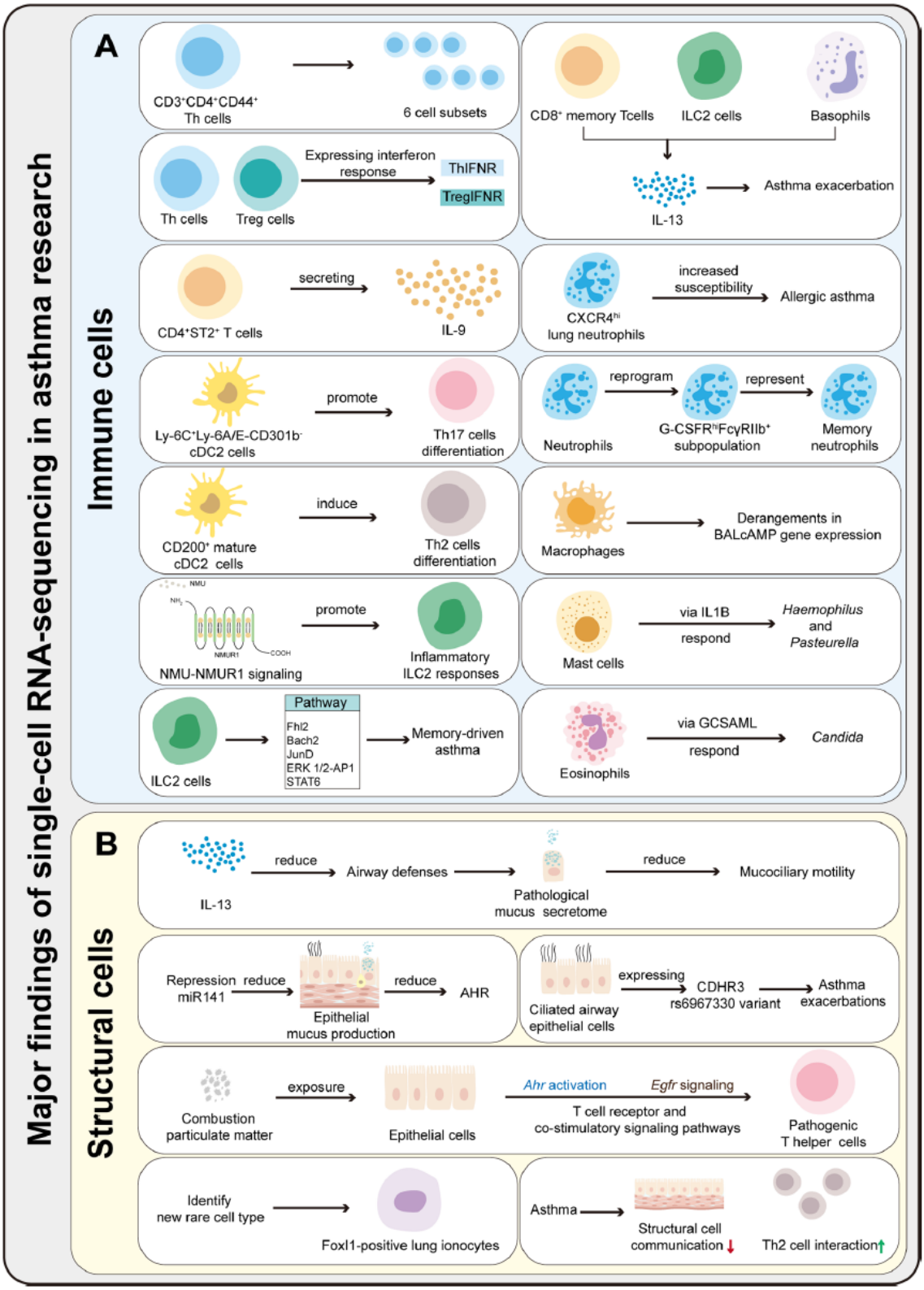


FIGURE 7 - Summary of the major findings resulting from scRNA-seq application in asthma research. (A) ScRNA-seq can be used to characterize the heterogeneity and function of immune cells, identify rare cells and new cell subsets, and explores immune cells differentiation mechanisms and pathways in asthma. (B) ScRNA-seq can be used to characterize the function of structural cells, identify rare cells, and explore the communication between immune and structural cells. From Tang et al. 2022 (68)

1.2.3 Challenges

As a recently developed technique, scRNA-seq represents pioneering work, necessitating resolution of unanswered questions that arise during experimental processes, data acquisition, and analysis. These unforeseen challenges demand additional optimization steps and process validation to overcome the inherent complexities. Firstly, accessibility to scRNA-seq remains limited. Secondly, researchers must be cautious in population selection and study design to avoid introducing biases. The complexity of the data also brings computational challenges. Lastly, data interpretation is currently impeded by the relative scarcity of preexisting scRNA-seq data and limitations of software tools.

a. Accessibility

The accessibility of scRNA-seq is constrained due to its cost, the need for specific infrastructure, and the expertise necessary for data acquisition and analysis. The cost of scRNA-seq encompasses expenses related to equipment, personnel, reagents, sequencing, and data analysis. Data analysis costs include expert bioinformatics work and access to high-performance computing (HPC) facilities. Despite the continuous decrease in sequencing costs, scRNA-seq remains an expensive technique.

ScRNA-seq requires specialized equipment and expertise in sample handling and sequencing-library preparation, necessitating trained personnel. HPC resources and substantial storage capacity are crucial for efficient data analysis. These resources demand regular maintenance and are predominantly available in major research centers and universities. Data analysis entails troubleshooting novel challenges, requiring significant expertise and surpassing basic programming skills. The involvement of informaticians with comprehensive knowledge of biological processes or close collaboration with biologists is essential for accurate data handling (65).

Despite the demanding cost and expertise prerequisites, we anticipate that scRNA-seq will become a standard technique in the forthcoming years. The market is rapidly evolving, with companies now offering researchers with limited resources the option to perform the entire scRNA-seq analysis process, from sample processing to data analysis. The development of user-friendly software tools for data analysis is expected to decrease reliance on expert bioinformatics work, although cooperation with bioinformaticians will probably remain indispensable.

b. Biological biases (55)

The experimental protocol of scRNA-seq involves a greater number of steps compared to bulk RNA analysis, resulting in an increased likelihood of introducing bias. Biases can arise from both biological and technical factors. At the biological level, biases can occur at three different levels: the sample (individual) level, the cell level, and the molecular (RNA) level.

Interindividual differences in the study population contribute to a major source of variability in the dataset. Fixed variables such as age, sex, and breed, as well as variable parameters like epigenetic changes, environmental stimuli, microbiome, and diseases, can introduce variability in the data. This becomes particularly relevant when studying non-conventional models like horses, which comprise a certain genetic heterogeneity and for which no genetically homogenous lines exist.

At the cell level, variability exists in terms of cell size, RNA content, cell cycle phase, and differentiation stage. Depending on the focus of the experiment, this variability can be considered either relevant biological variation or noise.

Molecular variability arises from the phenomenon of transcriptional bursting, also known as transcriptional pulsing, which is inherent to gene expression. Transcription occurs in stochastic "pulses," leading to potential differences in the transcriptome of samples collected or processed at different times from the same individual.

The presence of high biological variability can make it challenging to detect the biological signal of interest. To enhance the signal-to-noise ratio, efforts should be made to limit variability within the study population by selecting individuals with homogeneous characteristics. Additionally, the condition being studied should be well-defined within the population. This entails selecting individuals who exhibit strong characteristics associated with the respective condition.

Biases at the cellular and molecular levels cannot be entirely eliminated, as they represent natural processes with potential biological relevance. However, if desired, their effect can be regressed out during the data analysis phase.

c. Technical biases

Technical biases can be introduced at many steps of the process: during sample collection, sample processing, sample storage, scRNA-seq library preparation and sequencing. Potential technical artifacts include cell bursting leading to RNA leakage, selective RNA degradation, formation of multiplets, and batch effects (54).

Cell and RNA loss

During sample collection and processing, certain cell types may be more susceptible to alterations than others, leading to an enrichment of specific cell populations. For example, centrifugation steps may favor the recovery of larger, heavier cells over smaller, lighter cells. Tissue digestion methods may also selectively recover less delicate cell types, as cell rupture can lead to RNA degradation and lower RNA recovery (65). Additionally, cell types with high RNAse content, like granulocytes, pose challenges for RNA recovery and may be underrepresented in scRNA-seq datasets. Gentle processing techniques and automated tissue processing can help limit technical variability and minimize changes in cell composition (65). The addition of RNAse inhibitors in cell suspensions may improve RNA recovery from RNAse-rich cells, although this has not been critically assessed. It is essential to thoroughly document sample handling methods in publications to facilitate comparisons between studies and enhance consistency (64).

Although scRNA-seq methods were initially designed for freshly isolated cells, immediate sample processing can be challenging due to infrastructure limitations or the unavailability of specialized equipment. Preservation of samples offers a solution by disconnecting the sampling time and location from downstream processing steps (56). However, sample preservation may introduce technical biases, with up to 20% RNA loss per freeze-thaw cycle (69). Despite this, studies have shown that the impact of freezing on single-cell transcriptomic profiles is generally minimal (70). Nevertheless, certain cell types, particularly T cells, may be more affected by freezing, leading to a decrease in their population and increased expression of stress-related genes over time (71,72).

The bias introduced by preservation is considered minimal compared to the potential major biases associated with processing in multiple batches or delays between collection and analysis. When analyzing scRNA-seq data, direct comparisons between fresh and frozen samples should be avoided, as the effects of cryopreservation can overshadow the signal of

interest. However, employing the same preservation method for all analyzed samples introduces a systematic bias that affects all samples equally, enabling valid comparative analysis (70).

Significant loss of cells and RNA occurs at each step of the process, including encapsulation, RT, sequencing, and data analysis. It is estimated that less than 20% of all transcripts from each cell are efficiently captured and sequenced (57). Consequently, scRNA-seq may not be optimal for detecting lowly expressed genes, particularly at low sequencing depths. Transcription is a stochastic process, and the amount of transcript present within a single cell is minimal. Therefore, achieving comprehensive capture of the transcriptome in scRNA-seq relies on sequencing multiple single cells from the same population. Consequently, the number of expressed genes detected in single cells is typically lower compared to bulk RNA measurements. However, the large number of cells analyzed in parallel partially compensates for the low RNA recovery per cell (73). Ongoing efforts aim to increase the overall efficiency of scRNA-seq methods.

Formation of multiplets

The formation of doublets or multiplets during cell encapsulation poses a significant challenge in scRNA-seq. The encapsulation of cells and barcoded beads into a water-in-oil emulsion droplet has been shown to follow a Poisson distribution of the following possibilities (57):

a) No cell, no bead

b) ≥ 1 cell(s), no bead

c) No cell, ≥ 1 bead(s)

d) ≥ 1 cell(s), ≥ 1 bead(s)

The microfluidic platforms are optimized to minimize the occurrence of possibilities a and b. Computational methods can address the encapsulation of multiple beads by removing cells with dual barcodes (scenario c). However, encapsulating two or more cells with a unique barcode can lead to erroneous biological interpretations (62), such as the false identification of a hybrid cell type (74) or the inadvertent removal of *bona fide* cell-cell complexes through overzealous filtration (75).

While computational methods are being developed to better identify doublets, such as those based on the detection of dual expression profiles (76), currently, technical doublets

cannot be confidently distinguished from true dual lineage cells using scRNA-seq (74,75). Therefore, the experimental protocol should be optimized to limit the formation of multiplets. This involves preparing a true single-cell solution, avoiding cell clumps, and selecting an appropriate cell concentration, as the rate of multiplet formation increases with higher cell concentrations. The multiplet rate should be maintained below 5%, although it involves a trade-off between cell-capture efficiency and the occurrence of multiplets (57). In the coming years, dedicated efforts should be directed towards limiting multiplet formation and enhancing their detection. This can be achieved through various approaches, such as optimizing microfluidics platforms, developing more robust computational tools, and employing complementary experimental methods.

Batch effects

ScRNA-seq platforms can only analyze a limited number of samples at a time (eight samples per Chromium chip for the 10X Genomics platform). When samples from the same condition are processed together, it results in a confounded experiment where technical variability masks biological variation, commonly known as batch effect (Figure 8). This can entirely hamper the ability to detect biological signals in scRNA-seq datasets. Although computational methods can partially address batch effect through integration techniques, none of these methods can reliably separate biological and technical variation. To mitigate batch effects, careful consideration must be given to designing a balanced study that evenly distributes technical biases across the conditions being investigated. As depicted in figure 8, if three different conditions are being investigated, each processing batch should include samples from all three conditions. Furthermore, attribution to each batch should be random (77). Adopting good practices such as having the same investigator process all samples, using a pre-planned, reproducible protocol, minimizing sample handling steps, and utilizing dedicated computational tools can further minimize batch effects.

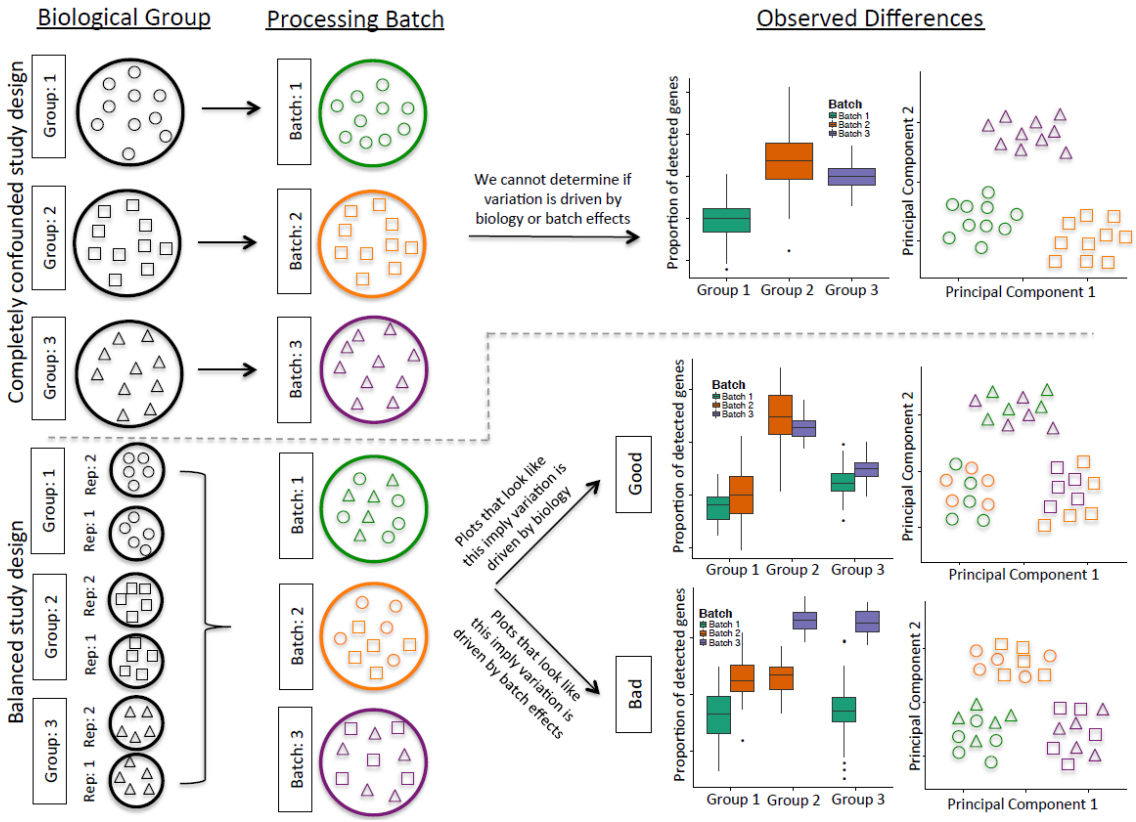


FIGURE 8. The problem of confounding biological variation and batch effects. The top section depicts a completely confounded study design of processing individual cells from three biological groups (represented by shapes) in three separate batches (represented by colors). In this case, we cannot determine if biology or batch effects drive the observed variation. The bottom section depicts a balanced study design consisting of multiple replicates (rep) split and processed across multiple batches. The use of multiple replicates allows observed variation be attributed to biology (cells cluster by shape) or batch effects (cells cluster by color). From Hicks et al. 2018 (78)

d. Computational challenges

Data sparsity

ScRNA-seq datasets have a high proportion of zeros or drop-outs, known as sparsity. This occurs because individual cells often have no detectable RNA due to their low RNA content. It is common for over 50% of entries in the cell-gene count matrix to be zeros. These zeros can be categorized as biological or artificial. Biological zeros indicate the genuine absence of transcripts in cells, while artificial zeros arise from technical issues during library preparation or sequencing. During library preparation, transcripts may fail to be captured by the poly-d(T) tail or not amplify properly. Additionally, if the sequencing depth is insufficient, transcripts that are present in the library may go undetected. Artificial zeros can be either systematic, such as sequence-specific mRNA degradation during cell lysis, or random, as seen with lowly

expressed transcripts that may or may not be detected due to sampling variation. The degree of sparsity depends on the scRNA-seq platform, sequencing depth, and gene expression level (79).

The occurrence of drop-outs poses significant computational challenges. Most of the computational tools used for scRNA-seq data analysis are derived from bulk RNA data science and are ill-equipped to handle the negative binomial distribution observed in single-cell expression measurements (55). To address data sparsity, it is important to sample a sufficient number of cells, which can provide comparable sensitivity to bulk RNA sequencing (73). Ongoing efforts are dedicated to developing mathematical models that effectively address data sparsity in scRNA-seq (79).

Standardization of scRNA-seq analysis pipelines

The lack of standardization in processing pipelines for scRNA-seq analysis poses a significant challenge. This issue not only makes it difficult for novice researchers to select the appropriate tools but also hampers the comparability of study results. Choosing the right computational methods for analysis is nontrivial, as an appropriate analysis pipeline can have a large impact on detecting a biological signal, similar in magnitude to quadrupling the sample size (80). The lack of software standardization directly stems from the relative immaturity of the field; however, there is an anticipated shift toward standardization as the field advances. Open access to raw data and analysis code enables extensive benchmarking of these tools, which will facilitate the identification of the most efficient and reliable methods. However, one should bear in mind that there is no, and probably never will be, a gold standard scRNA-seq analysis pipeline. Indeed, the multi-dimensional nature of scRNA-seq data allows for various interrogation strategies, which requires flexibility and adaptability in the analysis pipeline.

e. Challenges in data interpretation

Interpreting scRNA-seq data presents significant difficulties due to its high dimensionality and complexity, surpassing the comprehension capacity of the human brain. Computational processing offers a means to capture certain aspects of the data and extract relevant information. However, the absence of established guidelines or a gold standard for data interpretation hinders progress in the field.

Incomplete reference annotations

A significant hurdle in data interpretation is the lag in available reference annotations compared to the advancements in scRNA-seq technology. Until recently (81), the genome

assembly for the domestic horse, EquCab3.0, heavily relied on short-read RNA-seq, making the annotation of 5'- and 3'-UTR regions difficult. Mapping scRNA-seq data derived from transcript 3'-ends to a reference assembly with deficient or incorrect 3'-UTR annotation significantly hinders our ability to identify cell types and interpret gene expression changes. Manual extension of the 3'-ends of the reference annotation has been proposed as a means to enhance read mapping (82). Furthermore, tissue-specific transcript isoforms, particularly those specific to BALF, are likely not annotated, as the construction of EquCab3.0 relied on data generated from equine PBMCs. In horses, the isoform-to-gene ratio of the existing NCBI and Ensembl annotations of the EquCab3.0 genome assembly is 2.3 and 2.0, respectively, compared to 4 for the human genome (81).

To address this, the Functional Annotation of Animal Genome (FAANG) initiative is undertaking the colossal task of constructing a comprehensive atlas of species-specific and tissue-specific functional annotations for large animals. A tissue bank comprising more than 80 tissues from four healthy horses is being progressively sequenced through a large-scale collaborative effort (81). The FAANG project leverages long-read sequencing to enhance the reliability of transcriptome assembly. Notably, this approach can identify more than 50% of isoforms that are not covered by traditional short read sequencing datasets (83).

The two major platforms used to generate long-read transcripts are Nanopore sequencing from Oxford Nanopore Technologies and SMRT (single molecule real-time) sequencing from Pacific Bioscience (83). Nanopore sequencing employs oligo-nucleotide threading through a minuscule protein pore to analyze RNA. By applying voltage across the pore, an ion flow is induced, resulting in a detectable current. A motor protein unravels the RNA while regulating the speed of translocation. Sequential nucleotides entering the pore yield a discernible alteration in current, enabling translation into the corresponding nucleotide sequence (83). The SMRT method relies on a polymerase situated at the base of Zero-Mode Waveguides (ZMWs), where fluorescence-labeled nucleotides are incorporated, as depicted in figure 8. ZMWs are nanostructures that can guide light energy into a very small volume compared to the wavelength of the light. These structures are assembled on a flow cell, enabling the concurrent detection of fluorescence signals from a multitude of molecules. Furthermore, the duration of nucleotide incorporation provides insights into specific base modifications. To reduce the error rate, SMRT sequencing implements a high-precision protocol known as circular consensus sequencing. The circularized DNA templates undergo multiple readings by the polymerase, enhancing the accuracy from approximately 90% to 99.8% (83).

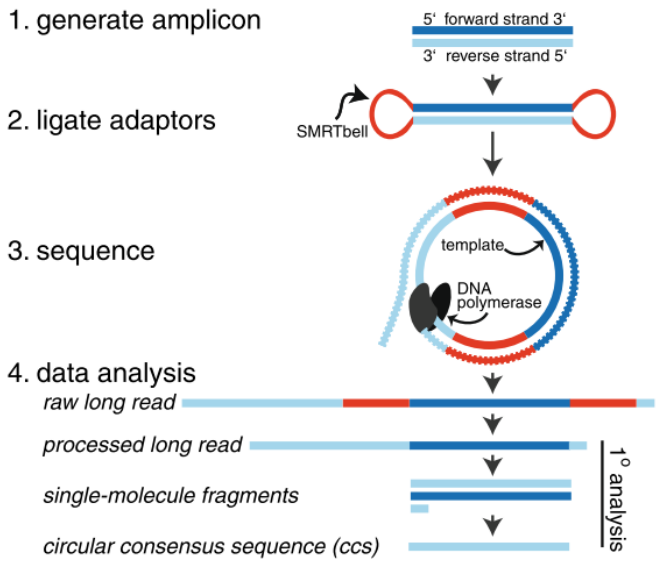


FIGURE 9. Illustration of Pacific Bioscience (PacBio) sequence generation. Adaptors (SMRTbells) are first ligated to each amplicon, and after a sequencing primer is annealed to the SMRTbell template, DNA polymerase is bound to the complex. This polymerase-amplicon-adaptor complex is then loaded into zero-mode waveguides (ZMWs) where replication occurs, producing nucleotide-specific fluorescence. Circular consensus sequencing (ccs) allows the polymerase to repeatedly replicate the circularized strand, producing one long read with randomly distributed errors. Post-run, the SMRTbell sequences are bioinformatically trimmed away, single-molecule fragments are aligned, and a consensus sequence is generated. The single-molecule coverage and accuracy of resulting ccs reads are amplicon- and read-length dependent, with smaller amplicons and longer reads giving higher single-molecule coverage and thus higher ccs read accuracy. From Fichot & Norman 2013 (84)

Recent long-read sequencing of nine different equine tissues improved the current equine transcriptome with 39,625 novel transcripts (81). The integration of a BALF-specific dataset into this combined transcriptome may facilitate future transcriptomics investigations of equine respiratory diseases.

Cell cluster annotation

Automated and manual approaches for cell cluster annotation come with their own caveats. Automated annotation is rapid but lacks the ability to label poorly characterized cells or identify new cell types or states. Additionally, there may be a lack of suitable reference datasets for the specific tissue and species of interest. In the absence of a reference dataset, transferring a human or animal model reference through orthology is possible but may yield inconsistent results (85). Careful examination of the reference dataset is crucial as mistakes in the reference can propagate to the query dataset. Furthermore, similarities in experimental methods between datasets are essential to ensure optimal performance of cell annotation. Automated annotation is better suited for obtaining a general understanding of major cell types,

but manual inspection of markers in the dataset is mandatory to validate the annotation and identify cell subtypes (60).

The process of manual annotation is a time-consuming task that demands specialized expertise. Cell type markers are typically available only for well-characterized organisms and cell types. Consequently, markers from different species are frequently utilized with the assumption that they can be applied to other species. However, a single marker is generally insufficient to define a cell type, necessitating the curation of marker panels from the literature. Cell-surface proteins commonly used in cell classification methods like flow cytometry are often ineffective, likely due to weak correlations between mRNA and protein levels (86). Master transcription factors that drive cell fate often make better gene expression markers. Additionally, the interpretation of gene expression patterns is somewhat subjective, hindering cross-study comparisons where similar cell types may receive different labels. When cell clusters cannot be confidently annotated, caution should be exercised in labeling to avoid propagating erroneous annotations to future studies, particularly in challenging cell subtypes with transient expression gradients. One approach is to decrease clustering resolution, leading to less cluster fragmentation, and assign a more general label (60). The generalization of scRNA-seq studies and data sharing among researchers will contribute to the development of a comprehensive single-cell atlas for diverse biological specimens, detailing the most valuable mRNA markers for each cell population. This will enhance the reproducibility of annotation across studies.

Some experts recommend validating cell annotation using alternative experimental techniques such as in vitro functional assays or imaging experiments (60). However, alternative techniques often fall short in resolution compared to scRNA-seq. Complementary single-cell genomic methods that combine multiple experimental techniques may offer a solution. For instance, CITE-seq from 10X Genomics enables simultaneous immunophenotyping of cell surface proteins and scRNA-seq (86). Although currently available only for humans and mice, this technique may become available for other species in the future.

Despite its challenges, scRNA-seq proves to be a powerful technique that allows for the examination of sample heterogeneity, the discovery of novel cell populations, and the exploration of perturbation effects with remarkable precision. The field of biology research is currently vibrant with enthusiasm, and it is certain that many of these challenges will be effectively addressed in the coming years. Gradually, limitations will be surmounted through the development of new computational tools as scRNA-seq becomes increasingly prevalent in

biological research. Considering that scRNA-seq experiments and analyses are still in their early stages, numerous parameters require adjustment and fine-tuning to enhance the quality of results. The learning curve for researchers interested in utilizing scRNA-seq is undeniably steep and likely endless, but the wealth of insights gained into cellular physiology justifies the effort invested.

Hypotheses and Aims of Thesis

The overall aim of the thesis was to characterize the cell-specific alterations in gene expression associated with SEA using scRNA-seq in equine bronchoalveolar cells. Our expectation was to uncover potential candidate genes that could be explored in the future as disease biomarker or therapeutic targets.

Study 1

Aim: To demonstrate the suitability of scRNA-seq for analyzing cryopreserved BALF cells of horses.

Hypotheses: We hypothesized that scRNA-seq technology could effectively be applied to equine BALF cells, enabling the identification of the five major leukocyte populations that are typically observed on cytological examination (macrophages, lymphocytes, neutrophils, mast cells, and eosinophils). Additionally, we expected to detect specific cell subtypes, such as Th1, Th2, Th17, or Treg lymphocyte subpopulations.

Study 2

Aim: To generate a custom transcriptome annotation for equine BALF cells using long-read sequencing, with the purpose of enhancing the mapping accuracy of scRNA-seq transcripts obtained in Study 1.

Hypotheses: We hypothesized that through long-read sequencing, we would be able to identify novel BALF-specific isoforms and improve the annotation of transcript 3'-ends. We anticipated that mapping our scRNA-seq transcripts to this custom BALF long-read transcriptome would yield higher alignment rates compared to the currently available reference annotation, leading to an overall improvement in the quality of our scRNA-seq data.

Study 3

Aim: To investigate the differences in the composition and transcriptional phenotypes of bronchoalveolar cells between severely asthmatic and healthy horses.

Hypotheses: We hypothesized that the distribution of BALF cells and their gene expression profiles would be altered in horses with SEA. We anticipated that compared to control horses, those with SEA would exhibit a relative elevation in the proportion of Th2 and Th17 cells. Furthermore, we expected several changes in the BALF transcriptome, including impaired cell cycle regulation, upregulation of genes associated with the innate immune response, and downregulation of genes involved in inhibiting lung inflammation in resident alveolar macrophages.

Results

2.1 Study 1: Single-cell gene expression analysis of cryopreserved equine bronchoalveolar cells

Authors: **Sophie E. Sage**, Pamela Nicholson, Laureen M. Peters, Tosso Leeb, Vidhya Jagannathan* and Vinzenz Gerber*

* These authors contributed equally to this work.

Status: published 29 August 2022 in *Frontiers in Immunology*.

Personal contribution:

- Study design
    - Elaboration of clinical scoring sheets
    - Optimization of experimental protocol
- Sample collection with the assistance of Dr. med. vet. Michelle Wyler
- Laboratory work
    - Development of cryopreservation protocol
    - Sample processing before cryopreservation
    - Sample processing before scRNA-seq
- Data analysis
    - Downstream analysis using R, with the assistance of Dr. Vidhya Jagannathan
    - Cell cluster annotation
- Manuscript
    - Draft preparation
    - Figures (Figure 3A with the help of Dr. Vidhya Jagannathan)
    - Finalization and submission
    - Corresponding author

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fimmu.2022.929922

OPEN ACCESS

EDITED BY
Falko Steinbach,
University of Surrey, United Kingdom

REVIEWED BY
Brad Rosenberg,
Icahn School of Medicine at Mount Sinai, United States
Carrie Finno,
University of California, Davis, United States

*CORRESPONDENCE
Sophie E. Sage
sophie.sage@vetsuisse.unibe.ch

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

RECEIVED 27 April 2022
ACCEPTED 08 August 2022
PUBLISHED 29 August 2022

CITATION
Sage SE, Nicholson P, Peters LM, Leeb T, Jagannathan V and Gerber V (2022) Single-cell gene expression analysis of cryopreserved equine bronchoalveolar cells.
Front. Immunol. 13:929922.
doi: 10.3389/fimmu.2022.929922

Single-cell gene expression analysis of cryopreserved equine bronchoalveolar cells

Sophie E. Sage[1*], Pamela Nicholson[2], Laureen M. Peters[3], Tosso Leeb[2,4], Vidhya Jagannathan[4†] and Vinzenz Gerber[1†]

[1]Swiss Institute of Equine Medicine, Department of Clinical Veterinary Medicine, Vetsuisse Faculty, University of Bern, Bern, Switzerland, [2]Next Generation Sequencing Platform, University of Bern, Bern, Switzerland, [3]Clinical Diagnostic Laboratory, Department of Clinical Veterinary Medicine, Vetsuisse Faculty, University of Bern, Bern, Switzerland, [4]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

The transcriptomic profile of a cell population can now be studied at the cellular level using single-cell mRNA sequencing (scRNA-seq). This novel technique provides the unprecedented opportunity to explore the cellular composition of the bronchoalveolar lavage fluid (BALF) of the horse, a species for which cell type markers are poorly described. Here, scRNA-seq technology was applied to cryopreserved equine BALF cells. Analysis of 4,631 cells isolated from three asthmatic horses in remission identified 16 cell clusters belonging to six major cell types: monocytes/macrophages, T cells, B/plasma cells, dendritic cells, neutrophils and mast cells. Higher resolution analysis of the constituents of the major immune cell populations allowed deep annotation of monocytes/macrophages, T cells and B/plasma cells. A significantly higher lymphocyte/macrophage ratio was detected with scRNA-seq compared to conventional cytological differential cell count. For the first time in horses, we detected a transcriptomic signature consistent with monocyte-lymphocyte complexes. Our findings indicate that scRNA-seq technology is applicable to cryopreserved equine BALF cells, allowing the identification of its major (cytologically differentiated) populations as well as previously unexplored T cell and macrophage subpopulations. Single-cell gene expression analysis has the potential to facilitate understanding of the immunological mechanisms at play in respiratory disorders of the horse, such as equine asthma.

KEYWORDS

single-cell mRNA sequencing, cell cryopreservation, equine respiratory system, equine immunology, cell annotation

1 Introduction

Single-cell sequencing technologies have brought a fresh impetus to biological research. With single-cell mRNA sequencing (scRNA-seq), it is now possible to study global gene expression at the single cell level in many complex tissues and heterogeneous cell populations. ScRNA-seq is a powerful tool for the generation of novel hypotheses. The applications of single-cell transcriptomics across the biomedical fields range from the identification of new cell types to the exploration of disease-specific pathological processes, unveiling novel biomarkers and potential therapeutic targets (1). Recently, scRNA-seq gave invaluable insights into the pathological mechanisms leading to respiratory decompensation in SARS-CoV-2 patients (2–4).

Lower airway diseases are common in horses and have a major impact on the equine industry and animal welfare. The processes at play in their pathogenesis are still ill defined due to their complexity, but also due to technical limitations. For the characterization of equine asthma, the most prevalent of these disorders, cytological examination of bronchoalveolar lavage fluid (BALF) is the most widely used technique in both clinical and experimental settings. It is, however, inherently subjective and only allows for the differentiation of five distinct leukocyte populations: macrophages, lymphocytes, neutrophils, mast cells and eosinophils (5). Antibody-based techniques such as immunohistochemistry and flow cytometry permit differentiation of further subpopulations (e.g. within lymphocytes), but in the horse they are restricted to a few cell types due to the limited pool of validated antibodies (6, 7). Individual mRNA transcripts can be measured in BALF mainly by RT-PCR to investigate the influence of various factors on the regulation of specific genes (8–11). This hypothesis-driven approach suffers from a low throughput and a significant investigator bias. In contrast, global transcriptomics is an unbiased, high throughput technique. However, critical differences between individual cells are obscured when performing bulk RNA sequencing of mixed cell populations. This is an important limitation knowing that the cellular composition of the lower respiratory tract, particularly when assessed by BALF sampling, is substantially affected by health status (12). ScRNA-seq enables the description of the different cell populations present in a sample concurrently with their individual transcriptome. Moreover, cell types can be identified without *a priori* knowledge of marker genes.

Single-cell gene expression analysis of bronchoalveolar cells has been successfully performed in humans (2–4), mice (13), ferrets (14) and dogs (15), but not in horses. Before scRNA-seq can be applied to the study of equine respiratory diseases, it is crucial to demonstrate its feasibility on BALF and to build an equine lung-specific reference database. Published scRNA-seq experiments on BALF use fresh samples to optimize cell viability and to avoid the transcriptional changes that could be associated with storage and processing. Being able to store equine BALF cells before scRNA-seq would have several advantages including reducing potential batch effects and facilitating large-scale longitudinal studies.

Here, we applied scRNA-seq to cryopreserved equine BALF cells as a proof of concept, with the aim to differentiate and characterize cell populations based on their transcriptional signatures. It should be noted that the potential effect of cryopreservation on gene expression was not assessed in this study.

2 Material and methods

2.1 Ethics statement

All animal experiments were performed according to the local regulations. This study was approved by the Animal Experimentation Committee of the Canton of Bern, Switzerland (BE07/19).

2.2 Study population

The study was carried out in February 2020. Three horses belonging to the university teaching herd were included. Characteristics of the study population are listed in Table 1. A standard physical examination was performed to assess systemic health. These horses suffered from mild-to-moderate equine asthma, but were in clinical remission at the time of the study. No medication other than alpha-2 agonists (sedation for teaching purposes) were administered for at least a month before the experiment. Lower airway inflammation status was assessed *via* clinical scoring (16), bronchoscopy and BALF cytology. The experimental workflow is summarized in Figure 1.

2.3 Sample collection

A bronchoalveolar lavage was performed with endoscopic guidance under light sedation. Briefly, a flexible endoscope (VET-OR1200HD, Medical Solution GMBH, Wil, SG, Switzerland) was inserted into the pharynx *via* the nasal passages and passed down into the trachea. The endoscope was then advanced into the lower airways *via* the right mainstem bronchus until it wedged against a distal bronchus. Sterile 0.9% NaCl (250 mL) was instilled through the endoscope channel using 60-mL syringes. The fluid was then re-aspirated and the endoscope pulled out. The syringes' content was pooled

TABLE 1 Characteristics of the study population.

	Horse 1	Horse 2	Horse 3
Sex	Gelding	Gelding	Gelding
Age (years)	11	12	10
Breed	Trotteur Français	Franches-Montagnes	Selle Français
HOARSI	3	3	3
Clinical score (/23)	0	0	0
Tracheal mucus score (/5)	1	0	2
BALF yield* (%)	11	8	16

HOARSI, Horse Owner Assessed Respiratory Signs Index; BALF, Bronchoalveolar Lavage Fluid.
*BALF yield % = (Volume saline re-aspirated (mL)/Volume saline instilled (mL)) x 100.

in a cooled silicone-coated glass bottle. The BALF was then filtered through a 40-µm cell strainer (BD Falcon™, Biosciences, USA, cat.352340) and kept on ice until processing.

2.4 Cytology

Cytocentrifuge preparations of the BALF were prepared within 30 minutes following collection. Briefly, BALF was centrifuged at 600 rpm for 8 minutes in a cytocentrifuge (Tharmac Cellspin® I), and slides were subsequently stained with Wright-Giemsa stain (Hematek Stain Pak, Siemens Healthineers, Erlangen, Germany) on an automated slide stainer (Hematek 3000, Siemens Healthineers, Erlangen, Germany). The cell suspensions later obtained after sample thawing (sample used for scRNA-seq) were similarly processed. A manual differential cell count (macrophages, lymphocytes, neutrophils, eosinophils and mast cells) was performed on the cytocentrifuge preparations of the stained BALF and cell suspension using a minimum of 400 cells and 4 different microscopic fields at 1000x magnification with oil immersion.

2.5 ScRNA-seq

2.5.1 Cryopreservation

The protocol used to freeze and subsequently thaw the BALF cells was adapted from a 10X Genomics protocol (CG00039 Rev D) intended for human peripheral blood mononuclear cells (PBMCs). The detailed laboratory protocol can be found in the Supplementary material. Samples were kept on ice throughout the freezing protocol. Cell count and viability were determined with a Moxi GO II™ cell counter (Witec AG, Sursee, LU, Switzerland) using propidium iodide 5:1000. The BALF was initially centrifuged at 300 rcf for 5 minutes at 4°C and the supernatant was removed. Cells were resuspended in RPMI (Gibco™ RPMI 1640 cat. 11875093) containing 40% fetal bovine serum (Gibco™ FBS cat.16000044) and 1 U/µL RNAse inhibitor (Roche® Protector RNase Inhibitor) to achieve a concentration of $20x10^6$ cells/mL. An equivalent volume of freezing medium was added to achieve a concentration of $10x10^6$ cells/mL. The freezing medium consisted of RPMI (Gibco™ RPMI 1640 cat.11875093) with 30% dimethylsulfoxid (MP Biomedicals DMSO ≥99% cat.0219141880) and 40% fetal bovine serum (Gibco™ FBS cat.16000044) added. Cell suspension aliquots were dispensed into cryovials. The cryovials were placed into a pre-cooled cell freezing container (Corning™ CoolCell™ cat. 432005) at -80°C for 4 hours, before

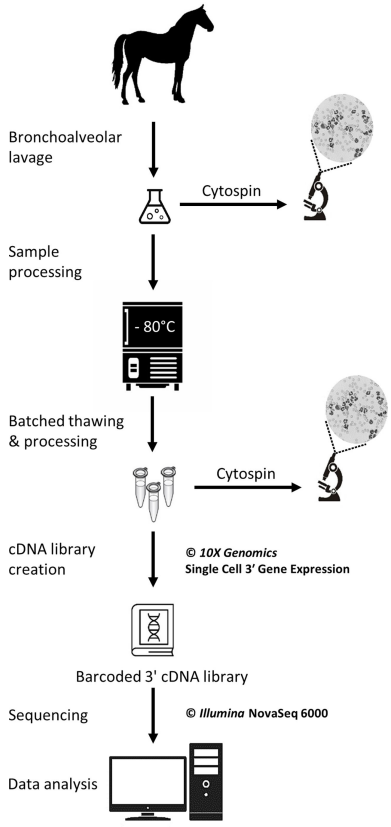


FIGURE 1
Experimental workflow for the single-cell mRNA sequencing of cryopreserved equine bronchoalveolar cells.

being transferred into a cryobox for storage at -80°C. The timespan between BALF collection and freezing was less than 2 hours for all samples.

2.5.2 Thawing and resuspension

Samples were stored at -80°C for 5 to 6 days before thawing and further processing for scRNA-seq. Cryovials (one 1-mL aliquot per horse) were rapidly thawed in a water bath at 37°C. Cells were resuspended 5 times by incremental 1:1 volume addition of complete growth medium, which consisted of RPMI (Gibco™ RPMI 1640 cat.11875093) with 10% fetal bovine serum (Gibco™ FBS cat.16000044). Cells were centrifuged at 300 rcf for 5 min at room temperature. The supernatant was removed except for 1 mL, in which the cells were resuspended. Complete growth medium was added to achieve a total volume of approximately 10 mL. The cell concentration and viability were determined as previously described. A volume of suspension containing $6x10^5$ cells was transferred into a new tube and centrifuged 5 min at 300 rcf at room temperature. The supernatant was removed and the cells were resuspended into 400 µL resuspension solution, which consisted of phosphate-buffered saline (Gibco™ DPBS cat.14190094) containing 0.04% bovine serum albumin (Invitrogen™ UltraPure™ BSA cat.AM2616) and 0.8 U/µL RNAse inhibitor (Roche® Protector RNase Inhibitor). The centrifugation step was repeated and the supernatant discarded. Resuspension solution was added with the goal of achieving a cell concentration between 700 and 1,200 cells/µL. Final cell concentration and viability were measured as described above. Cells were kept on ice until loading into the Chromium™ Controller (10X Genomics, Pleasanton, CA, USA). About 200 µL of the final cell suspension were used to prepare the cytospin slides (see section 2.4).

2.5.3 Single-cell cDNA library preparation and scRNA-seq

GEM generation & barcoding, reverse transcription, cDNA amplification and 3' gene expression library generation steps were all performed according to the Chromium Next GEM Single Cell 3' Reagent Kits v3 User Guide (10x Genomics CG000383 Rev C) with all stipulated 10x Genomics reagents. Nuclease-free water was added to the cell suspensions to reach a total volume of 46.6 µL each, for a targeted cell recovery of 5,000 cells (see details in Supplementary Table 1). GEM generation was followed by a GEM-reverse transcription incubation, a clean-up step and 12 cycles of cDNA amplification. The resulting cDNA was evaluated for quantity and quality using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32851) and an Advanced Analytical Fragment Analyzer System using a Fragment Analyzer NGS Fragment Kit (Agilent, DNF-473), respectively. Thereafter, 3′ gene expression libraries were constructed using a sample index PCR step of 15 cycles. Later, using the same double stranded cDNA, dual indexed libraries were also constructed using a library preparation kit and dual Index kit TT Set (10x Genomics part numbers 1000190 and 300041, respectively). These new libraries were constructed due to the dual indexing upgrade and the upgrade in NovaSeq 6000 reagent kits from v1 to v1.5 and they were generated following the relevant parts of 10 x Genomics User Guide CG000315. The aim of this upgrade in library type was to ensure that this pilot study is more compatible with future data. Any generated cDNA libraries were tested for quantity and size using fluorometry and capillary electrophoresis as described above. The cDNA libraries were pooled and sequenced with a loading concentration of 300 pM, paired end and either single or dual indexed, on an Illumina NovaSeq 6000 sequencer using a shared NovaSeq 6000 S1 Reagent Kit v1.0 (100 cycles; Illumina, 20012865) or a S4 Reagent Kit v1.5 (200 cycles; Illumina, 20028313). The quality of the sequencing runs were assessed using Illumina Sequencing Analysis Viewer (Illumina version 2.4.7) and all base call files were demultiplexed and converted into FASTQ files using Illumina bcl2fastq conversion software v2.20. At least 50,000 reads/cell were generated for each sample. All steps were performed at the Next Generation Sequencing Platform, University of Bern.

2.6 Computational analysis

The workflow followed for the computational analysis is illustrated in Figure 2.

2.6.1 Pre-processing

Raw sequencing data (*fastq* files) were converted to a count matrix of gene expression values using the Cell Ranger (v6.0) standard workflow. The annotations for 3'-untranslated regions of the genes in the reference genome (Equus caballus NCBI annotation release 103) were extended by 2 kb using a custom Python script and manual curation. All transcripts were extended except when the extension overlapped a neighboring gene. The summary metrics of the detected cells are provided in Table 2.

2.6.2 Quality control and data normalization

Quality control and downstream analysis were carried out using the R software package Seurat (v4.0) (17). Based on visual data inspection, cells that contained less than 200 genes or more than 6,500 gene features and/or greater than 15% mitochondrial genes were filtered (Supplementary Figure 1). The raw data were normalized using global scale normalization.

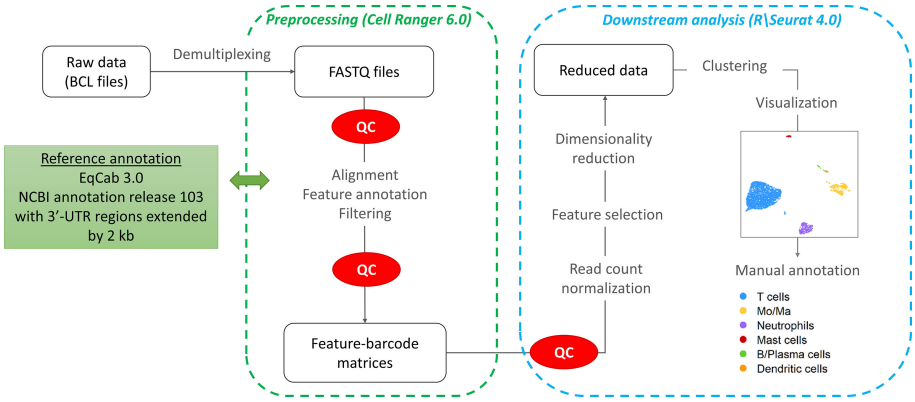


FIGURE 2
scRNA-seq data analysis workflow.

2.6.3 Principal component analysis (PCA) and cell clustering

After variance stabilizing transformation, the 2,000 most variable features were selected for dimension reduction and clustering. Dimensionality reduction was conducted using Principal Component Analysis (PCA). The number of principal components (PCs) was chosen based on an elbow plot. Clustering was performed on the first 16 PCs using the default Louvain algorithm ("FindNeighbors" function in Seurat). The optimal clustering resolution was chosen by visualizing the granularity with the "clustree" R package. The "FindClusters" function (Seurat) was used with a clustering resolution of 1.0. The clusters were visualized with the Uniform Manifold Approximation and Projection method (UMAP). After data integration, cluster visualization was only marginally improved. Cluster membership remained the same, except for a very small proportion of cells (0.5%) which were assigned to Mo/Ma instead of DC. Hence we elected to perform the downstream analysis without integration to avoid introduction of bias due to non-linear data integration approaches (18).

In order to better distinguish cell subpopulations, we independently reanalyzed three of the major cell populations using the "subset" function in Seurat. All previous steps were repeated. We selected 13 PCs with a 1.2 clustering resolution for monocyte/macrophages, 11 PCs with a 0.5 clustering resolution for T cells, and 8 PCs with a 0.7 clustering resolution for B/plasma cells.

2.6.4 Cell cycle analysis

To investigate whether the cell cycle stage affected clustering, we used the "CellCycleScoring" function (Seurat package). Briefly, the lists of human markers for the G2M phase and the S phase ("*cc.genes.updated.2019*" from Seurat) were converted to their equine orthologs using the Biomart R package. Cells were divided into cycling (G2M phase) or resting (S phase) based on the score obtained.

2.6.5 Cell cluster annotation

The cell clusters were annotated based on the expression of canonical markers and subsequently merged into major cell types. Cell cluster annotation was confirmed by subjective analysis of the list of markers identified with the "FindAllMarkers" Seurat function, using an adjusted *P*-value <0.05 and an average log2 fold change > 0.25. Expression of specific cell type markers was visualized using the "DotPlot", "VlnPlot" and "FeaturePlot" functions. Cell type specificity of the markers was evaluated based on the information provided on the Human Protein Atlas version 21.0 database[a]. To facilitate cell cluster annotation, expression scores for cell-specific group of genes were calculated using the Seurat function "AddModuleScore". Gene expression scores of Mo/Ma 2, Mo/Ma 4 and T cells were compared using a Kruskall-Wallis test, with *P*-value < 0.05 considered statistically significant. The differentially expressed genes (DEGs) for each of the 16 clusters can be found in Supplementary Table 2.

3 Results

Three university-owned horses known to be affected by mild-to-moderate equine asthma were included. They were considered systemically healthy based on a complete physical examination. The diagnosis of mild-to-moderate equine asthma was confirmed based on the following three criteria: a Horse Owner Assessed Respiratory Signs Index (HOARSI) score of 3 (19, 20); BALF cytology values of

a The Human Protein Atlas. https://www.proteinatlas.org/ [Accessed May 23, 2022]

TABLE 2 Summary metrics of the detected cells for each sample.

	Horse 1	Horse 2	Horse 3
Estimated number of cells	1,460	1,931	2,017
Fraction reads in cells	86.0%	91.1%	78.8%
Mean reads per cell	115,356	71,580	88,570
Median UMIs per cell	2,922	2,637	2,376
Median genes per cell	1,340	1,256	1,061
Total genes detected	18,287	18,910	18,893
Sequencing saturation	88.7%	81.1%	84.5%
Reads mapped confidently to genome	92.5%	92.6%	91.7%
Reads mapped confidently to exonic regions	57.4%	58.6%	57.0%
Reads mapped confidently to transcriptome	53.6%	55.0%	57.0%

Data generated with CellRanger v6.0 using EqCab 3.0 NCBI annotation release 103 with 3'-UTR regions extended by 2 kb.

>10% and <25% neutrophils, >5% mast cells or >5% eosinophils; absence of increased breathing effort at rest (12). The low clinical score and tracheal mucus score indicated that the horses were in clinical remission at the time of inclusion (16, 21). Key characteristics of the study population can be found in Tables 1, 3.

A total of 5,408 cells were sequenced, of which 777 were filtered after quality control. Downstream analysis was thus performed on 4,631 cells. Unsupervised graph-based clustering produced 16 clusters, which were grouped into six major cell populations based on the expression of canonical markers (Figure 3). The differentially expressed genes (DEGs) for each of the 16 clusters can be found in (Supplementary Figure 2), Cells did not cluster based on the individual sample, therefore data integration was not performed. Clusters 7, 8, 9 and 15 expressing *CD163* and *CD68* were identified as monocytes/macrophages (Mo/Ma). Cluster 13 was annotated as dendritic cells (DCs) based on *CD83* and *CCR7* expression (22). These were most likely myeloid DCs, based on the upregulation of *FSCN1*, a gene not expressed in plasmacytoid DCs (23). Clusters 0, 1, 2, 3, 4, 5, 12 and 14 were annotated as T cells based on *CD2, CD3D, CD3E* and *CD3G* expression. Cluster 11 was annotated as B/Plasma cells based on *MS4A1*, *CD79A* and *CD79B* expression. Cluster 6 was annotated as neutrophils based on *TG*, *RGS2*, *LILRA5* and *CSF3R* expression. Lastly, cluster 10 was identified as mast cells based on *LTC4S*, *HPGDS*, *GCSAML* and *MS4A2* expression. Ribosomal protein (RP) genes were highly expressed in T cells and B/plasma cells (Supplementary Figure 2). After merging the clusters into these six major cell populations, the (DEGs) for each population were extracted (Supplementary Table 3). Analysis of these markers supported our annotations.

3.1 Cell distribution

The distribution of the six major cell populations identified with scRNA-seq was similar for the three horses, except for an increased mast cell proportion in horse 1 (Figure 3E). Next, we inspected the distribution of the five cytologically distinguishable leukocyte types (macrophages, lymphocytes, neutrophils, eosinophils and mast cells). We considered that under light microscopy, T cells and B/plasma cells are counted as lymphocytes, while Mo/Ma and DCs are counted as macrophages. To assess the effect of sample processing and cryopreservation on differential cell counts

TABLE 3 Differential cell counts (DCC) obtained with conventional cytology and single-cell RNA sequencing (scRNA-seq).

	Horse 1			Horse 2			Horse 3		
	BALF cytology	Cell suspension cytology	scRNA-seq	BALF cytology	Cell suspension cytology	scRNA-seq	BALF cytology	Cell suspension cytology	scRNA-seq
Lymphocytes %	55.8	60.8	85.3	40.5	39.5	74.7	39.5	39.5	71.3
Macrophages %	25.8	22.8	7.4	44.0	48.0	13.7	46.0	48.5	16.9
Neutrophils %	5.8	4.0	3.4	14.5	12.0	10.5	14.0	10.0	11.0
Mast cells %	12.8	12.3	3.8	1.0	0.5	1.1	0.5	2.0	0.8
Eosinophils %	0	0	0	0	0	0	0	0	0
Lymphocytes/ Macrophages	2.2	2.7	11.5	0.9	0.8	5.5	0.9	0.8	4.2

BALF cytology was performed on a fresh BALF sample shortly after collection. Cell suspension cytology was performed on an aliquot of the thawed sample used for scRNA-seq. ScRNA-seq DCC was calculated based on the number of cells present in each of the six major cell populations listed, with B and T cell clusters counted as lymphocytes and DCs counted as macrophages.

(DCC), we first compared the cytological DCC performed on fresh BALF to the cytological DCC performed on the final cell suspension. The cell distribution was mostly unchanged, suggesting cryopreservation did not significantly affect DCC. Next, we compared the cytological DCC of the cell suspension to the DCC calculated with scRNA-seq. The lymphocyte/macrophage ratio was found to be markedly higher with scRNA-seq compared to cytology. The proportions of the other cell populations were similar (Figure 3D and Table 3).

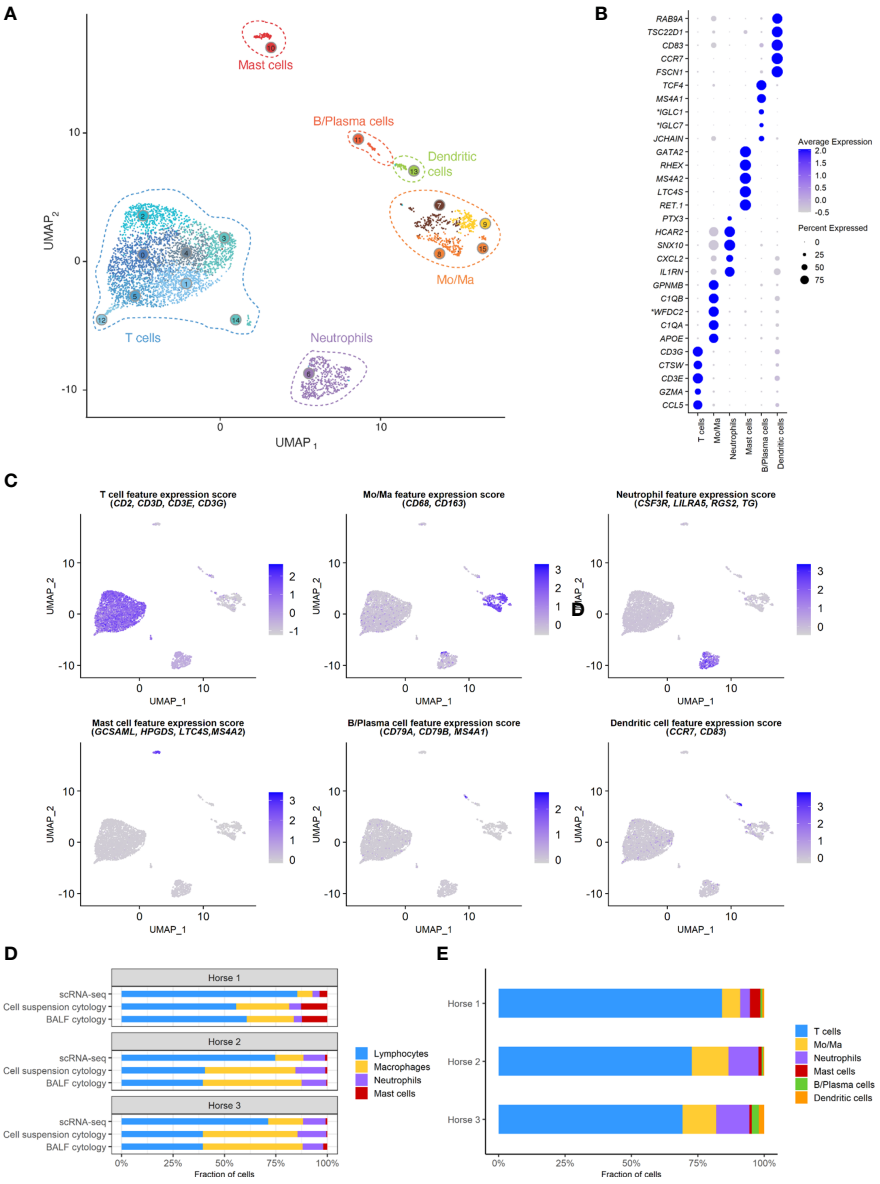
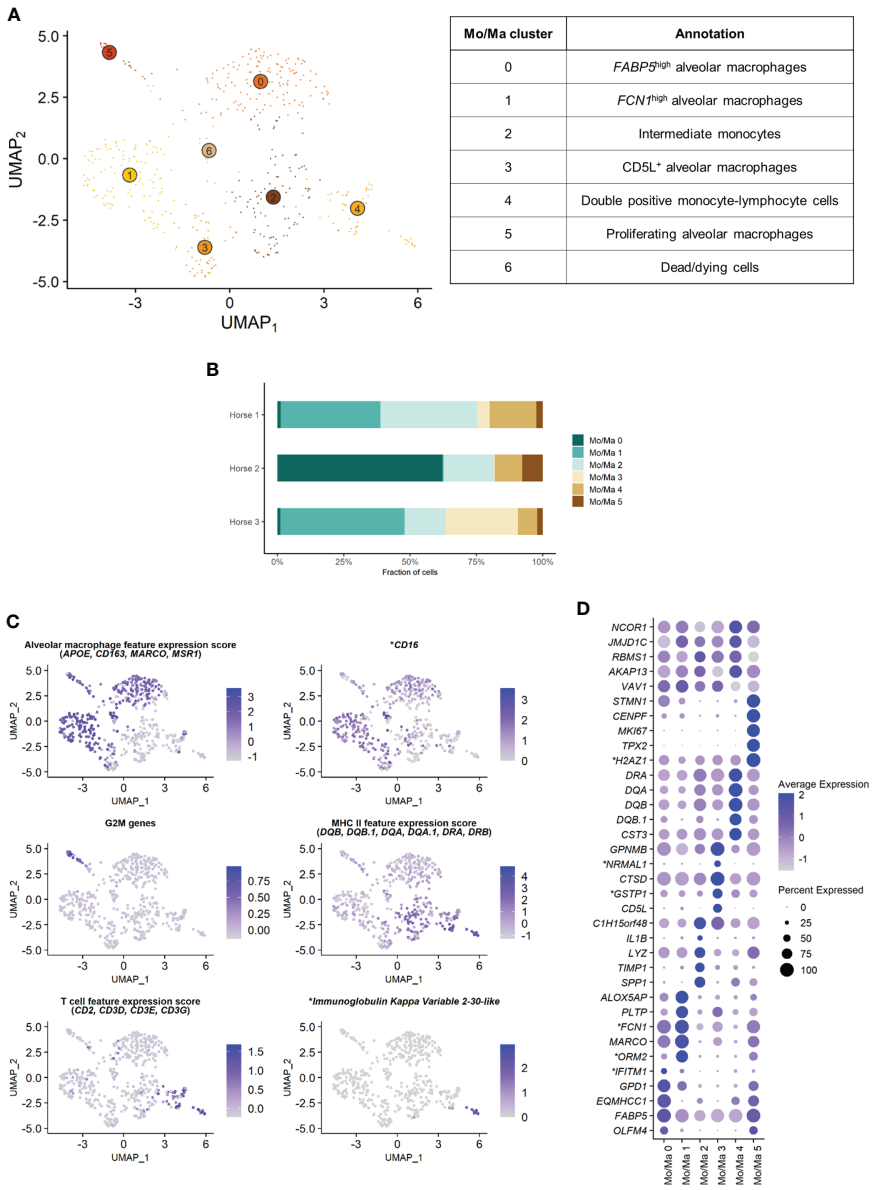


FIGURE 3
scRNA-seq analysis of 4,631 cryopreserved equine bronchoalveolar cells isolated from 3 horses. *Mo/Ma, monocytes/macrophages.* **(A)** The 16 clusters identified (indicated by numbers) are grouped into 6 major cell populations (UMAP visualization). **(B)** Dot plot of the 5 most upregulated genes in each major cell population. Dot size is proportional to the percentage of cells expressing the gene. Dot color intensity represents average gene expression. **Gene ID LOC102147726 annotated as Immunoglobulin Lambda-1 Light Chain (IGLC1), LOC100060608 as Immunoglobulin Lambda Constant 7 (IGLC7) and LOC102148710 as WAP four-disulfide core domain protein 2 (WFDC2) (NCBI EqCab3.0 v103).* **(C)** Gene expression patterns used for cell type assignment. **(D)** Distributions of the major bronchoalveolar cell populations obtained with cytology on BALF, with cytology on the cell suspension (post cryopreservation) and with scRNA-seq on the cell suspension. T cells and B/plasma cells are counted as lymphocytes, while Mo/Ma and DCs are counted as macrophages. **(E)** Distribution of the major bronchoalveolar cell populations for each horse, based on scRNA-seq analysis.



Mo/Ma cluster	Annotation
0	*FABP5*^{high} alveolar macrophages
1	*FCN1*^{high} alveolar macrophages
2	Intermediate monocytes
3	CD5L^{+} alveolar macrophages
4	Double positive monocyte-lymphocyte cells
5	Proliferating alveolar macrophages
6	Dead/dying cells

FIGURE 4
Independent analysis of the monocytes-macrophages (Mo/Ma) population (n=529 cells) **(A)** UMAP visualization of the 7 Mo/Ma clusters identified, with suggested cluster annotation. **(B)** Distribution of the Mo/Ma clusters for each horse, based on scRNA-seq analysis. **(C)** Gene expression patterns used for cell subtype assignment. **LOC100051526 annotated as CD16 (NCBI EqCab3.0 v103).* **(D)** Dot plot of the 5 most upregulated genes in each cluster. Dot size is proportional to the percentage of cells expressing the gene. Dot color intensity represents average gene expression. Mo/Ma 6 (dead cells) is excluded. **LOC100146489 annotated as H2A.Z variant histone 1 (H2AZ1), LOC100067916 as NmrA like redox sensor 1 (NRMAL1), LOC100059533 as glutathione S-transferase pi 1 (GSTP1), LOC100069029 as ficolin 1 (FCN1), LOC100050560 as interferon induced transmembrane protein 1 (IFITM1) and LOC100050100 as orosomucoid 2 (ORM2) (NCBI EqCab3.0 v103).*

3.2 Composition of the monocytes-macrophages population

The Mo/Ma subset was reanalyzed independently to better resolve putative cell subtypes. Unsupervised graph-based clustering identified seven distinct Mo/Ma clusters (Figure 4A). Some of the gene expression patterns used for cell subtype assignment are displayed in (Figure 4C). The five most upregulated genes in each cluster are shown in (Figure 4D). The DEGs list for each cluster can be found in (Supplementary Table 4). Mo/Ma 6 represented dead or dying cells, based on a low RNA and feature count (Supplementary Figure 3).

3.2.1 Alveolar macrophages

Clusters Mo/Ma 0, Mo/Ma 1, Mo/Ma 3 and Mo/Ma 5 overexpressed the macrophage-specific markers *MARCO*, *APOE*, *MSR1* and *CD163*, indicating that these cells were mature alveolar macrophages (AM). Mo/Ma 0 and Mo/Ma 1 had the highest expression levels for *MARCO* and *APOE*. These cells also overexpressed complement genes (e.g. *C1QB*, *C4BPA*, *C1QC* and *C1QA*), suggesting these were activated AMs (24).

Mo/Ma 0 overexpressed *OLFM4*, *FAPB5* and *ANXA2*. This cell cluster was annotated as *FABP5*^{high} AMs, in reference to the previously described *FABP4*^{high} (25) or *FABP4*$^{+}$ (4) AM populations.

Mo/Ma 1 expressed high levels of *FCN1* and *ORM2* (*LOC100050100*). Transcripts coding for lipid mediators (*ALOX5AP*, *PLTP*, *LTA4H* and *APOE*) were overrepresented. This cell cluster was annotated as *FCN1*^{high} AMs in line with a previously published classification (4, 25).

Mo/Ma 3 had somewhat lower levels of AM specific markers compared to Mo/Ma 0 and Mo/Ma 1, perhaps indicating that this cell cluster was at an earlier differentiation stage. This cell cluster was characterized by upregulation of *CD5L*. Numerous anti-oxidant genes were also upregulated (e.g. *FTH1*/*LOC111767398*, *GSTP1*/*LOC100059533*, *TXN*, *SOD2* and *SRXN1*). We detected upregulation of *MARCKSL1*, involved in macrophage migration (26). This cell cluster was annotated as *CD5L*$^{+}$ AMs.

The expression profile of Mo/Ma 5 was dominated by mitosis-associated genes (e.g. *CENPF*, *MKI67*, *TOP2A* and *TPX2*). This cluster was therefore annotated as proliferating AMs (Figure 4C).

3.2.2 Monocytes

Mo/Ma 2 and Mo/Ma 4 showed an overall low expression of AM markers. In particular, *APOE*, *MARCO*, *CD163* and complement genes were downregulated compared to the other Mo/Ma clusters. These cells were thus labeled as monocytes. Both Mo/Ma 2 and Mo/Ma 4 expressed high levels of MHCII-associated genes and downregulated *CD16* (Figure 4C).

Mo/Ma 2 overexpressed genes associated with macrophage chemotaxis (*SPP1* and *CCL15*) and cell migration (*CD44*, *MMP9*) (27–29). We hypothesized that these cells were blood-derived intermediate monocytes, analogous to the extravascular *CD14*$^{+}$*CD16*$^{-}$ HLA-DR^{high} monocytes described by Evren and colleagues (30). The human classical monocyte marker *CD14* was not upregulated in Mo/Ma 2, but it was overall sparsely expressed in our dataset. We annotated Mo/Ma 2 as intermediate monocytes.

The monocytes Mo/Ma 4 upregulated *PLAC8*, considered a signature gene for patrolling monocytes (31). Most importantly, we detected a clear T cell signature in this cluster (*CD2*, *CD3D*, *CD3E*, *CD3G*, *CD5* and *CD7* expression) (Figure 4C). Mo/Ma 4 presented an expression profile "halfway" between Mo/Ma2 (intermediate monocytes) and T cells (Supplementary Figure 5). Several genes responsible for lymphocyte activation (e.g. *CTSW*, *PTPRCAP* and *LTB*) were also overexpressed. A subpopulation of Mo/Ma 4 overexpressed *LOC100630729*. This gene codes for the immunoglobulin kappa variable 2-30-like protein (Figure 4C) expressed in plasma cells. Mo/Ma 4 was annotated as double positive monocyte-lymphocyte cells.

After exclusion of the dead cells (Mo/Ma 6), monocytes (Mo/Ma 2 and Mo/Ma 4), AMs (Mo/Ma 0, Mo/Ma 1, Mo/Ma 3) and proliferating AMs (Mo/Ma 5) represented 30.9%, 64.4% and 4.7% of the Mo/Ma population. While the proportion of monocytes and proliferating AMs was fairly similar among the horses, the distribution of the three distinct AM populations showed a high interindividual variability (Figure 4B).

3.3 Composition of the T cell population

Independent reanalysis of the T cell population identified nine clusters (Figure 5A). Some of the gene expression patterns used for cell subtype assignment are displayed in (Figure 5C). The five most upregulated genes in each cluster are shown in (Figure 5D). The DEGs list for each cluster can be found in (Supplementary Table 5). T7 constituted dead or dying cells, based on a low RNA count and upregulation of mitochondrial genes (Supplementary Figure 3).

Clusters T1, T2 and T6 were CD4^{+} T cells (T helpers), as shown by upregulation of the *CD4* gene and downregulation of the *CD8a* and *CD8b* genes. The remaining clusters T0, T3, T4 and T5 were CD8^{+} T cells (cytotoxic T cells) based on the upregulation of the *CD8a* and/or *CD8b* genes and the downregulation of the *CD4* gene (Supplementary Figure 4). The tissue resident marker *ITGAE* was upregulated in the T cell clusters 0, 2, 3 and 6.

T1 represented naïve CD4^{+} T cells, based on the overexpression of *TCF7* (32, 33). Transcripts coding for ribosomal proteins (RP) were overrepresented. Expression of RP genes was not correlated with the cell cycle state. Increased ribosome biogenesis could instead be explained by the high requirement for protein synthesis during cell differentiation in a naïve cell population as previously observed in stem cells (34).

T2 showed high levels of the tissue resident marker *ITGAE* (33) and the T helper marker *CD40LG*. Upregulation of *DUSP1* and *ANXA1* were consistent with antigen-experienced, activated T cells (35, 36). T2 was thus annotated as CD4^{+} tissue-resident memory (T_{RM}) cells. Interestingly, the top differentially expressed gene was *KLRF1*, a cytotoxicity regulator whose expression is associated with exhaustion of human memory CD4^{+} T cells (37). *RGS1*, a marker of exhaustion in CD8^{+} T cells (38), was also upregulated in this cluster.

Cluster T6 overexpressed Treg cells canonical markers *FOXP3*, *CTLA4* and *IL2RA* (33, 39) and key transcription

factors *FOXO1* (40), *MAF* (41, 42) and *FOXP1* (43), as well as the Treg differentiation regulator *CD27* (44). These cells were annotated as Treg cells.

The expression profiles of the $CD8^+$ clusters T0 and T3 were very alike. Both overexpressed the tissue-resident marker *ITGAE* (33), the cytotoxicity effectors *PRF1* and *CTSW*, the granzyme genes *GZMA*, *GZMK* and *GZMH*, and *EOMES* (45), suggesting they were T_{RM} $CD8^+$ T cells. One substantial difference between the two clusters was the upregulation of the IFN-stimulated genes *IFI6* and *IRF7* and other genes involved in antiviral response (*PLAC8* and *LY6E*) in T0 (46, 47). These genes code for proteins involved in response to viral infection and are upregulated in activated T cells. The chemokines *CCL4* and *CCL5*, secreted by pathogen-specific effector T cells (48), were also upregulated. We consequently annotated T0 as activated $CD8^+$ T_{RM}. On the other hand, T3 was characterized by overexpression of tissue-intrinsic genes participating in cytoskeletal structure (*VIM*, *TUBA1A*) and modulation (*S100A4*, *ANXA2*, *EZR*), cell-cell or cell-matrix interactions (*LGALS1, LGALS3*), and membrane scaffolding (*ITM2B*) (33). Additionally, RP genes were downregulated in this cluster, suggesting it was composed of fully differentiated, tissue-adapted cells (49). The DEGs for T3 included *PRF1*, *NKG7*, *HOPX*, *LAG3* and *CCL4* (33), as well as MHCII-associated genes, consistent with activated T cells. We suspected that T3 represented $CD8^+$ terminally differentiated effector cells (T_{EMRA}) based on the upregulation of *PTPRC* (*CD45*). However, PTPRC average expression was overall low in the cluster and, in absence of PTRPC isoform data in horses, this annotation could not be ascertained. T3 was therefore annotated as $MHCII^{high}$ $CD8^+$ T_{RM} cells.

The ubiquitous expression of the T cell markers *CD2*, *CD3* and *NCAM1* indicated that our dataset did not contain true NK cells. However, T4 still demonstrated NK-specific features. Top upregulated genes included genes from the *NKG2* family (*KLRC1*, *KLRK1*) and the related *KLRD1*, *CD160* and *NKG7* genes, all associated with NK function (Figures 5C, D). *TYROBP* was also overexpressed, in accordance with previous description of human peripheral NKT cells (50). We thus annotated T4 as NKT cells (*CD1d*-restricted invariant natural killer T cells).

Cluster T5 was annotated as γδ T cells. The second most strongly upregulated gene in this cluster was *LOC100066851*, coding for a *SCART1*-like protein (Figures 5C, D). SCARTs are surface receptors found primarily on γδ T cells (51, 52). This cluster shared many DEGs with the equine PBMC cluster annotated as γδ T cells by Patel and colleagues (53). Of note, our BALF γδ T cells also overexpressed genes associated with cytotoxicity such as *KLRB1*, *GNLY* or *KLRF1*.

Cluster T8 was annotated as proliferating T cells based on the high levels of mitosis marker (e.g. *CENP*, *HMGB2*, *TOP2A*) and other markers of cell proliferation such as *MKI67* (Figures 5C, D). This cluster comprised both $CD4^+CD8^-$ and $CD4^-CD8^+$ T cells.

The distribution of the nine T cell clusters was comparable in horses 1 and 3. Horse 2 presented a higher proportion of $CD4^+$ T_{RM} (T2), $MHCII^{high}$ $CD8^+$ T_{RM} (T3, most markedly), and proliferating T cells (T8) (Figure 5B).

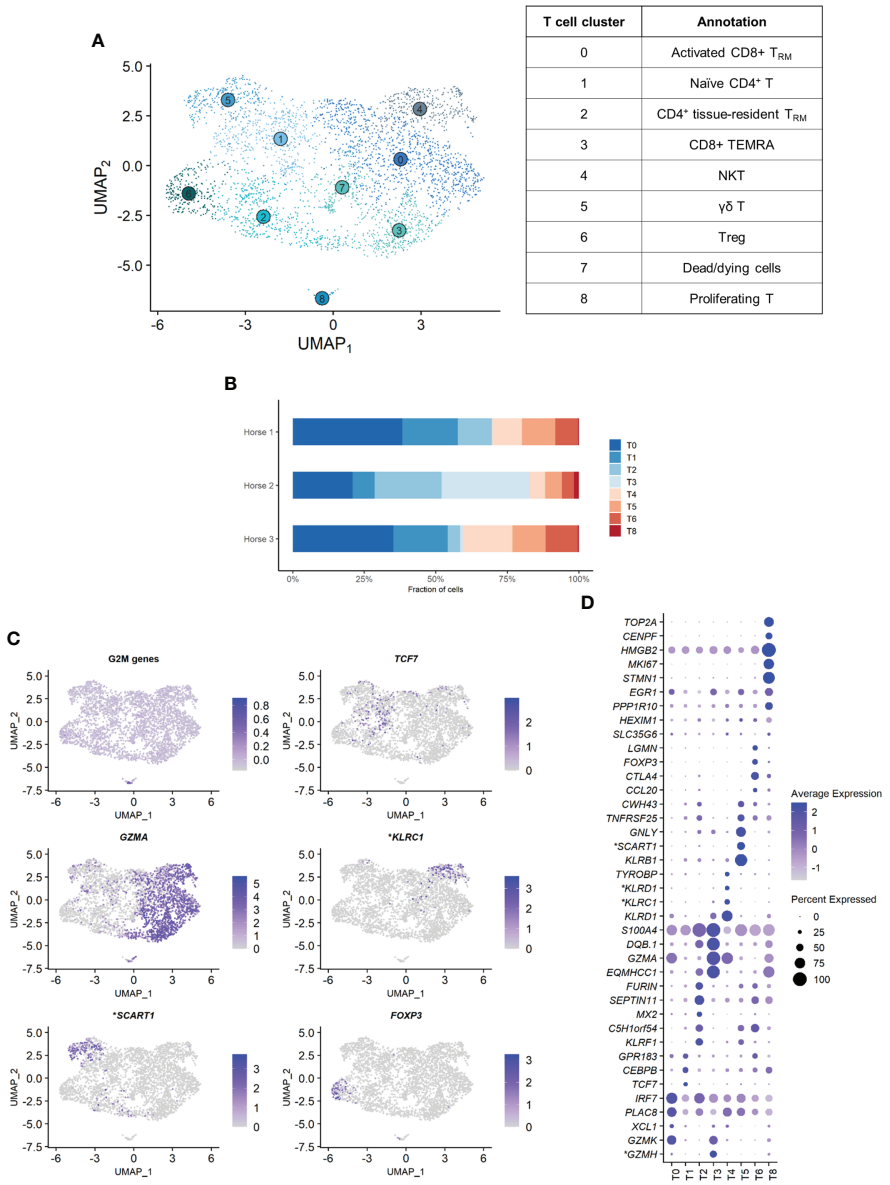
3.4 Composition of the B/plasma cell population

Independent analysis of the B/plasma cell population revealed two distinct cell clusters: B/plasma cell 0 and B/plasma cell 1 and B/plasma cell 1 (Figure 6A). Some of the gene expression patterns used for cell subtype assignment are displayed in (Figure 6C). The DEGs list for each cluster can be found in (Supplementary Table 6). B/plasma 1 overexpressed several genes associated with humoral response, some of which being upregulated in equine blood antibody-secreting B cells (*DERL3*, *HSP90B1*, *PPIB* and *SSR3*) (53). *JCHAIN*, coding for the joining chain of multimeric IgA and IgM, was also overexpressed. B/plasma 1 was thus annotated as plasma cells. On the other hand, B/plasma 0 was characterized by high expression of the canonical B lymphocyte markers *CCR1*, *CD74*, *BANK1*, *MS4A1* and *CD79B*. This cluster was annotated as B cells. The two top upregulated genes for B cells were the MHC class II components *DRB (*ortholog of human *HLA-DRB1)* and *CD74*. MHCII expression is lost during differentiation to plasma cells, further supporting our annotation (54). The relative distribution of B and plasma cells differed between horses, with no plasma cells detected in horse 2 (Figure 6B).

4 Discussion

Our proof of concept single-cell analysis of equine bronchoalveolar cells allowed the identification of the major immune cell populations present in the BALF of adult horses without the need for conventional microscopic cytology or labeled antibodies. We were able to distinguish transcriptionally distinct Mo/Ma and T cell subpopulations relevant for the characterization of different types of immune responses. An unexpected but important finding was the presence of cells or cell pairs expressing both lymphocyte and monocyte markers.

To the best of our knowledge, this is the first scRNA-seq experiment on cryopreserved BALF cells, not only in horses but in any species. We successfully demonstrated that equine bronchoalveolar cells can undergo cryopreservation at -80°C before scRNA-seq. We were able to detect most of the major immune cell types present in equine BALF, with the exception of eosinophils. This was most likely due to their absence or sparsity in the samples, since they were not found on the cytological preparations either. Of note, we found a significant proportion of neutrophils, cells that are notoriously difficult to detect with



T cell cluster	Annotation
0	Activated CD8+ T_{RM}
1	Naïve CD4+ T
2	CD4+ tissue-resident T_{RM}
3	CD8+ TEMRA
4	NKT
5	γδ T
6	Treg
7	Dead/dying cells
8	Proliferating T

FIGURE 5
Independent analysis of the T cell population (n=3,465 cells) **(A)** UMAP visualization of the 9 T cell clusters identified, with suggested cluster annotation. T_{RM}, *tissue-resident memory*. **(B)** Distribution of the T cell clusters for each horse, based on scRNA-seq analysis. **(C)** Gene expression patterns used for cell subtype assignment. **LOC100062846 annotated as killer cell lectin like receptor C1 (KLRC1), LOC100066851 as scavenger receptor family member expressed on T cells 1 (SCART1) (NCBI EqCab3.0 v103).* **(D)** Dot plot of the 5 most upregulated genes in each cluster (one ribosomal RNA gene and two ribosomal protein genes removed). Dot size is proportional to the percentage of cells expressing the gene. Dot color intensity represents average gene expression. T7 (dead cells) is excluded. **LOC100066851 annotated as SCART1, LOC101910264 as killer cell lectin like receptor D1 (KLRD1), both LOC100062823 and LOC100062846 as KLRC1, both LOC100051986 and LOC100147522 as granzyme H (GZMH) (NCBI EqCab3.0 v103).*

scRNA-seq due to their high RNAse content. The addition of RNAse inhibitor during sample processing may have prevented RNA degradation. Detection and correct identification of granulocyte types is important for the characterization of equine asthma phenotypes, highlighting the potential of scRNA-seq to investigate this disease.

The major cell populations were identified with a high level of confidence through the analysis of canonical markers and inspection of the top DEGs for each cluster. The analysis of the cell subtypes was more difficult, with their annotation being subject to interpretation. Automated annotation using SingleR with three different reference datasets [the Human Primary Cell

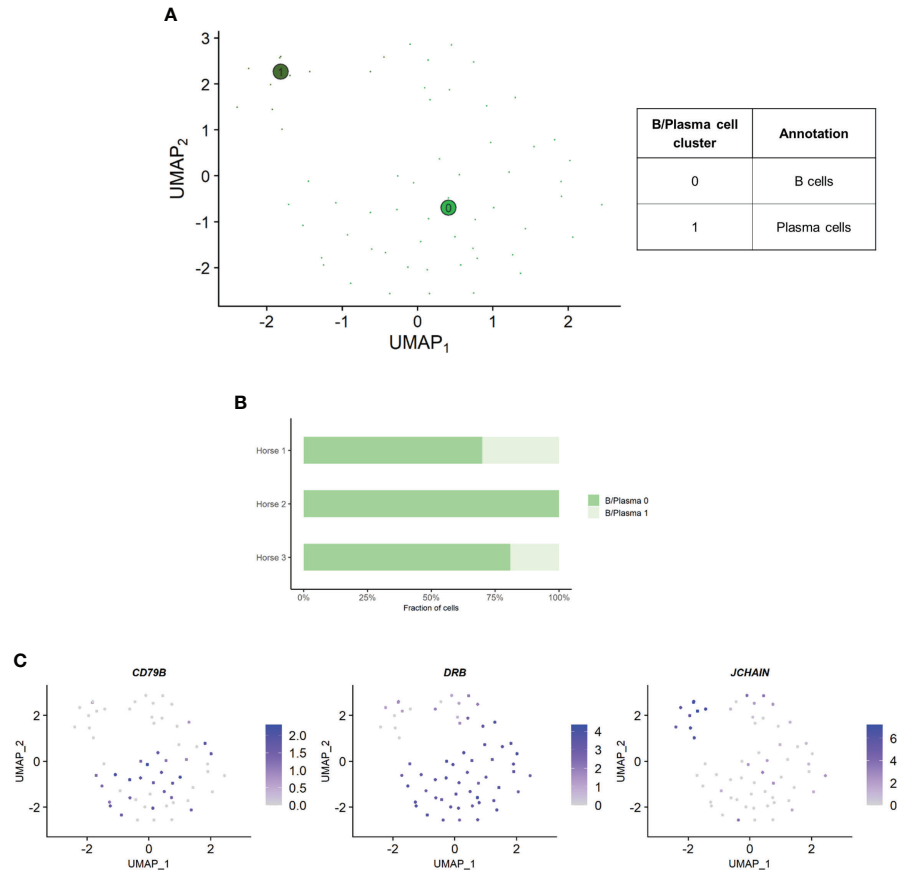


FIGURE 6
Independent analysis of the B/plasma cell population (n=62 cells) **(A)** UMAP visualization of the two B/plasma cell clusters identified, with suggested cluster annotation. **(B)** Distribution of the B/plasma cell clusters for each horse, based on scRNA-seq analysis. **(C)** Gene expression patterns used for cell subtype assignment. **LOC100147624 annotated as derlin 3 (DERL3) (NCBI EqCab3.0 v103).*

Atlas, a human lung cell scRNA-seq dataset (55) and a ferret BALF cell dataset (14)] did not improve the annotation of cell subtypes (data not shown). ScRNA-seq experiments have introduced unprecedented levels of complexity to the classification of cells, challenging the traditional definition of a cell type. Cell types can be defined with various criteria, including their phenotypes, lineages and states (56). Microscopy can only identify a few cell populations based on their morphologic features. On the other hand, use of labeled antibody techniques allow the exploration of cell types at a much higher resolution. These methods, however, rely on the presence or absence of a cell surface marker, shaping our definition of what is a cell type. Additionally, the range of cell types that can be identified in horses is hampered by the limited pool of available equine-specific antibodies. With scRNA-seq, we can appreciate various degrees of expression of a plethora of marker genes, blurring the lines between transcriptionally similar cell populations. The different layers formed by the lineage origin, the differentiation stage, or the activation state can be difficult to untangle. The interesting concept of a cell periodic table was recently proposed (57). Such a table could reconciliate the various strati detected with single-cell analysis to redefine cell types and states. Our dataset could contribute to the assembly of such a periodic cell type table for horses. It will also aid in the construction of an equine lung atlas, similar to the human lung atlas initiative[b].

We observed a marked discrepancy of the lymphocyte/macrophage ratio between cytology and scRNA-seq. This resulted from differences between the techniques used for counting, since the cytological DCCs before and after sample cryopreservation were comparable. We analyzed the data provided by Fastrès et al. and noticed that the scRNA-seq analysis of healthy dogs' BALF cells was similarly biased toward a higher lymphocyte count compared to cytology (15). The lymphocyte/macrophage ratio was five times higher with scRNA-seq, akin to our findings. The droplet-based sequencing method could theoretically bias the observed cell

b The Human Cell Atlas. https://www.humancellatlas.org/ [Accessed May 23, 2022]

distribution through preferential selection of smaller cells. To our knowledge, this has never been reported in immune cells. Another potential explanation for the lower macrophage proportion is the high production of RNAses by granulocytes, leading to greater mRNA degradation and thus lower transcript recovery (58). In such cases however, we would have expected no or low detection of neutrophils and mast cells with scRNA-seq. An alternative and more plausible hypothesis is that conventional cytological DCC diverges stronger from the biological reality than the scRNA-seq based classification. Studies comparing flow cytometry with manual counting have revealed that conventional cytology led to lymphocyte count underestimation (59) or macrophage count overestimation (60). Explanations put forward were the limited number of cells counted under light microscopy, inaccurate classification due to apoptotic changes and uneven cellular distribution on cytological preparations. Indeed, cytocentrifugation tends to propel the lymphocytes to the margins of cytological slides, leading to their underestimation during manual counting (61). Furthermore, in our experience, single epithelial cells can be mistaken for macrophages in light microscopy, potentially leading to an overestimation of the latter in the DCC. In summary, the lymphocyte/macrophage ratio in our cell population was markedly higher with scRNA-seq compared to cytology. Whether scRNA-seq or cytological cell distribution best reflects the biological reality remains to be determined. Since current guidelines to characterize lower airway inflammation in horses are based on cytological DCC (12), we recommend that cell distribution is concomitantly assessed *via* conventional cytology when performing scRNA-seq on equine BALF cells.

Independent analysis of the Mo/Ma cluster enabled a clear distinction between monocytes and AMs based on the expression of a selection of macrophage-specific genes. Previous scRNA-seq studies on BALF from other species (14, 55) and on equine peripheral blood (53) used *CD14* to identify monocytes. The expression of this canonical marker was barely detectable in our dataset, similar to previous observations in dog BALF (62). Yet, equine BALF monocytes have been described as $CD14^+$ cells based on flow cytometric data (63). We suspect that *CD14* expression was blunted by the higher expression of other genes in our cell population. The major DEGs for the three distinct quiescent AM clusters (Mo/Ma 0, 1 and 3) were mostly associated with regulation of inflammation.Cluster Mo/Ma 0 ($FABP5^{high}$ AMs) overexpressed *OLFM4*, *FAPB5* and *ANXA2*. *OLFM4*, a marker of severe lung disease, may be involved in the inflammatory response regulation (64, 65). *FABP5* has anti-inflammatory properties in allergic lung inflammation (66). *ANXA2* negatively regulates TLR4-triggered inflammatory responses, thus preventing excessive inflammation (67). Moreover, the myeloid cell-derived proteins *S100A8* and *S100A9*, involved in lung protective mechanisms (68), were also upregulated in this cluster. The concurrent downregulation of MHCII-related genes in Mo/Ma 0 further supports an anti-inflammatory phenotype (69). Cluster Mo/Ma 1 ($FCN1^{high}$ AMs) was characterized by upregulation of the *FCN1* gene coding for ficolin 1, a pattern-recognition receptor involved in innate immunity. This cluster also expressed high levels of *ORM2* (*LOC100050100*), an acute-phase protein preventing inflammation in adipose and neural tissue (70, 71). Mo/Ma 3 ($CD5L^+$ AMs) had somewhat lower levels of AM specific markers compared to Mo/Ma 0 and Mo/Ma 1, perhaps indicating that this cell cluster was at an earlier differentiation stage. We detected upregulation of *MARCKSL1*, which plays a crucial role in macrophage migration (26). Mo/Ma 3 could thus represent recently migrated AMs. This cell cluster also overexpressed *CD5L*, which promotes M2 macrophage polarization (72, 73). However, none of the clusters displayed a clear M1 or M2 phenotype. AM phenotypes deviating from the classic M1 or M2 phenotypes have already been identified with flow cytometry (74) and RNA-sequencing (75). It has been suggested that M1 and M2 actually represent extremes of a polarization state continuum, rather than stable phenotypes (76, 77). Similar to our findings, previous BALF scRNA-seq studies on ferrets (14) and dogs (15, 62) detected transcriptionally distinct AM clusters. These may represent discrete cell subtypes or different activation states of the same cell type.

We annotated Mo/Ma 2 as intermediate monocytes in reference to a recently published developmental map of human lung macrophages (30). In this study, $HLA\text{-}DR^{high}$ $CD14^+CD16^-$ monocytes egressing from peripheral blood were shown to be AM precursors. Until recently, it was assumed that the pool of AMs was mostly maintained through local self-renewal, with minimal contribution of blood-derived monocytes [reviewed in (78)]. While initial studies were performed in mice, latter work in healthy humans suggests that human AMs are mostly derived from circulating monocytes (78). This discrepancy could stem from the constant exposure of humans to inhaled antigens, in contrast to the germ-free environment of laboratory mice (78). Our findings support the idea that equine AMs mostly arise from the differentiation of peripherally-derived monocytes, similar to what is described in humans. This may reflect a constant renewal of AMs to adapt the lung immune response to the dynamic antigenic load and to the non-specific irritants horses are exposed to *via* the inhalation route in their typical stable environments (12). Single-cell analysis of lung cells collected at sequential developmental stages in the horse may definitely elucidate the origin of equine AMs.

Cluster Mo/Ma 4 stood out by the co-expression of lymphocyte and monocyte markers. Similar double positive cells were detected in three other BALF scRNA-seq studies (4, 30, 62). This was attributed to either ambient RNA contamination (30, 62) or to doublet formation (4). We considered several different hypotheses to explain this unusual expression pattern, including engulfing macrophages, a novel dual lineage cell population, technical multiplets (capture of several cells in a single droplet) and immune cell complexes. The engulfing macrophages hypothesis was not convincing. First, we assume these cells are monocytes

based on downregulated macrophage-specific genes, including those required for phagocytosis (e.g. *CD163*) (79). Second, we did not detect markers specific for neutrophils, the cells that are most likely to be phagocytosed by macrophages (80). While the presence of a novel dual lineage cell population expressing both lymphocyte and monocyte markers cannot be definitely dismissed, it seems unlikely that several flow cytometry-based studies on equine BALF (59, 63, 81, 82) would have failed to detect them. Lymphocytes and monocytes originate from separate cell lineages, namely lymphoid and myeloid. A new dual lineage cell type would thus call into question our current understanding of cell ontology. Mo/Ma 4 could represent technical doublets and/or multiplets. However, this does not explain why only lymphocytes and monocytes signatures are combined, and not signatures for other cell types. Indeed, encapsulation of two or more cells within a droplet during libraries generation should be random. Therefore we propose that the Mo/Ma 4 cluster represents mostly monocyte-T cell complexes, as recently described in human peripheral blood (83, 84). We hypothesize that the Mo/Ma 4 subset overexpressing the Igκ-like protein gene corresponds to monocyte-B cell complexes. Immunological stimuli such as immunization or disease affect the frequency and the phenotype of immune cell complexes (83). Studying their formation and specificity may help unveil the underlying mechanisms of specific equine respiratory conditions such as equine asthma. To determine whether monocyte-lymphocyte complexes are truly present in equine BALF, we suggest combining scRNA-seq with imaging flow cytometry.

We found the annotation of T cell subsets challenging. Similarly to Patel and colleagues working with equine PBMCs (53), we could not identify the Th1, Th2 and Th17 phenotypes, as their specific chemokines were only sparsely expressed in our dataset. Assigning a cell identity to $CD8^+$ clusters was also difficult, probably because gene expression changes in a linear fashion throughout differentiation (45, 85). Further single-cell studies and replication experiments should help to better delineate the transcriptional signatures of the different T cell populations. We annotated one of the T cell cluster as NKT cells. The existence of NKT cells in horses was already reported (86, 87). While NKT cells are well characterized in humans and mice, their transcriptional signature and pattern of surface receptors remain to be defined in horses. A cell cluster with an expression profile similar to our NKT cell population was detected using scRNA-seq on equine PBMCs, but was annotated as NK cells despite being $CD3^+$ (53).Genes coding for RP were overexpressed in the two lymphocyte cell populations (T and B/plasma cells), a common observation in scRNA-seq of lymphocytes. This is likely a consequence of minimal transcriptional activity of most other genes, resulting in an apparent relative increase of abundantly expressed transcripts, such as those encoding RP. This could also reflect the high translational activity required for the cell types respective function, such as cytokine or antibody production (88–90). Even though the DEG list of some T cell clusters were dominated by RP genes, filtering them did not significantly affect cell clustering (data not shown). Moreover, we found that the differential expression of RP genes between T cell clusters corroborated our cluster annotation, with upregulation at the early differentiation stage ($CD4^+$ naïve T cells) and downregulation at the terminal differentiation stage ($MHCII^{high}$ $CD8^+$ T_{RM}) (34, 49).

Three horses from our teaching herd were included in this pilot study out of convenience. These horses were affected with mild-to-moderate equine asthma, but sampled in a phase of clinical remission. Interestingly, horse 2 presented higher proportions of *FABP5*^{high} AMs, $CD4^+$ T_{RM}, $MHCII^{high}$ $CD8^+$ T_{RM} and proliferating T cells, and no plasma cells. This may reflect intrinsic individual differences of the immune response and/or in the environmental exposures encountered throughout lifetime.

A first pre-processing analysis found a low percentage of reads mapping to transcriptome (mean 33.5%, data not shown). We suspected that this resulted from the poor quality of the horse genome annotation, especially at the untranslated 3'-ends of genes. Therefore, we manually extended the annotations of the 3'-untranslated regions for all transcripts listed in the reference genome by 2 kb, as described elsewhere (91). The mapping to transcriptome substantially improved (mean 53.4%), reaching levels comparable with humane or murine studies. This highlights the need for an improved annotation of the equine reference genome.

While a large amount of new information was gained from this study, we acknowledge some limitations. Our study population was small, due to the high cost of scRNA-seq. The number of cells sequenced was, however, sufficient to gather meaningful initial data. While our ability to detect rare cells (e.g. eosinophils) may have been hindered, we could still demonstrate that equine BALF cells can be successfully used for scRNA-seq after cryopreservation. For the present proof-of-concept study, we sampled BALF from asthmatic horses in remission available on site. The gene expression profiles obtained may thus differ from those of horses free of respiratory diseases. A larger-scale study using healthy horses is required to define a single-cell atlas of equine BALF cells. Most single-cell studies are conducted on fresh samples in order to optimize cell viability and prevent transcriptional modifications associated with sample handling. We wanted to demonstrate the feasibility of cryopreservation before sequencing to facilitate the design of larger, potentially multi-centric single-cell studies. Cryopreservation will allow collection of BALF from different horses on separate days for later batched library preparation and sequencing. Analyzing several samples in one single run minimizes batch effect and significantly reduces sequencing cost. Cryopreservation of cells and tissues have a minimal effect on transcriptional profiles obtained with scRNA-seq (92, 93). However, T cells may be more affected by cold preservation, with populations declining over time and an expression profile biased toward cytotoxicity genes (92, 93). Potential confounding effects of cryopreservation on the expression profiles of specific clusters may have been missed in our dataset due to the currently limited

knowledge of cell type-specific transcriptional signatures. The present study was not designed to assess the effect of cryopreservation on gene expression. This should be investigated in future experiments comparing directly fresh and cryopreserved equine BALF samples at the single-cell level. A major limitation of scRNA-seq studies is the difficulty to validate the annotation of cell clusters. Considerable efforts are put into the construction of extensive standardized single-cell atlases (94). Unfortunately, complementary methods that could corroborate the assigned cell clusters (e.g. flow cytometry) fall short in resolution compared to scRNA-seq. The development of new single-cell platforms combining different experimental approaches, such as concurrent gene expression and surface protein labeling, will help to fill this gap.

5 Conclusion

Our findings indicate that scRNA-seq technology is applicable to cryopreserved equine BALF cells, enabling the identification of its major immune cell populations. The sample processing protocol developed for this study may be applied to equine BALF cells and thus allow large-scale single-cell sequencing experiments in horses. Here we provide the single-cell gene expression profiles of the bronchoalveolar cells collected from asthmatic horses in remission. Our collection of signature genes will facilitate cell clustering in forthcoming equine scRNA-seq investigations. We anticipate that single-cell transcriptomic studies will generate novel paradigms in equine respiratory research. Importantly, scRNA-seq will be a powerful tool for the identification of equine asthma endotypes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/ena, PRJEB51962. The R code used for data analysis is available at https://github.com/vetsuisse-unibe/ScRNAseq-BALF-paper.

Ethics statement

The animal study was reviewed and approved by Animal Experimentation Committee of the Canton of Bern, Switzerland.

Author contributions

VJ and VG supervised the study. SS, TL, and VG contributed to the study conception and design. VG acquired funding. SS examined the horses, collected and processed the samples. LP processed and read the cytological preparations. PN coordinated the library preparations and acquired the scRNA-seq data, which was then analyzed by SS and VJ. SS interpreted the data and prepared the original manuscript draft. All authors contributed to manuscript revision, read and approved the submitted version.

Funding

Swiss National Science Foundation (Grant No. 31003A-162548/1); Internal Research Fund of the Swiss Institute of Equine Medicine, Bern, Switzerland (ISMEquine Research No. 33-890).

Acknowledgments

The authors would like to thank med. vet. Michelle Wyler and med. vet. Nicole Altermatt for their assistance in sample collection. We thank the Next Generation Sequencing Platform of the University of Bern for performing the sequencing experiments and the Interfaculty Bioinformatics Unit of the University of Bern for providing the high performance computing infrastructure. The authors furthermore acknowledge the laboratory technicians of the Central Diagnostic Laboratory of the Vetsuisse Faculty for their assistance in sample processing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.929922/full#supplementary-material

References

1. Nayak R, Hasija Y. A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. *Genomics* (2021) 113:606–19. doi: 10.1016/j.ygeno.2021.01.007

2. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat Biotechnol* (2020) 38:970–9. doi: 10.1038/s41587-020-0602-4

3. He J, Cai S, Feng H, Cai B, Lin L, Mai Y, et al. Single-cell analysis reveals bronchoalveolar epithelial dysfunction in COVID-19 patients. *Protein Cell* (2020) 20:680–87. doi: 10.1007/s13238-020-00752-4

4. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* (2020) 26:842–4. doi: 10.1038/s41591-020-0901-9

5. Couetil LL, Thompson CA. Airway diagnostics: Bronchoalveolar lavage, tracheal wash, and pleural fluid. *Vet Clin North Am - Equine Pract* (2020) 36:87–103. doi: 10.1016/j.cveq.2019.12.006

6. Tomlinson JE, Wagner B, Felippe MJB, Van de Walle GR. Multispectral fluorescence-activated cell sorting of b and T cell subpopulations from equine peripheral blood. *Vet Immunol Immunopathol* (2018) 199:22–31. doi: 10.1016/j.vetimm.2018.03.010

7. Davis EG, Wilkerson MJ, Rush BR. Flow cytometry: Clinical applications in equine medicine. *J Vet Intern Med* (2002) 16:404–10. doi: 10.1111/j.1939-1676.2002.tb01257.x

8. Vargas A, Boivin R, Cano P, Murcia Y, Bazin I, Lavoie JP. Neutrophil extracellular traps are downregulated by glucocorticosteroids in lungs in an equine model of asthma. *Respir Res* (2017) 18:1–11. doi: 10.1186/s12931-017-0689-4

9. Lavoie JP, Cesarini C, Lavoie-Lamoureux A, Moran K, Lutz S, Picandet V, et al. Bronchoalveolar lavage fluid cytology and cytokine messenger ribonucleic acid expression of racehorses with exercise intolerance and lower airway inflammation. *J Vet Intern Med* (2011) 25:322–9. doi: 10.1111/j.1939-1676.2010.0664.x

10. Hughes KJ, Nicolson L, Da Costa N, Franklin SH, Allen KJ, Dunham SP. Evaluation of cytokine mRNA expression in bronchoalveolar lavage cells from horses with inflammatory airway disease. *Vet Immunol Immunopathol* (2011) 140:82–9. doi: 10.1016/j.vetimm.2010.11.018

11. Beekman L, Tohver T, Léguillette R. Comparison of cytokine mRNA expression in the bronchoalveolar lavage fluid of horses with inflammatory airway disease and bronchoalveolar lavage mastocytosis or neutrophilia using REST software analysis. *J Vet Intern Med* (2012) 26:153–61. doi: 10.1111/j.1939-1676.2011.00847.x

12. Couëtil LL, Cardwell JM, Gerber V, Lavoie J -P., Léguillette R, Richard EA. Inflammatory airway disease of horses–revised consensus statement. *J Vet Intern Med* (2016) 30:503–15. doi: 10.1111/jvim.13824

13. Mould KJ, Jackson ND, Henson PM, Seibold M, Janssen WJ. Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. *JCI Insight* (2019) 4:1–17. doi: 10.1172/jci.insight.126556

14. Lee JS, Koh JY, Yi K, Kim YIL, Park SJ, Kim EH, et al. Single-cell transcriptome of bronchoalveolar lavage fluid reveals sequential change of macrophages during SARS-CoV-2 infection in ferrets. *Nat Commun* (2021) 12:1–13. doi: 10.1038/s41467-021-24807-0

15. Fastrès A, Pirottin D, Fievez L, Marichal T, Desmet CJ, Bureau F, et al. Characterization of the bronchoalveolar lavage fluid by single cell gene expression analysis in healthy dogs: A promising technique. *Front Immunol* (2020) 11:1707. doi: 10.3389/fimmu.2020.01707

16. Lavoie JP, Bullone M, Rodrigues N, Germim P, Albrecht B, von Salis-Soglio M. Effect of different doses of inhaled ciclesonide on lung function, clinical signs related to airflow limitation and serum cortisol levels in horses with experimentally induced mild to severe airway obstruction. *Equine Vet J* (2019) 51:779–86. doi: 10.1111/evj.13093

17. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177:1888–902.e21. doi: 10.1016/j.cell.2019.05.031

18. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* (2019) 15:e8746. doi: 10.15252/msb.20188746

19. Ramseyer A, Gaillard C, Burger D, Straub R, Jost U, Boog C, et al. Effects of genetic and environmental factors on chronic lower airway disease in horses. *J Vet Intern Med* (2007) 21:149–56. doi: 10.1111/j.1939-1676.2007.tb02941.x

20. Laumen E, Doherr MG, Gerber V. Relationship of horse owner assessed respiratory signs index to characteristics of recurrent airway obstruction in two warmblood families. *Equine Vet J* (2010) 42:142–8. doi: 10.2746/042516409X479586

21. Couetil L, Cardwell JM, Leguillette R, Mazan M, Richard E, Bienzle D, et al. Equine asthma: Current understanding and future directions. *Front Vet Sci* (2020) 7:450. doi: 10.3389/fvets.2020.00450

22. Moyo NA, Marchi E, Steinbach F. Differentiation and activation of equine monocyte-derived dendritic cells are not correlated with CD206 or CD83 expression. *Immunology* (2013) 139:472–83. doi: 10.1111/imm.12094

23. Worah K, Mathan TSM, Vu Manh TP, Keerthikumar S, Schreibelt G, Tel J, et al. Proteomics of human dendritic cell subsets reveals subset-specific surface markers and differential inflammasome function. *Cell Rep* (2016) 16:2953–66. doi: 10.1016/j.celrep.2016.08.023

24. Ogawa T, Shichino S, Ueha S, Ogawa S, Matsushima K. Complement protein C1q activates lung fibroblasts and exacerbates silica-induced pulmonary fibrosis in mice. *Biochem Biophys Res Commun* (2022) 603:88–93. doi: 10.1016/j.bbrc.2022.02.090

25. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J* (2019) 54:1802441. doi: 10.1183/13993003.02441-2018

26. Chun K-R, Bae EM, Kim J-K, Suk K, Lee W-H. Suppression of the lipopolysaccharide-induced expression of MARCKS-related protein (MRP) affects transmigration in activated RAW264. *7 Cells Cell Immunol* (2009) 256:92–8. doi: 10.1016/j.cellimm.2009.01.011

27. Shimizu Y, Dobashi K. CC-chemokine CCL15 expression and possible implications for the pathogenesis of IgE-related severe asthma. *Mediators Inflammation* (2012) 2012:1–7. doi: 10.1155/2012/475253

28. Hollingsworth JW, Li Z, Brass DM, Garantziotis S, Timberlake SH, Kim A, et al. CD44 regulates macrophage recruitment to the lung in lipopolysaccharide-induced airway disease. *Am J Respir Cell Mol Biol* (2007) 37:248–53. doi: 10.1165/rcmb.2006-0363OC

29. Wei J, Marisetty A, Schrand B, Gabrusiewicz K, Hashimoto Y, Ott M, et al. Osteopontin mediates glioblastoma-associated macrophage infiltration and is a potential therapeutic target. *J Clin Invest* (2018) 129:137–49. doi: 10.1172/JCI121266

30. Evren E, Ringqvist E, Tripathi KP, Sleiers N, Rives IC, Alisjahbana A, et al. Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* (2021) 54:259–75.e7. doi: 10.1016/j.immuni.2020.12.003

31. Schyns J, Bai Q, Ruscitti C, Radermecker C, De Schepper S, Chakarov S, et al. Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. *Nat Commun* (2019) 10:3964. doi: 10.1038/s41467-019-11843-0

32. Kim C, Jin J, Weyand CM, Goronzy JJ. The transcription factor TCF1 in T cell differentiation and aging. *Int J Mol Sci* (2020) 21:6497. doi: 10.3390/ijms21186497

33. Szabo PA, Levitin HM, Miron M, Snyder ME, Senda T, Yuan J, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun* (2019) 10:4706. doi: 10.1038/s41467-019-12464-3

34. Sanchez CG, Teixeira FK, Czech B, Preall JB, Zamparini AL, Seifert JRK, et al. Regulation of ribosome biogenesis and protein synthesis controls germline stem cell differentiation. *Cell Stem Cell* (2016) 18:276–90. doi: 10.1016/j.stem.2015.11.004

35. Zhang Y, Reynolds JM, Chang SH, Martin-Orozco N, Chung Y, Nurieva RI, et al. MKP-1 is necessary for T cell activation and function. *J Biol Chem* (2009) 284:30815–24. doi: 10.1074/jbc.M109.052472

36. Gavins FNE, Hickey MJ. Annexin A1 and the regulation of innate and adaptive immunity. *Front Immunol* (2012) 3:354. doi: 10.3389/fimmu.2012.00354

37. Truong K-L, Schlickeiser S, Vogt K, Boës D, Stanko K, Appelt C, et al. Killer-like receptors and GPR56 progressive expression defines cytokine production of human CD4+ memory T cells. *Nat Commun* (2019) 10:2263. doi: 10.1038/s41467-019-10018-1

38. Bai Y, Hu M, Chen Z, Wei J, Du H. Single-cell transcriptome analysis reveals RGS1 as a new marker and promoting factor for T-cell exhaustion in multiple cancers. *Front Immunol* (2021) 12:767070. doi: 10.3389/fimmu.2021.767070

39. Ohkura N, Sakaguchi S. Transcriptional and epigenetic basis of treg cell development and function: its genetic anomalies or variations in autoimmune diseases. *Cell Res* (2020) 30:465–74. doi: 10.1038/s41422-020-0324-7

40. Schumann K, Raju SS, Lauber M, Kolb S, Shifrut E, Cortez JT, et al. Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. *Nat Immunol* (2020) 21:1456–66. doi: 10.1038/s41590-020-0784-4

41. Zhang H, Madi A, Yosef N, Chihara N, Awasthi A, Pot C, et al. An IL-27-Driven transcriptional network identifies regulators of IL-10 expression across T helper cell subsets. *Cell Rep* (2020) 33:108433. doi: 10.1016/j.celrep.2020.108433

42. Wheaton JD, Yeh C-H, Ciofani M. Cutting edge: c-maf is required for regulatory T cells to adopt RORγt + and follicular phenotypes. *J Immunol* (2017) 199:3931–6. doi: 10.4049/jimmunol.1701134

43. Ren J, Han L, Tang J, Liu Y, Deng X, Liu Q, et al. Foxp1 is critical for the maintenance of regulatory T-cell homeostasis and suppressive function. *PloS Biol* (2019) 17:e3000270. doi: 10.1371/journal.pbio.3000270

44. Remedios KA, Zirak B, Sandoval PM, Lowe MM, Boda D, Henley E, et al. The TNFRSF members CD27 and OX40 coordinately limit T h 17 differentiation in regulatory T cells. *Sci Immunol* (2018) 3:1–14. doi: 10.1126/sciimmunol.aau2042

45. van Aalderen MC, van den Biggelaar M, Remmerswaal EBM, van Alphen FPJ, Meijer AB, ten Berge IJM, et al. Label-free analysis of CD8+ T cell subset proteomes supports a progressive differentiation model of human-Virus-Specific T cells. *Cell Rep* (2017) 19:1068–79. doi: 10.1016/j.celrep.2017.04.014

46. Slade CD, Reagin KL, Lakshmanan HG, Klonowski KD, Watford WT. Placenta-specific 8 limits IFNγ production by CD4 T cells *in vitro* and promotes establishment of influenza-specific CD8 T cells *in vivo*. *PloS One* (2020) 15: e0235706. doi: 10.1371/journal.pone.0235706

47. Pfaender S, Mar KB, Michailidis E, Kratzel A, Boys IN, V'kovski P, et al. LY6E impairs coronavirus fusion and confers immune control of viral disease. *Nat Microbiol* (2020) 5:1330–9. doi: 10.1038/s41564-020-0769-y

48. Eberlein J, Davenport B, Nguyen TT, Victorino F, Jhun K, van der Heide V, et al. Chemokine signatures of pathogen-specific T cells I: Effector T cells. *J Immunol* (2020) 205:2169–87. doi: 10.4049/jimmunol.2000253

49. Athanasiadis EI, Botthof JG, Andres H, Ferreira L, Lio P, Cvejic A. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun* (2017) 8:2045. doi: 10.1038/s41467-017-02305-6

50. Zhang J-Y, Wang X-M, Xing X, Xu Z, Zhang C, Song J-W, et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat Immunol* (2020) 21:1107–18. doi: 10.1038/s41590-020-0762-x

51. Holm D, Fink DR, Steffensen MA, Schlosser A, Nielsen O, Moeller JB, et al. Characterization of a novel human scavenger receptor cysteine-rich molecule SCART1 expressed by lymphocytes. *Immunobiology* (2013) 218:408–17. doi: 10.1016/j.imbio.2012.05.025

52. Kisielow J, Kopf M, Karjalainen K. SCART scavenger receptors identify a novel subset of adult γδ T cells. *J Immunol* (2008) 181:1710–6. doi: 10.4049/jimmunol.181.3.1710

53. Patel RS, Tomlinson JE, Divers TJ, Van de Walle GR, Rosenberg BR. Single-cell resolution landscape of equine peripheral blood mononuclear cells reveals diverse cell types including T-bet+ b cells. *BMC Biol* (2021) 19:1–18. doi: 10.1186/s12915-020-00947-5

54. Yoon HS, Scharer CD, Majumder P, Davis CW, Butler R, Zinzow-Kramer W, et al. ZBTB32 is an early repressor of the CIITA and MHC class II gene expression during b cell differentiation to plasma cells. *J Immunol* (2012) 189:2393–403. doi: 10.4049/jimmunol.1103371

55. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* (2020) 587:619–25. doi: 10.1038/s41586-020-2922-4

56. Morris SA, Klein A, Treutlein B. The evolving concept of cell identity in the single cell era. *Development* (2019) 146(12):dev169748. doi: 10.1242/dev.169748

57. Xia B, Yanai I. A periodic table of cell types. *Development* (2019) 146:1–9. doi: 10.1242/dev.169854

58. Koczera P, Martin L, Marx G, Schuerholz T. The ribonuclease a superfamily in humans: Canonical RNases as the buttress of innate immunity. *Int J Mol Sci* (2016) 17:1278. doi: 10.3390/ijms17081278

59. Kang H, Bienzle D, Lee GKC, Piché É, Viel L, Odemuyiwa SO, et al. Flow cytometric analysis of equine bronchoalveolar lavage fluid cells in horses with and without severe equine asthma. *Vet Pathol* (2022) 59:91–9. doi: 10.1177/03009858211042588

60. Hodge SJ, Hodge GL, Holmes M, Reynolds PN. Flow cytometric characterization of cell populations in bronchoalveolar lavage and bronchial brushings from patients with chronic obstructive pulmonary disease. *Cytom Part B - Clin Cytom* (2004) 61:27–34. doi: 10.1002/cyto.b.20020

61. Pickles K, Pirie RS, Rhind S, Dixon PM, McGorum BC. Cytological analysis of equine bronchoalveolar lavage fluid. part 2: comparison of smear and cytocentrifuged preparations. *Equine Vet J* (2010) 34:292–6. doi: 10.2746/042516402776186155

62. Fastrès A, Pirottin D, Fievez L, Tutunaru AC, Bolen G, Merveille AC, et al. Identification of pro-fibrotic macrophage populations by single-cell transcriptomic analysis in West highland white terriers affected with canine idiopathic pulmonary fibrosis. *Front Immunol* (2020) 11:611749. doi: 10.3389/fimmu.2020.611749

63. Karagianni AE, Kapetanovic R, Summers KM, McGorum BC, Hume DA, Pirie RS. Comparative transcriptome analysis of equine alveolar macrophages. *Equine Vet J* (2017) 49:375–82. doi: 10.1111/evj.12584

64. Brand HK, Ahout IML, de Ridder D, van Diepen A, Li Y, Zaalberg M, et al. Olfactomedin 4 serves as a marker for disease severity in pediatric respiratory syncytial virus (RSV) infection. *PloS One* (2015) 10:e0131927. doi: 10.1371/journal.pone.0131927

65. Gong F, Li R, Zheng X, Chen W, Zheng Y, Yang Z, et al. OLFM4 regulates lung epithelial cell function in sepsis-associated ARDS/ALI *via* LDHA-mediated NF-κB signaling. *J Inflammation Res* (2021) 14:7035–51. doi: 10.2147/JIR.S335915

66. Kobayashi S, Tayama S, Phung HT, Kagawa Y, Miyazaki H, Takahashi Y, et al. Fatty acid-binding protein 5 limits ILC2-mediated allergic lung inflammation in a murine asthma model. *Sci Rep* (2020) 10:16617. doi: 10.1038/s41598-020-73935-y

67. Zhang S, Yu M, Guo Q, Li R, Li G, Tan S, et al. Annexin A2 binds to endosomes and negatively regulates TLR4-triggered inflammatory responses *via* the TRAM-TRIF pathway. *Sci Rep* (2015) 5:15859. doi: 10.1038/srep15859

68. Hiroshima Y, Hsu K, Tedla N, Wong SW, Chow S, Kawaguchi N, et al. S100A8/A9 and S100A9 reduce acute lung injury. *Immunol Cell Biol* (2017) 95:461–72. doi: 10.1038/icb.2017.2

69. Viola A, Munari F, Sánchez-Rodríguez R, Scolaro T, Castegna A. The metabolic signature of macrophage responses. *Front Immunol* (2019) 10:1462. doi: 10.3389/fimmu.2019.01462

70. Jo M, Kim JH, Song GJ, Seo M, Hwang EM, Suk K. Astrocytic orosomucoid-2 modulates microglial activation and neuroinflammation. *J Neurosci* (2017) 37:2878–94. doi: 10.1523/JNEUROSCI.2534-16.2017

71. Lee YS, Choi JW, Hwang I, Lee JW, Lee JH, Kim AY, et al. Adipocytokine orosomucoid integrates inflammatory and metabolic signals to preserve energy homeostasis by resolving immoderate inflammation. *J Biol Chem* (2010) 285:22174–85. doi: 10.1074/jbc.M109.085464

72. Sanjurjo L, Amézaga N, Aran G, Naranjo-Gómez M, Arias L, Armengol C, et al. The human CD5L/AIM-CD36 axis: A novel autophagy inducer in macrophages that modulates inflammatory responses. *Autophagy* (2015) 11:487–502. doi: 10.1080/15548627.2015.1017183

73. Sanjurjo L, Aran G, Téllez É, Amézaga N, Armengol C, López D, et al. CD5L promotes M2 macrophage polarization through autophagy-mediated upregulation of ID3. *Front Immunol* (2018) 9:480. doi: 10.3389/fimmu.2018.00480

74. Mitsi E, Kamng'ona R, Rylance J, Solórzano C, Jesus Reiné J, Mwandumba HC, et al. Human alveolar macrophages predominately express combined classical M1 and M2 surface markers in steady state. *Respir Res* (2018) 19:1–4. doi: 10.1186/s12931-018-0777-0

75. Takiguchi H, Yang CX, Yang CWT, Sahin B, Whalen BA, Milne S, et al. Macrophages with reduced expressions of classical M1 and M2 surface markers in human bronchoalveolar lavage fluid exhibit pro-inflammatory gene signatures. *Sci Rep* (2021) 11:1–11. doi: 10.1038/s41598-021-87720-y

76. Karagianni AE, Lisowski ZM, Hume DA, Scott Pirie R. The equine mononuclear phagocyte system: The relevance of the horse as a model for understanding human innate immunity. *Equine Vet J* (2021) 53:231–49. doi: 10.1111/evj.13341

77. Ross EA, Devitt A, Johnson JR. Macrophages: The good, the bad, and the gluttony. *Front Immunol* (2021) 12:708186. doi: 10.3389/fimmu.2021.708186

78. Byrne AJ, Powell JE, O'Sullivan BJ, Ogger PP, Hoffland A, Cook J, et al. Dynamics of human monocytes and airway macrophages during healthy aging and after transplant. *J Exp Med* (2020) 217:1–11. doi: 10.1084/jem.20191236

79. Schulz D, Severin Y, Zanotelli VRT, Bodenmiller B. In-depth characterization of monocyte-derived macrophages using a mass cytometry-based phagocytosis assay. *Sci Rep* (2019) 9:1925. doi: 10.1038/s41598-018-38127-9

80. Silva MT. Macrophage phagocytosis of neutrophils at inflammatory/infectious foci: A cooperative mechanism in the control of infection and infectious inflammation. *J Leukoc Biol* (2011) 89:675–83. doi: 10.1189/jlb.0910536

81. Niedzwiedz A, Jaworski Z, Tykalowski B, Smialek M. Neutrophil and macrophage apoptosis in bronchoalveolar lavage fluid from healthy horses and horses with recurrent airway obstruction (RAO). *BMC Vet Res* (2014) 10:29. doi: 10.1186/1746-6148-10-29

82. Lee Y, Kiupel M, Soboll Hussey G. Characterization of respiratory dendritic cells from equine lung tissues. *BMC Vet Res* (2017) 13:313. doi: 10.1186/s12917-017-1240-z

83. Burel JG, Pomaznoy M, Lindestam Arlehamn CS, Weiskopf D, da Silva Antunes R, Jung Y, et al. Circulating T cell-monocyte complexes are markers of immune perturbations. *Elife* (2019) 8:1–21. doi: 10.7554/eLife.46045

84. Burel JG, Pomaznoy M, Lindestam Arlehamn CS, Seumois G, Vijayanand P, Sette A, et al. The challenge of distinguishing cell–cell complexes from singlet cells in non-imaging flow cytometry and single-cell sorting. *Cytom Part A* (2020) 97:1127–35. doi: 10.1002/cyto.a.24027

85. Restifo NP, Gattinoni L. Lineage relationship of effector and memory T cells. *Curr Opin Immunol* (2013) 25:556–63. doi: 10.1016/j.coi.2013.09.003

86. Looringh van Beeck FA, Reinink P, Hermsen R, Zajonc DM, Laven MJ, Fun A, et al. Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants. *Mol Immunol* (2009) 46:1424–31. doi: 10.1016/j.molimm.2008.12.009

87. Dossa RG, Alperin DC, Garzon D, Mealey RH, Brown WC, Jervis PJ, et al. In contrast to other species, α-galactosylceramide (α-GalCer) is not an immunostimulatory NKT cell agonist in horses. *Dev Comp Immunol* (2015) 49:49–58. doi: 10.1016/j.dci.2014.11.005

88. Tan TCJ, Knight J, Sbarrato T, Dudek K, Willis AE, Zamoyska R. Suboptimal T-cell receptor signaling compromises protein translation, ribosome biogenesis, and proliferation of mouse CD8 T cells. *Proc Natl Acad Sci* (2017) 114: E6117–26. doi: 10.1073/pnas.1700939114

89. Tang R, Liu X, Liang C, Hua J, Xu J, Wang W, et al. Deciphering the prognostic implications of the components and signatures in the immune microenvironment of pancreatic ductal adenocarcinoma. *Front Immunol* (2021) 12:648917. doi: 10.3389/fimmu.2021.648917

90. Yang B, Fan J, Huang J, Guo E, Fu Y, Liu S, et al. Clinical and molecular characteristics of COVID-19 patients with persistent SARS-CoV-2 infection. *Nat Commun* (2021) 12:3501. doi: 10.1038/s41467-021-23621-y

91. Kaelin CB, McGowan KA, Barsh GS. Developmental genetics of color pattern establishment in cats. *Nat Commun* (2021) 12:5127. doi: 10.1038/s41467-021-25348-2

92. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol* (2020) 21:1. doi: 10.1186/s13059-019-1906-x

93. Guillaumet-Adkins A, Rodríguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, et al. Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol* (2017) 18:1–15. doi: 10.1186/s13059-017-1171-9

94. Schiller HB, Montoro DT, Simon LM, Rawlins EL, Meyer KB, Strunz M, et al. The human lung cell atlas: A high-resolution reference map of the human lung in health and disease. *Am J Respir Cell Mol Biol* (2019) 61:31–41. doi: 10.1165/rcmb.2018-0416TR

2.2 Study 2: Long read transcriptome of equine bronchoalveolar cells

Authors: **Sophie E. Sage**, Pamela Nicholson, Tosso Leeb, Vinzenz Gerber, Vidhya Jagannathan

Status: published 25 September 2022 in *Genes*.

Personal contribution:

- Laboratory work
    - Optimization of RNA isolation protocol
    - RNA extraction (4 samples)
- Manuscript
    - Writing of "Discussion" section
    - Draft revision

genes

MDPI

Article

Long-Read Transcriptome of Equine Bronchoalveolar Cells

Sophie Elena Sage [1], Pamela Nicholson [2], Tosso Leeb [3], Vinzenz Gerber [1] and Vidhya Jagannathan [3,*]

[1] Swiss Institute of Equine Medicine, Department of Clinical Veterinary Medicine, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland
[2] Next Generation Sequencing Platform, University of Bern, 3001 Bern, Switzerland
[3] Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland
* Correspondence: vidhya.jagannathan@vetsuisse.unibe.ch

Abstract: We used Pacific Biosciences long-read isoform sequencing to generate full-length transcript sequences in equine bronchoalveolar lavage fluid (BALF) cells. Our dataset consisted of 313,563 HiFi reads comprising 805 Mb of polished sequence information. The resulting equine BALF transcriptome consisted of 14,234 full-length transcript isoforms originating from 7017 unique genes. These genes consisted of 6880 previously annotated genes and 137 novel genes. We identified 3428 novel transcripts in addition to 10,806 previously known transcripts. These included transcripts absent from existing genome annotations, transcripts mapping to putative novel (unannotated) genes and fusion transcripts incorporating exons from multiple genes. We provide transcript-level data for equine BALF cells as a resource to the scientific community.

Keywords: Iso-Seq; lung; bioinformatics; annotation; asthma; horse; *Equus caballus*

Citation: Sage, S.E.; Nicholson, P.; Leeb, T.; Gerber, V.; Jagannathan, V. Long-Read Transcriptome of Equine Bronchoalveolar Cells. *Genes* **2022**, *13*, 1722. https://doi.org/10.3390/genes13101722

Academic Editor: Scott Pratt

Received: 14 July 2022
Accepted: 21 September 2022
Published: 25 September 2022

1. Introduction

Equine asthma is a naturally occurring disease of the horse with many similarities to human asthma. The etiology and pathophysiology of equine and human asthma are not completely understood. Bronchoalveolar lavage fluid (BALF) cytology is typically used to confirm the diagnosis in horses [1]. BALF cells are often used to study gene expression in models of asthma. Compared to mice or humans, the collection of BALF in horses is less invasive, as it can be performed in a standing animal with light sedation. This makes horses a valuable animal model for the discovery of novel pathways implicated in asthmatic inflammation [2–4]. To take full advantage of the comparative gene expression data gained from the equine asthma model, it is critical to have an accurate and comprehensive annotation of the genes expressed in equine BALF cells.

Short-read RNA sequencing (RNA-seq) is commonly used to study the changes in gene expression associated with a disease (e.g., equine asthma) [5–7]. This method allows for quantification of transcript abundance in a biological sample at the gene level, and sometimes even at the isoform level. However, short-read data are not sufficient to infer accurate full-length transcript structures, as transcripts with multiple alternative splicing sites cannot be completely covered by the individual short reads. Complete cDNA molecules can now be sequenced using long-read sequencing platforms such as Pacific Biosciences [8] and Oxford Nanopore [9]. In contrast to short-read sequencing, long-read-based techniques allow unambiguous mapping to a reference assembly, conservation of isoform information, as well as discovery of novel genes, fusion transcripts and non-coding RNA. This has led to significant improvement of the transcriptome annotation in a few domestic species such as chicken [10], rabbits [11] or pigs [12]. The FAANG consortium, whose goal is to create reference functional maps of livestock animal genomes [13] is already taking advantage of long-read sequencing. Recently, Peng et al. [14] applied the PacBio Iso-Seq sequencing technique to nine different horse tissues, including lung. They combined their Iso-Seq transcriptome with the existing NCBI and Ensembl EquCab3.0 reference annotations. The

resulting FAANG equine transcriptome comprises 36,239 protein-coding genes and 153,492 transcripts, a considerable improvement compared to the current NCBI annotation, which contains only 21,129 protein-coding genes and 72,102 transcripts. The discovery of novel genes and isoforms with potential biological relevance will benefit to all genomics studies.

Advanced transcriptomics are becoming increasingly popular to study the pathophysiology of equine asthma. Our group recently performed the first single-cell mRNA sequencing of equine BALF [15], a technique that is widely used to study the transcriptome of individual cells and compare them. Peng et al. [14] have also showed that major isoform (based on relative expression values) differs between tissue type, hence a BALF-specific reference transcriptome is highly desirable.

In this study, we applied the Pacific Biosciences Iso-Seq method [8] to generate a full-length transcriptome of BALF cells originating from a healthy and an asthmatic horse. The dataset obtained is a valuable resource for future transcriptomic studies in horses, including investigations of equine asthma.

2. Materials and Methods

2.1. Study Animals

Privately owned horses were prospectively enrolled for a study on equine asthma. Suitable candidates were identified through a validated owner questionnaire (HOARSI) [16]. Requirements for study enrollment were an age $\geq$ 5 years old, a longer than two-month history of being fed hay, no history of immunotherapy, no history of upper airway disease (e.g., roarer), no evidence of current systemic disease and a rectal temperature $\leq$ 38.5 °C on the day of the exam. Additionally, the horses should not have received any corticosteroids, bronchodilators or anti-histaminic administration nor suffered from a respiratory infection in the two weeks preceding the examination. Owners were asked to bring a healthy horse (without respiratory signs) from the same barn along with their asthmatic horse to the Bern ISME equine clinic. Horses underwent the following standard diagnostic procedures to characterize their respiratory status: physical examination with clinical scoring [17], lower airway endoscopy with tracheal mucus scoring [18], and bronchoalveolar lavage. Selection of the samples for RNA extraction was based on the BALF quality (yield $\geq$ 30%, foamy, normal CASY cell counter profile (OMNI Life Science, Bremen, Germany)) and on the horses' historical, clinical and laboratory features, with the goal of selecting the most archetypal phenotypes for both control and severe equine asthma (sEA). RNA was extracted from four BALF samples (two control and two sEA). One control and one SEA sample were then selected for sequencing based on the quality of RNA (see paragraph 2.5). The main characteristics of the sequenced horses are shown in Table 1. Detailed clinical features of these horses can be found in Table S1.

Table 1. Characteristics of the horses used for the study.

Sample	1	2
Health status	Asthmatic	Control
Breed	Selle Français	Swiss Warmblood
Sex	Mare	Mare
Age (years)	21	12
HOARSI *	4	1
Weight (kg)	634	613

* The Horse Owner Respiratory Signs Index (HOARSI) was used assess the equine asthma severity. HOARSI 1 indicates a healthy, and HOARSI 4 a severely asthmatic horse.

2.2. Sample Collection

Horses were sedated with detomidine 0.01 mg/kg IV (Equisedan ®, Graeub, Bern, Switzerland) and butorphanol 0.01–0.02 mg/kg IV (Morphasol-10 ®, Graeub, Bern, Switzerland). A sterile BALF tube (300 cm; 10 mm outer diameter; Bivona Medical Technologies, Gary, Indiana, USA) was passed into the nose down to the trachea. Twenty ml of lidocaine

2% were instilled and the tube was further inserted until wedged against a peripheral bronchus. The cuff was insufflated with 5 mL of air and 0.9% NaCl (250 mL) was infused using 60-mL syringes. The fluid was re-aspirated until no more was yielded, at which point the cuff was deflated and the tube pulled out. The syringes content was pooled in a cooled silicone-coated glass bottle. The BALF was filtered through a 40-µm Falcon cell strainer and kept on ice until processing.

2.3. Sample Cryopreservation and Thawing

The protocol used to freeze and subsequently thaw the BALF cells has been described in detail in a previous study from our group [15]. The four samples from which RNA was extracted were stored at −80 °C for a median of 321 days (interquartile range [IQR] 50) before thawing.

2.4. RNA extraction and cDNA Library Preparation

Total RNA was extracted with the RNeasy mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The quantity and quality of the isolated RNA was assessed using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit RNA BR Assay Kit (Q10211, Thermo Fisher Scientific, Reinach, Switzerland) and an Advanced Analytical Fragment Analyzer System with a Fragment Analyzer RNA Kit (DNF-471, Agilent, Basel, Switzerland). The RNA was also tested by spectrophotometry using a Denovix DS-11 FX (M/F) spectrophotometer/fluorometer to assess the purity of the RNA (DeNovix, Wilmington, DE, USA). The median RIN for the four samples was 9.85 (IQR 0.13). The two highest quality samples (RIN $\geq$ 9.9), representing one asthmatic and one control horse, were equally pooled based on their RNA concentration.

2.5. PacBio Iso-Seq Long-Read Sequencing

Once all quality control tests confirmed high quality of the pooled RNA sample, the Procedure & Checklist—Iso-Seq Express Template Preparation for Sequel and Sequel II Systems was followed (PN 101-763-800 Version 02 (October 2019)) using the standard protocol (Pacific Biosciences, Menlo Park, CA, USA). In short, full-length cDNA from 300 ng of total RNA was prepared using the NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module Kit. Thereafter, a SMRTbell express template prep kit 2.0 was used to prepare the library. The resulting library was Single Molecule Real-Time (SMRT) sequenced using a Sequel binding plate 3.0, sequel sequencing plate 3.0 with a 20 h movie time on a PacBio Sequel system using a SMRT cell 1M v3 LR. The 2.3 kb library was loaded at 4.5 pM and generated 28.6 Gb and 419,674 polymerase reads. Next, the Circular Consensus Sequencing (CCS) application was run in SMRT Link v10.1 (Pacific Bioscience, Menlo Park, CA, USA) using the continuous long reads sub-read dataset and default parameters. This resulted in 313,563 HiFi reads and 805,344,539 bp of HiFi yield. The HiFi mean read length was 2568 bp and 19 HiFi passes was recoded (mean). These HiFi reads were used to run the Iso-Seq analysis pipeline in SMRT Link v10.1. This generated 317,797 Full-Length Non-Concatamer (FLNC) reads with 5′ and 3′ primers and poly-A tails. The mean length of the FLNC reads was 2536 bp. Reads identified as full-length non-chimeric (FLNC) were considered for de novo clustering to generate unique isoforms. All steps, listed above, including RNA quality control assessments, were conducted at the Next Generation Sequencing Platform, University of Bern.

2.6. Isoform Sequence Analysis

Unique high-quality isoforms (supported by at least two FLNC) were mapped to the equine genome (version EquCab3.0) using minimap2 [19]. The following parameters were used with minimap2: (-ax splice -t 30 -uf –secondary=no -C5). The mapped data in SAM format were annotated using cDNA_Cupcake (https://github.com/Magdoll/cDNA_Cupcake;accessed on 14 February 2022) and SQANTI3 [20] (https://github.com/ConesaLab/SQANTI3; accessed on 14 February 2022) pipelines. The mapped data in SAM

format were input to Cupcake in order to collapse redundant isoforms into transcript loci. 5′-differences were not considered when collapsing the isoforms. 5′-degraded isoforms were removed using the filter_away_subset function of cDNA_cupcake. The resulting isoforms were annotated and compared with EquCab3.0 NCBI annotation release 103 using SQANTI3 with default parameters. The known and novel isoforms were categorized by SQANTI3 into nine different structural classes (see Table 2). A schematic overview of the bioinformatics workflow is shown in Figure 1.

Table 2. Transcript structural classes provided by SQANTI3.

Full splice match (FSM)	Isoforms matching perfectly to annotated transcripts
Incomplete splice match (ISM)	Isoforms matching to a subsection of an annotated transcript
Novel in catalog (NIC)	Isoforms with a new combination of annotated splice sites
Novel not in catalog (NNIC)	Isoforms with at least one novel splice site along with annotated splice sites
Intergenic	Isoforms mapping to intergenic region
Genic intron	Isoforms contained within an intron
Genic genomic	Isoforms overlapping with exons and introns
Fusion	Isoforms spanning two annotated loci
Antisense	Isoforms mapping to the complementary strand of an annotated transcript

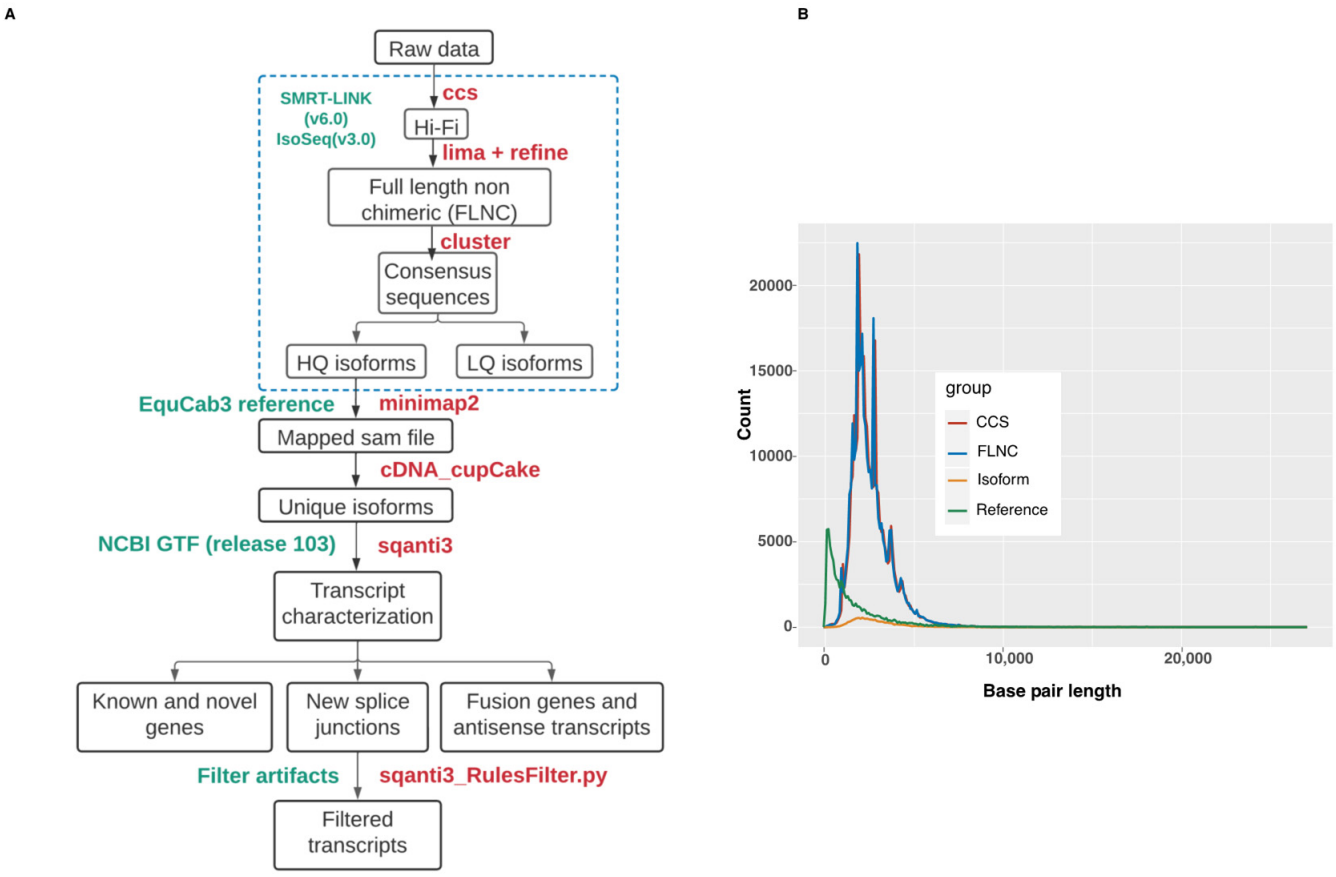


Figure 1. (**A**) Schematic overview of the bioinformatics workflow used to generate full-length transcript annotations. (**B**) Length comparison between reference transcript length and Iso-Seq data.

Template switching during reverse transcription of cDNA is a known source of false non-canonical splice junctions [21,22]. Secondary structures in RNA template or direct repeats can cause reverse transcriptase (RT) to switch from one template to the other, creating gaps in the cDNA sequence. The gaps are interpreted as splicing-like events by algorithms, thereby artificially inflating for non-canonical splice sites. To detect this RT switching event, SQANTI3 looks for repeat sequences flanking the non-canonical splicing exon-intron boundary. The output of SQANTI3 classification was filtered for transcripts that had 1) a junction classified as a RT-template switching artifact and 2) 86% or higher adenosine content in the 20 nucleotides immediately downstream of the aligned 3′-end of the transcript, as well as a continuous run of eight or more A (i.e poly(A)), indicating a possible oligo(dT) intra-priming artifact. These transcripts were filtered using SQANTI3 filtering script sqanti3_Rulesfilter.py.

Saturation-discovery or rarefaction curves were produced by subsampling full-length reads at different depths. Full-length reads were randomly sampled and, for each subsample of reads, the number of unique genes or transcript isoforms detected was determined. For each sampling depth, 100 sampling iterations were performed before computing the average number of unique genes or isoforms observed. Only isoforms exactly matching the NCBI annotation release 103 were retained. The saturation-discovery curve analysis was produced with the 'subsample.py' and 'subsample_with_category.py' scripts available in the cDNA_cupcake repository.

2.7. BUSCO Analysis

Benchmarking Universal Single-Copy Orthologs (BUSCO) [23] looks for near-universal single-copy orthologs present in a whole transcriptome dataset. We used BUSCO to determine the percentage of orthologs present in equine BALF. BUSCO v4.1 was run with its default settings, using the BUSCO vertebrate database (https://busco.ezlab.org/; accessed on 8 June 2022).

2.8. Blastp Analysis

SQANTI3 uses the GeneMarkS-T (GMST) algorithm to predict ORFs from the transcripts. The predicted ORFs were mapped to the clustered nr protein database [24] using blastp with the parameters -max_target_seqs 5 -evalue 0.0001 -outfmt "6 qseqid qaccver qlen sseqid saccver slen length pident qcovs evalue bitscore stitle". The ORFs were also mapped to the PFAM database using the HMMER algorithm [25]. The predicted fusion transcripts were searched against the ConjoinG database [26]

2.9. Gene ontology (GO) Analysis

EnrichR online tool [27]-based gene enrichment analysis was performed with the top 500 most abundantly expressed genes. Total number of full-length read counts associated with genes was used as expression value. This value was obtained from the SQANTI3 classification file. The functional categories examined were: GO_Biological_Process 2021, GO_Cellular_Component 2021, GO_Molecular_Function 2021, Jensen TISSUES and PanglaoDB Augmented. Jensen tissues is a human database of gene–tissue associations. It records the expression of mRNA or corresponding protein in several tissues collected from multiple sources using various data types. PanglaoDB is a single cell RNA-seq database for mouse and human.

3. Results

3.1. Full-Length Transcripts

PacBio Iso-Seq data were generated from pooled RNA isolated from BALF cells of one healthy and one severely asthmatic horse (See Table 1 and S1 for details). We collected 12,487,670 subreads comprising ~51 Gb of raw sequencing data to generate consensus reads with the circular consensus sequencing (CCS) technology (Figure 1A). The CCS dataset consisted of 313,563 Hi-Fi reads with a mean length of 2568 bp (Figure 1B and Table 3)

and a median quality of Q37. Adapter and poly(A) tail removal resulted in 258,022 FLNC sequences. Clustering and polishing of FLNC sequences produced 20,462 high quality (HQ) sequences and 159 low quality consensus transcript sequences. The HQ sequences had an accuracy > 99% and were supported by at least 2 FLNC reads. The HQ sequences had a mean length of 2931 bp.

Table 3. Summary statistics of PacBio Sequel IIe transcript sequencing data.

Sequence Type	Total Number	Min Length	Average Length	Max Length	N50
Polymerase read	419,674	-	68,058	-	126,248
Subread	12,487,670	500	2244	236,862	2427
CCS	313,563	500	2568	15,078	2759
FLNC	258,022	500	2536	15,000	2689
HQ transcripts	20,462	502	2931	10,513	2937

Rarefaction analysis with subsampled full-length reads showed that sequencing saturation for transcript loci discovery was almost attained. Therefore, a higher sequencing depth would have been unlikely to detect more transcripts (Figure S1A,B).

3.2. Annotation of HQ Isoforms

All HQ isoforms were mapped to the EquCab3.0 reference genome. The resulting SAM file from the transcript mapping was used for collapsing redundant transcripts isoforms based on genomic location. The collapsed data showed 17,538 isoforms mapping to unique locations and 599 HQ transcripts that did not map, had low alignment length coverage (<99%) or had low alignment identity (<95%). Each of the 17,538 unique isoforms was supported by at least two full-length reads. Filtering of the 1484 isoforms with truncated 5′-ends left 16,054 isoforms for the next analysis steps.

3.3. Isoform Characterization

The Iso-Seq isoforms aligned to 58.5% (n = 1960) vertebrate orthologs, of which 36.1% (n = 1210) matched single-copy orthologs.

Transcripts with unreliable 3′-ends (intra-priming) and/or with junctions labeled as (i) RT-switch, (ii) without a minimum coverage of three reads were filtered (n = 1820). The numbers of filtered transcripts from each structural category are shown in Figure S2.

The final filtered Iso-Seq dataset consisted of 14,234 transcript isoforms covering 6880 previously annotated and 137 novel genes (Table 4). Of those, 6309 (90.0%) were protein-coding genes. Most of the transcripts identified were protein-coding (98.0%).

Table 4. Overview of the annotated long-read equine BALF cells transcriptome.

Characteristics	Number (Percentage)
Isoforms	14,234
Unique genes	7017
Previously annotated genes	6880 (98.0%)
Novel genes	137 (2.0%)
Genes with >1 isoform	3400 (48.4%)
Genes with >6 isoforms	164 (2.3%)
Known coding transcripts	10,665 (75.0%)
Known non-coding transcripts	141 (1.0%)
Novel coding transcripts	3235 (22.7%)
Novel non-coding transcripts	193 (1.3%)

Majority of transcripts were FSM isoforms (66.5%), while ISM isoforms represented 9.4% of the transcripts. NIC and NNIC accounted for 13.2% and 9.4% of the transcripts, respectively (Table 5). The 35 isoforms of the *CD163* gene transcript are depicted in

Figure 2A as an example. They consisted of four FSM, two ISM, 12 NIC and 17 NNC isoforms. Of the NIC and NNC isoforms, 20 showed intron retention.

Table 5. Structural classes of the 14,234 transcripts identified.

Structural Class	#Isoforms (Percentage)
All	14,234 (100%)
FSM	9473 (66.5%)
ISM	1333 (9.4%)
NIC	1886 (13.2%)
NNC	1341 (9.4%)
Genic genomic	6 (0.04%)
Antisense	38 (0.3%)
Fusion	48 (0.3%)
Intergenic	109 (0.8%)
Genic intron	0 (0%)

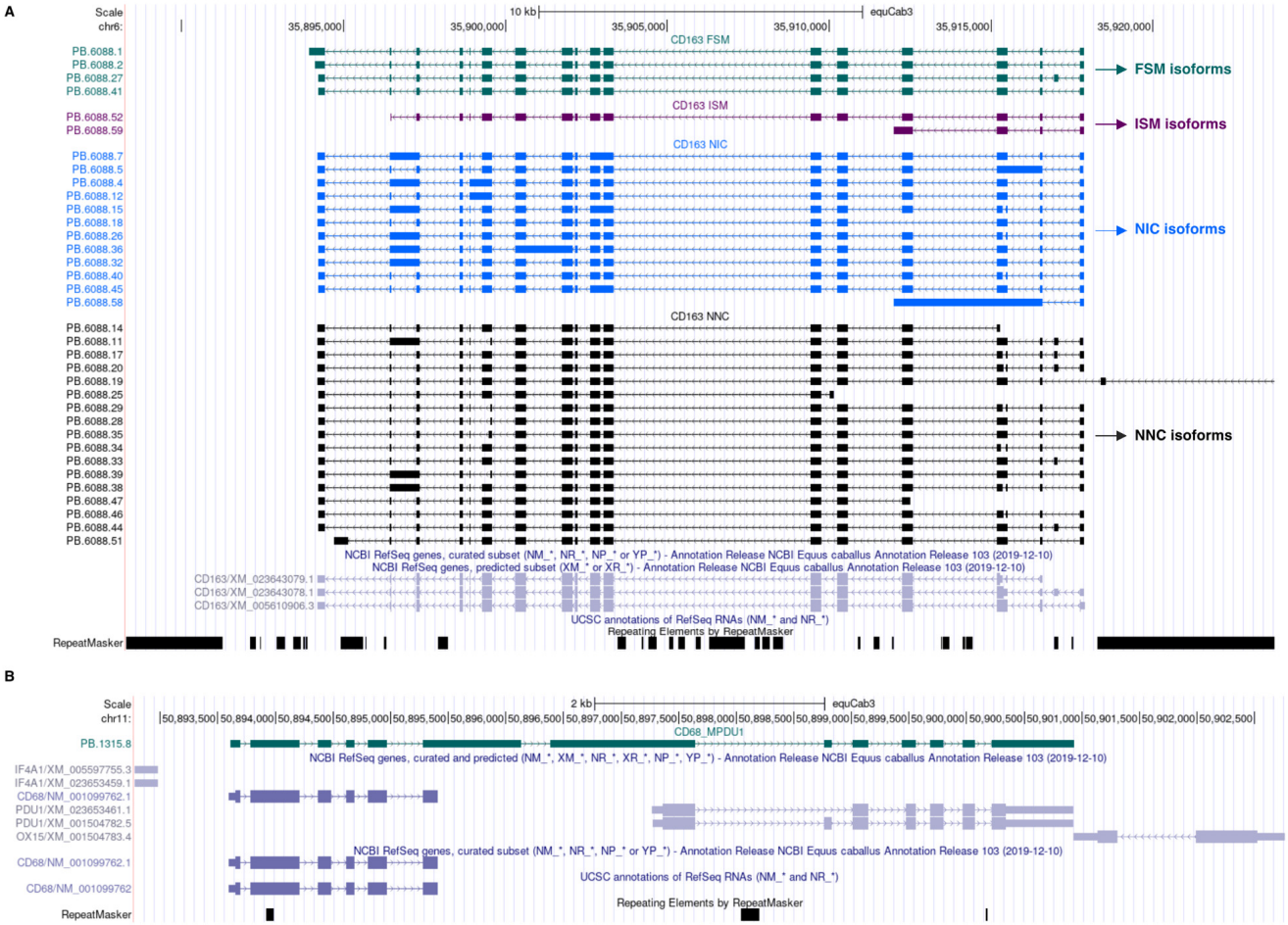


Figure 2. (**A**). UCSC genome browser track of *CD163* in equine BALF cells. (**B**). UCSC genome browser track showing the transcripts of the fusion gene *MPDU1_CD68* (green track). These transcripts cover both the genes *MPDU1* and *CD68*. Their biological significance is unknown at this time.

The full-length transcripts recovered had a length up to 10 kb (Figure 1B). The longest transcript was 10,514 bp in length and mapped to the gene *ABCA1*, which encodes a protein that exhibits ATPase-coupled transmembrane transporter activity. A total of 3372 (49%) genes were multi-exonic and gave rise to more than one isoform. A total of 163 (2.4%) of

the 6880 known genes identified had more than six isoforms. The length of the detected isoforms correlated with the number of exons (r = 0.43).

A small number of transcripts was qualified as antisense [n = 38; 0.3%], intergenic [n = 109; 0.8%] and genic genomic [n = 6; 0.04%]. All isoform categories were mostly represented by coding transcripts (96–98%), except for the antisense and intergenic categories. Antisense and intergenic transcripts had 60% and 75% predicted ORFs, respectively. Nonsense-mediated mRNA decay (NMD) prediction showed that novel transcripts (27% of the NIC and 21% of the NNC) were more likely to be targeted by the NMD pathway than known transcripts (3% of the FSM).

3.4. UTR Extensions

The 5′-end of 46% of the FSM transcripts overlapped completely or almost completely the transcriptional start site (TSS) of the matched reference transcripts (Figure 3A). Similarly, the 3′-end of 51% of the FSM transcripts overlapped completely or almost completely the transcriptional termination site (TTS) of the matched reference transcripts (Figure 3B). Seventeen percent of the FSM and 14% of the ISM transcripts extended beyond the known TSS, while 20% of the FSM and 11% of the ISM transcripts extended beyond the known TTS (Figure 3C,D). The 5′-end of the FSM transcripts were extended by an average of 72 bp (max: 3115 bp). The 3′-end of the FSM transcripts were extended by an average of 505 bp (max: 14,447 bp).

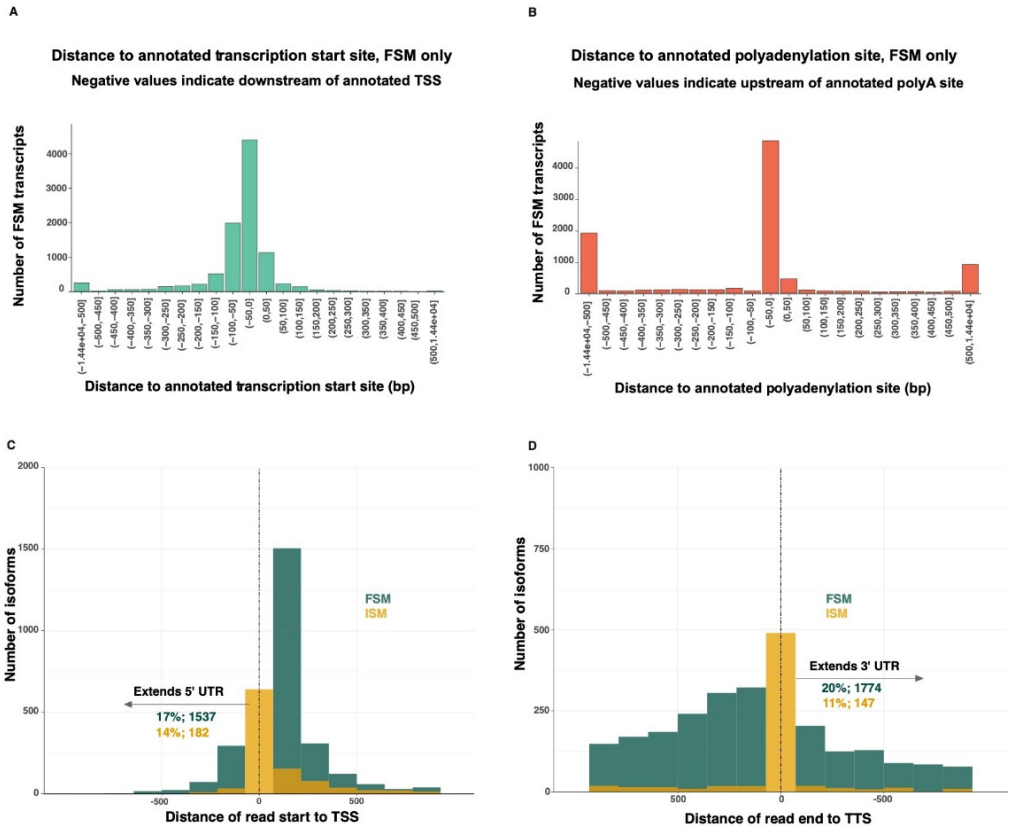


Figure 3. (**A**). Binned distances between the Iso-Seq transcription start sites (TSS) and the corresponding annotated TSS (FSM transcripts only). (**B**). Binned distances between the Iso-Seq transcript poly(A) site and the corresponding annotated poly(A) site (FSM transcripts only). (**C**). In contrast to FSM reads, most ISM reads do not extend beyond the 5′-UTRs of known transcripts. (**D**). FSM and ISM reads extending beyond the 3′-UTRs of known transcripts. TSS: transcriptional start site. TTS: transcriptional termination site.

3.5. Characterization of Splice Junctions

Novel genes were enriched for multi-exon transcripts, compared to known genes (Figure S3A). Of these multi-exonic novel genes, 44.3% (4702) were non-coding. The number of exons per isoform structural class was similar for both known and novel transcripts (Figure S3B). Only 623 transcripts were classified as mono-exonic. Their length was comparable to multi-exon transcripts (Figure S3C). Almost all splice junctions were canonical (GT-AG, GC-AG or AT-AC dinucleotide pairs) (Table 6). The ten most frequent splice junctions found in the Iso-Seq transcripts are shown in Figure 4A.

Table 6. Splice junction classification.

Splice Junction Category	Number (Percentage)
Known canonical	67,607 (96.9%)
Known non-canonical	30 (0.04%)
Novel canonical	2171 (3.1%)
Novel non-canonical	0 (0.0%)

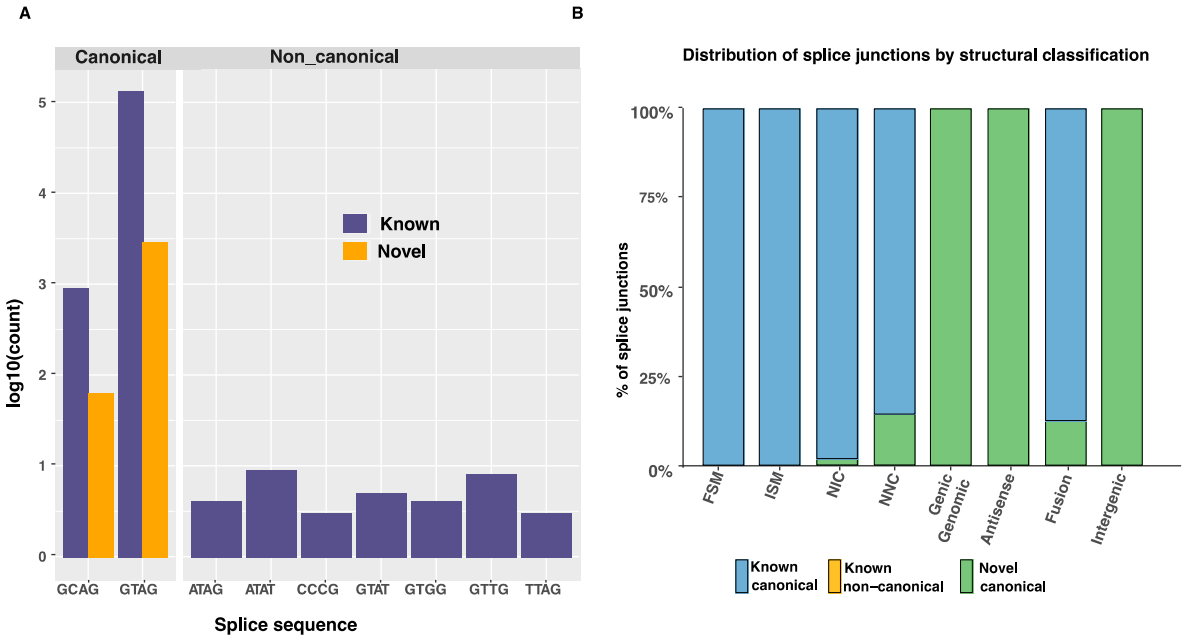


Figure 4. (**A**). Top ten splice sites identified in the known (previously annotated) and novel transcript isoforms. (**B**). Distribution of splice junction types across transcript structural classes.

Canonical splice junctions from previously annotated genes accounted for 96.89% of all splice junctions (Table 6). New combinations of donor and acceptor splicing sites accounted for 0.7% of the isoforms only. Many of those were found in the novel isoform category, as illustrated in Figure 4B. The top ten splice junctions are shown in Figure 4A.

3.6. Novel Isoforms

Novel isoforms comprised 24% of all transcripts. While these 3428 transcripts were classified as novel, 1886 (55%) could still be linked to previously annotated genes because they shared splice junctions with known gene transcripts. These were classified as NIC. The remaining novel transcripts, predominantly classified as NNC, had at least one novel donor or acceptor site (1341 [40.8%] of all novel transcripts]). Intron retention was observed in 34% of the novel isoforms.

The novel transcripts contained 3428 ORFs with a length $\geq$ 100 amino acids. A total of 105 protein-coding transcripts mapping to 147 novel genes were predicted. These protein-coding ORFs from novel genes had an average length of 327 amino acids. Fifty-four of

the predicted proteins matched proteins in the nr database with more than 85% identity and 80% coverage (Table S2). Table S2 shows the similarity between the novel transcripts identified and the orthologous gene transcripts. These transcripts are absent from the EquCab3.0 NCBI annotation release 103.

3.7. Fusion Genes

Forty-eight fusion transcripts were associated with 30 previously annotated genes, most of them originating from read-through transcription of two adjacent genes. The fusion gene *MPDU1_CD68* is displayed in Figure 4B as an example. The majority of fusion transcripts encompassed only two genes. ORFs were predicted in 45 of them. Twenty-six ORFs spanned the coding sequence of the two loci they overlapped. An NMD effect was predicted in 12 fusion transcripts. We also identified 13 fusion transcripts as potential "conjoined genes" based on their sequence matching with the ConjoinG database [20] (Table S2). Conjoined genes give rise to transcripts combining exons from two or more distinct genes lying on the same strand of a chromosome. The conjoined genes found in our dataset were supported by at least one mRNA or expressed sequence tag (EST) sequence.

3.8. Gene Ontology Analysis

The 500 most expressed genes in our dataset were predominantly involved in neutrophil degranulation (GO:0043312), neutrophil activation (GO:0002283) and neutrophil-mediated immunity (GO:0002446). Tissue ontology analysis (Jensen database [27]) showed these genes were overexpressed in trachea, blood and monocytes. Gene ontology analysis using the PanglaoDB database [27] suggested a high expression in macrophages, monocytes, alveolar macrophages and dendritic cells. The enrichment table is detailed in Table S2.

4. Discussion

We used PacBio long-read Iso-Seq RNA-seq to characterize full-length cDNA sequences in equine BALF cells. We were able to identify and map 14,234 transcripts with considerable isoform diversity, of which 4449 were not represented in the current equine genome NCBI annotation release 103. This highlights the need for a more comprehensive and tissue-specific transcriptome annotation in horses.

The coding transcripts identified in our Iso-Seq dataset covered 60% of the 21,129 coding transcripts from the EquCab3.0 NCBI annotation release 103. The levels of gene expression appear to be tissue dependent: using PacBio sequencing in chicken, Kuo et al. [10] identified 211,292 transcripts from the brain but only 14,776 transcripts from the embryo. Hence, our incomplete transcriptome representation most likely reflects a BALF cell-specific gene expression. The results of the gene ontology analyses using tissue (Jensen database) and cell type (PanglaoDB) databases also support this hypothesis.

Our dataset comprised 70% of FSM transcripts, indicating that RNA degradation was minimal. Still, ISM isoforms comprised 9.4% of the Iso-Seq transcripts associated with known genes, similar to the 8.7% reported in the equine FAANG transcriptome [14]. ISM transcripts were supported by reference transcripts with missing 5′- or 3′-exons. These ISM transcripts may represent technical artifacts, for example due to RNA degradation or incorrect priming, or real additional isoforms. These transcripts all had CCS sequences with a median ORF length of 450 bp, indicating they could encode proteins. To which extent these transcripts contribute to the BALF proteome needs to be further explored. A high content of adenosine nucleotides in the RNA strands can lead to mispriming during reverse transcription. However, only 241 of the 1333 ISM reads had 80% or more A nucleotides in the 20 bases flanking their mapped position. Almost all splice sites detected in the ISM transcripts were canonical, making false splice site predictions unlikely. Indeed, splice sites are easily predicted, since 98.7% of the canonical splice sites in mammalian genomes are known to be GT-AG [28]. In summary, ISM transcripts identified with Iso-Seq RNA-seq likely represented biologically true novel transcript isoforms.

The isoform-to-gene ratio of our Iso-Seq dataset was 2.2 when considering protein-coding genes only. This is comparable to the existing NCBI and Ensembl EquCab3.0 reference annotations with a ratio of 2.6 and 2.1, respectively. In contrast, the equine FAANG transcriptome [14] achieved an isoform-to-gene ratio of 4.2 by combining the two existing annotations with an Iso-Seq transcriptome. Peng et al. [14] used an interesting approach where the transcripts only differing at their 5′-ends were collapsed into a single transcript. They then used short-read and ATAC-seq data from the same tissues to assess the completeness of the 5′-end of the isoforms. In the absence of other type of datasets to compare to, we filtered out the transcripts with differing 5′-ends, which may have reduced the number of isoforms recovered.

The GO analysis using the Enrichr database highlighted biological processes that are closely related to the pathophysiology of equine asthma. One of the two horses sampled for this experiment was affected with neutrophilic severe equine asthma. This most likely explains the enrichment of genes associated with neutrophilic inflammation in our dataset. Our Iso-Seq transcripts will thus be valuable when interpreting RNA-seq data originating from asthmatic horses. In future experiments, it would be interesting to extend the panel of transcripts by sampling horses with other forms of asthma, such as mild/moderate mastocytic or eosinophilic equine asthma.

A small proportion (0.3%) of the transcripts were fusion transcripts, most of them due to read-through transcription of two adjacent genes. Peng et al. [14] similarly found them in small numbers (1.21% of all transcripts). The biological significance of fusion transcripts remains unclear. Their hits to the conjoined genes database suggest a functional role, as several of these conjoined genes are conserved among vertebrates [26].

While this study identified many equine BALF cell transcripts not referenced in the current EquCab3.0 NCBI annotation release 103, it also had several limitations. First, the Iso-Seq RNA-seq protocol entails poly(A) selection, resulting in a 3′-bias. Complementary techniques would be required to capture non-polyadenylated RNA such as ribosomal RNA. Consequently, our dataset contains mostly protein coding transcripts and polyadenylated long non-coding RNA transcripts (lncRNAs).

Another potential limitation of our experiment is the sequencing depth. While the sequencing saturation was appropriate for gene detection, it may not have been sufficient to capture the full extent of isoform diversity (see Figure S3). We propose to integrate this equine BALF-specific transcriptome to the recently built equine FAANG transcriptome [14]. In future studies, it will be beneficial to sequence RNA from additional tissues to cover a wider range of cell types and/or developmental stages with potentially different expression profiles [10,29].

Eventually, we acknowledge that the complexity of the different isoforms for some genes is daunting. It is difficult to determine whether this reflects the complexity of equine BALF cell biology, or if it arises from technical artifacts inherent to the Iso-Seq sequencing technology. The replication of long-read sequencing experiments in horses will make it possible to cross-check the data and to disentangle technical artifacts from biological reality. The combination of long-read sequencing with single cell sequencing should enable the generation of high-quality gene expression profiles at a superior resolution.

5. Conclusions

In summary, this study demonstrates the potential of long-read sequencing to improve the annotation of the equine transcriptome by providing full-length transcripts at the isoform level. Here, we provide a BALF cell-specific transcriptome that will be useful for future equine bulk or single cell RNA-seq studies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13101722/s1, Table S1: Clinical characteristics of horses used in the study; Table S2 (Sheet 1) SQANTI3 classification of transcripts; Table S2 (Sheet2): nr protein database results; Table S2 (Sheet 3): Alignment results of fusion genes again ConjoinG database; Table S2 (Sheet 4, 5, 6, 7,8): Enrichment of GO categories top 100 gene sets, Jensen tissue database and

Pangloa database; Figure S1. Saturation-discovery curve; Figure S2: Count of filtered isoforms from different structural categories; Figure S3: A. Exon distribution in novel and known genes. B. Exon counts in different structural category of transcripts. C. Length of mono- and multi-exonic transcripts. D Length of transcripts classified by structural category; Figure S4: RT-Switching templates predicted by SQANTI3 in the different isoform classes.

Author Contributions: Conceptualization, V.G., T.L., P.N. and S.E.S.; methodology, V.J, P.N. and S.E.S.; sample collection and processing, S.E.S.; software, V.J. and P.N.; validation, V.G., T.L. and S.E.S.; formal analysis, V.J.; investigation, V.J.; resources, V.G. and S.E.S.; data curation, V.J. and S.E.S.; writing—original draft preparation, V.J.; writing—review and editing, V.J., T.L., V.G., P.N. and S.E.S.; visualization, V.J.; supervision, T.L. and V.G.; project administration, V.G.; funding acquisition, V.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swiss National Science Foundation (Grant No. 31003A-162548/1) and the Internal Research Fund of the Swiss Institute of Equine Medicine, Bern, Switzerland (ISMEquine Research No. 33-890).

Institutional Review Board Statement: All animal experiments were performed according to the local regulations and with the consent of the horse owners. This study was approved by the Animal Experimentation Committee of the Canton of Bern, Switzerland (BE4/20+).

Informed Consent Statement: Informed written consent was obtained from the horse owners.

Data Availability Statement: The data presented in this study are available online at: https://www.ebi.ac.uk/ena/browser/view/PRJEB51962; accessed on 14 Februry 2022, (run accession: ERR9954028). The code used for the analysis and the final read annotation can be found at https://github.com/vetsuisse-unibe/EquCab3-Iso-seq-BALF-paper; accessed on 14 Februry 2022.

Acknowledgments: We thank the Next Generation Sequencing Platform of the University of Bern for performing the high-throughput experiments and the Interfaculty Bioinformatics Unit of the University of Bern for providing high-performance computing infrastructure.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Couetil, L.; Cardwell, J.M.; Leguillette, R.; Mazan, M.; Richard, E.; Bienzle, D.; Bullone, M.; Gerber, V.; Ivester, K.; Lavoie, J.P.; et al. Equine Asthma: Current Understanding and Future Directions. *Front. Vet. Sci.* **2020**, *7*, 450. [CrossRef]
2. Aun, M.V.; Bonamichi-Santos, R.; Arantes-Costa, F.M.; Kalil, J.; Giavina-Bianchi, P. Animal Models of Asthma: Utility and Limitations. *J. Asthma Allergy* **2017**, *10*, 293–301. [CrossRef]
3. Leclere, M.; Lavoie-Lamoureux, A.; Lavoie, J.P. Heaves, an Asthma-like Disease of Horses. *Respirology* **2011**, *16*, 1027–1046. [CrossRef]
4. Bullone, M.; Lavoie, J.P. Asthma "of Horses and Men"—How Can Equine Heaves Help Us Better Understand Human Asthma Immunopathology and Its Functional Consequences? *Mol. Immunol.* **2015**, *66*, 97–105. [CrossRef]
5. Pacholewska, A.; Jagannathan, V.; Drögemüller, M.; Klukowska-Rötzler, J.; Lanz, S.; Hamza, E.; Dermitzakis, E.T.; Marti, E.; Leeb, T.; Gerber, V. Impaired Cell Cycle Regulation in a Natural Equine Model of Asthma. *PLoS ONE* **2015**, *10*, e0136103. [CrossRef]
6. Tessier, L.; Côté, O.; Clark, M.E.; Viel, L.; Diaz-Méndez, A.; Anders, S.; Bienzle, D. Impaired Response of the Bronchial Epithelium to Inflammation Characterizes Severe Equine Asthma. *BMC Genom.* **2017**, *18*, 708. [CrossRef]
7. Hulliger, M.F.; Pacholewska, A.; Vargas, A.; Lavoie, J.P.; Leeb, T.; Gerber, V.; Jagannathan, V. An Integrative MiRNA-MRNA Expression Analysis Reveals Striking Transcriptomic Similarities between Severe Equine Asthma and Specific Asthma Endotypes in Humans. *Genes* **2020**, *11*, 1143. [CrossRef]
8. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [CrossRef] [PubMed]
9. Bayega, A.; Fahiminiya, S.; Oikonomopoulos, S.; Ragoussis, J. Current and Future Methods for MRNA Analysis: A Drive Toward Single Molecule Sequencing. *Methods Mol. Biol.* **2018**, *1783*, 209–241. [CrossRef]
10. Kuo, R.I.; Tseng, E.; Eory, L.; Paton, I.R.; Archibald, A.L.; Burt, D.W. Normalized Long Read RNA Sequencing in Chicken Reveals Transcriptome Complexity Similar to Human. *BMC Genom.* **2017**, *18*, 323. [CrossRef]
11. Chen, S.Y.; Deng, F.; Jia, X.; Li, C.; Lai, S.J. A Transcriptome Atlas of Rabbit Revealed by PacBio Single-Molecule Long-Read Sequencing. *Sci. Rep.* **2017**, *7*, 7648. [CrossRef] [PubMed]
12. Beiki, H.; Liu, H.; Huang, J.; Manchanda, N.; Nonneman, D.; Smith, T.P.L.; Reecy, J.M.; Tuggle, C.K. Improved Annotation of the Domestic Pig Genome through Integration of Iso-Seq and RNA-Seq Data. *BMC Genom.* **2019**, *20*, 344. [CrossRef]

13. Giuffra, E.; Tuggle, C.K. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annu. Rev. Anim. Biosci.* **2019**, *7*, 65–88. [CrossRef]
14. Peng, S.; Dahlgren, A.R.; Hales, E.N.; Barber, A.M.; Kalbfleisch, T.; Petersen, J.L.; Bellone, R.R.; Mackowski, M.; Cappelli, K.; Capomaccio, S.; et al. Long-Read RNA Sequencing Improves the Annotation of the Equine Transcriptome. *bioRxiv* **2022**. [CrossRef]
15. Sage, S.E.; Nicholson, P.; Peters, L.M.; Leeb, T.; Jagannathan, V.; Gerber, V. Single-Cell Gene Expression Analysis of Cryopreserved Equine Bronchoalveolar Cells. *Front. Immunol.* **2022**, *2022*, 4877. [CrossRef]
16. Ramseyer, A.; Gaillard, C.; Burger, D.; Straub, R.; Jost, U.; Boog, C.; Marti, E.; Gerber, V. Effects of genetic and environmental factors on chronic lower airway disease in horses. *J. Vet. Intern. Med.* **2007**, *21*, 149–156. [CrossRef]
17. Lavoie, J.-P.; Bullone, M.; Rodrigues, N.; Germim, P.; Albrecht, B.; von Salis-Soglio, M. Effect of different doses of inhaled ciclesonide on lung function, clinical signs related to airflow limitation and serum cortisol levels in horses with experimentally induced mild to severe airway obstruction. *Equine Vet. J.* **2019**, *51*, 779–786. [CrossRef]
18. Gerber, V.; Straub, R.; Marti, E.; Hauptman, J.; Herholz, C.; King, M.; Imhof, A.; Tahon, L.; Robinson, N.E. Gerber V, Straub R, Marti E, et al. Endoscopic scoring of mucus quantity and quality: Observer and horse variance and relationship to inflammation, mucus viscoelasticity and volume. *Equine Vet. J.* **2004**, *36*, 576–582. [CrossRef]
19. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef]
20. Tardaguila, M.; de La Fuente, L.; Marti, C.; Pereira, C.; Pardo-Palacios, F.J.; del Risco, H.; Ferrell, M.; Mellado, M.; Macchietto, M.; Verheggen, K.; et al. SQANTI: Extensive Characterization of Long-Read Transcript Sequences for Quality Control in Full-Length Transcriptome Identification and Quantification. *Genome Res.* **2018**, *28*, 396–411. [CrossRef]
21. Cocquet, J.; Chong, A.; Zhang, G.; Veitia, R.A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **2006**, *1*, 127–131. [CrossRef] [PubMed]
22. Houseley, J.; Tollervey, D. Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. *PLoS One* **2010**, *5*, e12271. [CrossRef] [PubMed]
23. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* **2019**, *1962*, 227–245. [CrossRef] [PubMed]
24. Bateman, A.; Martin, M.J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; Bursteinas, B.; et al. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]
25. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [CrossRef] [PubMed]
26. Prakash, T.; Sharma, V.K.; Adati, N.; Ozawa, R.; Kumar, N.; Nishida, Y.; Fujikake, T.; Takeda, T.; Taylor, T.D. Expression of Conjoined Genes: Another Mechanism for Gene Regulation in Eukaryotes. *PLoS ONE* **2010**, *5*, e13284. [CrossRef] [PubMed]
27. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; van Meirelles, G.; Clark, N.R.; Ma'ayan, A. Enrichr: Interactive and Collaborative HTML5 Gene List Enrichment Analysis Tool. *BMC Bioinform.* **2013**, *14*, 128. [CrossRef] [PubMed]
28. Burset, M.; Seledtsov, I.A.; Solovyev, V.V. SpliceDB: Database of Canonical and Non-Canonical Mammalian Splice Sites. *Nucleic Acids Res.* **2001**, *29*, 255–259. [CrossRef] [PubMed]
29. Tseng, E.; Underwood, J.G.; Evans Hutzenbiler, B.D.; Trojahn, S.; Kingham, B.; Shevchenko, O.; Bernberg, E.; Vierra, M.; Robbins, C.T.; Jansen, H.T.; et al. Long-Read Isoform Sequencing Reveals Tissue-Specific Isoform Expression between Active and Hibernating Brown Bears (Ursus Arctos). *G3 Genes Genomes Genet.* **2022**, *2022*, 12. [CrossRef] [PubMed]

2.3 Study 3: Single-cell profiling of bronchoalveolar cells reveals a Th17 signature in neutrophilic severe equine asthma

Authors: **Sophie E. Sage**, Tosso Leeb, Vidhya Jagannathan* and Vinzenz Gerber*

*These authors contributed equally to this work.

Status: published as preprint on bioRxiv.

NB: version provided below is the one submitted to a peer-reviewed scientific journal.

Personal contribution:

- Study design
    - Elaboration of clinical scoring sheets
    - Optimization of experimental protocol
- Sample collection with the assistance of Dr. med. vet. Michelle Wyler and Dr. med. vet. Nicole Altermatt
- Laboratory work
    - Optimization of cryopreservation protocol
    - Sample processing before cryopreservation
    - Sample processing before scRNA-seq
- Data analysis
    - Downstream analysis using R, with the assistance of Dr. Vidhya Jagannathan
    - Cell cluster annotation
    - Interpretation of the differential gene expression analysis
- Manuscript
    - Draft preparation
    - Figures
    - Finalization and submission
    - Corresponding author

aSingle-cell profiling of bronchoalveolar cells reveals a Th17 signature in neutrophilic severe equine asthma

Sophie E. Sage[1*], Tosso Leeb[2,3], Vidhya Jagannathan[2¶], Vinzenz Gerber[1¶]

[1] Swiss Institute of Equine Medicine, Department of Clinical Veterinary Medicine, Vetsuisse Faculty, University of Bern; Bern, Switzerland.
[2] Institute of Genetics, Vetsuisse Faculty, University of Bern; Bern, Switzerland
[3] Next Generation Sequencing Platform, University of Bern, Bern, Switzerland

[*] Corresponding author: sophie.sage@unibe.ch
[¶]These authors contributed equally to this work.

Conflict of Interest

Authors declare that they have no competing interests.

Funding

Swiss National Science Foundation, Grant No. 31003A-162548/1; Internal Research Fund of the Swiss Institute of Equine Medicine, Bern, Switzerland, No. 33-890

Ethics approval

All animal experiments were performed according to the local regulations and with the consent of horse owners. This study was approved by the Animal Experimentation Committee of the Canton of Bern, Switzerland (BE4/20+).

Data Availability Statement

The datasets generated for this study is available in the European Nucleotide Archive (ENA) repository https://www.ebi.ac.uk, under the accession number PRJEB51962. The R code used for data analysis can be found on https://github.com/vetsuisse-unibe/ScRNA-seq_BALF_SEA.

Abstract

Severe equine asthma (SEA) is a complex respiratory condition characterized by chronic airway inflammation. It shares many clinical and pathological features with human neutrophilic asthma, making it a valuable model for studying this condition. However, the immune mechanisms driving SEA have remained elusive. Although SEA has been primarily associated with a Th2 response, there have also been reports of Th1, Th17, or mixed mediated responses. To uncover the elusive immune mechanisms driving SEA, we performed single-cell mRNA sequencing (scRNA-seq) on cryopreserved bronchoalveolar cells from 11 Warmblood horses, five controls and six with SEA. We identified six major cell types, including B cells, T cells, monocytes-macrophages, dendritic cells, neutrophils, and mast cells. All cell types exhibited significant heterogeneity, with previously identified and novel cell subtypes. Notably, we observed monocyte-lymphocyte complexes and detected a robust Th17 signature in SEA, with CXCL13 upregulation in intermediate monocytes. Asthmatic horses exhibited expansion of the B cell population, Th17 polarization of the T cell populations, and dysregulation of genes associated with T cell function. Neutrophils demonstrated enhanced migratory capacity and heightened aptitude for neutrophil extracellular trap formation. These findings provide compelling evidence for a predominant Th17 immune response in neutrophilic SEA, driven by dysregulation of

monocyte and T cell genes. The dysregulated genes identified through scRNA-seq have potential as biomarkers and therapeutic targets for SEA and provide insights into human neutrophilic asthma.

1 Introduction

Equine asthma is a common respiratory disease of the horse characterized by bronchoconstriction, mucus production and bronchospasm.[1] Its severe form, severe equine asthma (SEA), presents with increased breathing effort at rest, airway remodeling and in most cases, airway neutrophilia.[1] Equine asthma is an active field of research, in part because of its negative impact on animal welfare and the horse industry but also due to its similarities with human asthma, making it a unique natural animal model.[2–4] In contrast to murine models with experimentally induced airway inflammation, horses develop asthma under natural conditions. Their longer lifespan enables the study of disease progression, particularly airway remodeling. Furthermore, their size facilitates collection of lower airway samples. For instance, collection of bronchoalveolar lavage fluid (BALF) is a routine procedure in horses, in contrast to humans and conventional laboratory animal models. Although promising asthma drugs have been identified based on murine studies, their limited clinical efficacy when applied to humans[4] may be attributed to disparities in the underlying pathophysiological mechanisms between experimentally induced and naturally occurring diseases.

In humans, asthma is considered an umbrella diagnosis encompassing a plethora of diseases with distinct pathophysiologic mechanisms (so-called endotypes). The advent of omics technologies has begun to unveil the diversity of human asthma endotypes.[5,6] SEA shares clinical and pathological features with several human asthma endotypes, including allergic, non-allergic and late-onset asthma.[2] As horse are exposed to high levels of dust in stables, they represent an ideal model for organic dust-induced asthma of agricultural workers.[7] While SEA has been mainly attributed to a Th2 response, there have also been reports of predominant Th1 and mixed Th1/Th2 phenotypes.[8] Furthermore, the Th17 pathway, typically associated with autoimmune diseases, has been implicated.[9–12] The complexity of the disease and limitations of experimental techniques may have contributed to these inconsistent findings. To address this knowledge gap, we leveraged the emerging single-cell mRNA sequencing (scRNA-seq) technology to dissect the immune mechanisms of SEA at the single-cell level.

In a previous experiment, we demonstrated that scRNA-seq can be successfully applied to fresh frozen equine BALF cells.[13] Here, we employed the scRNA-seq technology to characterize BALF cells from six horses with SEA and five control horses.

2 Materials and methods

In this observational case-control study, we recruited SEA-affected horses and controls based on their medical history. We selected six asthmatic and six control horses, using BALF quality, history of respiratory signs and BALF neutrophilia as inclusion criteria. We performed 10X Genomics 3'-end scRNA-seq on ~6000 cryopreserved bronchoalveolar cells per horse. One control horse was excluded due to low cell number and quality, leaving 11 horses for the data analysis. Our objectives were to assess the effect of SEA on i) the distribution of cell types and cell subtypes in the BALF and ii) the differential gene expression (DGE) within each of the cell types/subtypes identified (refer to supplementary materials and methods for details).

3 Results

3.1 Single-cell landscape of bronchoalveolar lavage fluid from asthmatic and control horses

We analyzed the scRNA-seq data obtained from the BALF cells collected from six asthmatic and five control horses. Characteristics of the study population are listed in Table 1.

Unsupervised clustering identified 19 distinct cell clusters (Fig. 1A). Through automated annotation using the top ten differentially expressed genes (DEGs) derived from major cell types identified in our pilot study,[13] we successfully predicted the identity of 99.6% of the cells. Cell cluster identities were validated using the expression of known canonical markers and the top DEGs specific to each cell group (Figure 1B). Subsequently, the cell clusters were consolidated into six major cells groups: B cells, dendritic cells (DC), mast cells, monocytes-macrophages (Mo/Ma), neutrophils and T cells. To explore the diversity of each major cell type, we re-analyzed them independently. We identified three three distinct B cell clusters (Figure 2A), three neutrophil clusters (Figure 3A), seven T cell clusters (Figure 4A), six Mo/Ma clusters (Figure 5A) and four DC clusters (Figure 6A). The mast cell population remained homogenous, without convincing subclustering. Cell clusters were annotated based on the calculated marker genes (Tables S2 – S8) and the expression of cell type marker genes (Figures 1B, 2B, 3B, 4B, 4C, 5B and 6C). The supplementary results provide supporting evidence for the annotation.

3.2 The BALF of asthmatic horses is enriched in B cells but specifically depleted in activated plasma cells

As expected, asthmatic horses showed a significantly higher proportion of neutrophils compared to the control horses (Table 2, Figures 1D and 1E). A novel finding was the B cell enrichment in the BALF of asthmatic horses (Figures 1D and 1E). Asthmatic horses exhibited approximately three times fewer activated plasma cells (B2 cluster) than control horses (Table 2, Figures 2D and 2E). This suggests that expansion of the naïve B cells and non-switched plasma cells primarily contributed to the increased B cell proportion in asthmatic horses. No significant differences were observed between asthmatic and control groups for other major cell types or subtypes (Table 2 and Figures 1E, 2E, 4F, 5E and 6F).

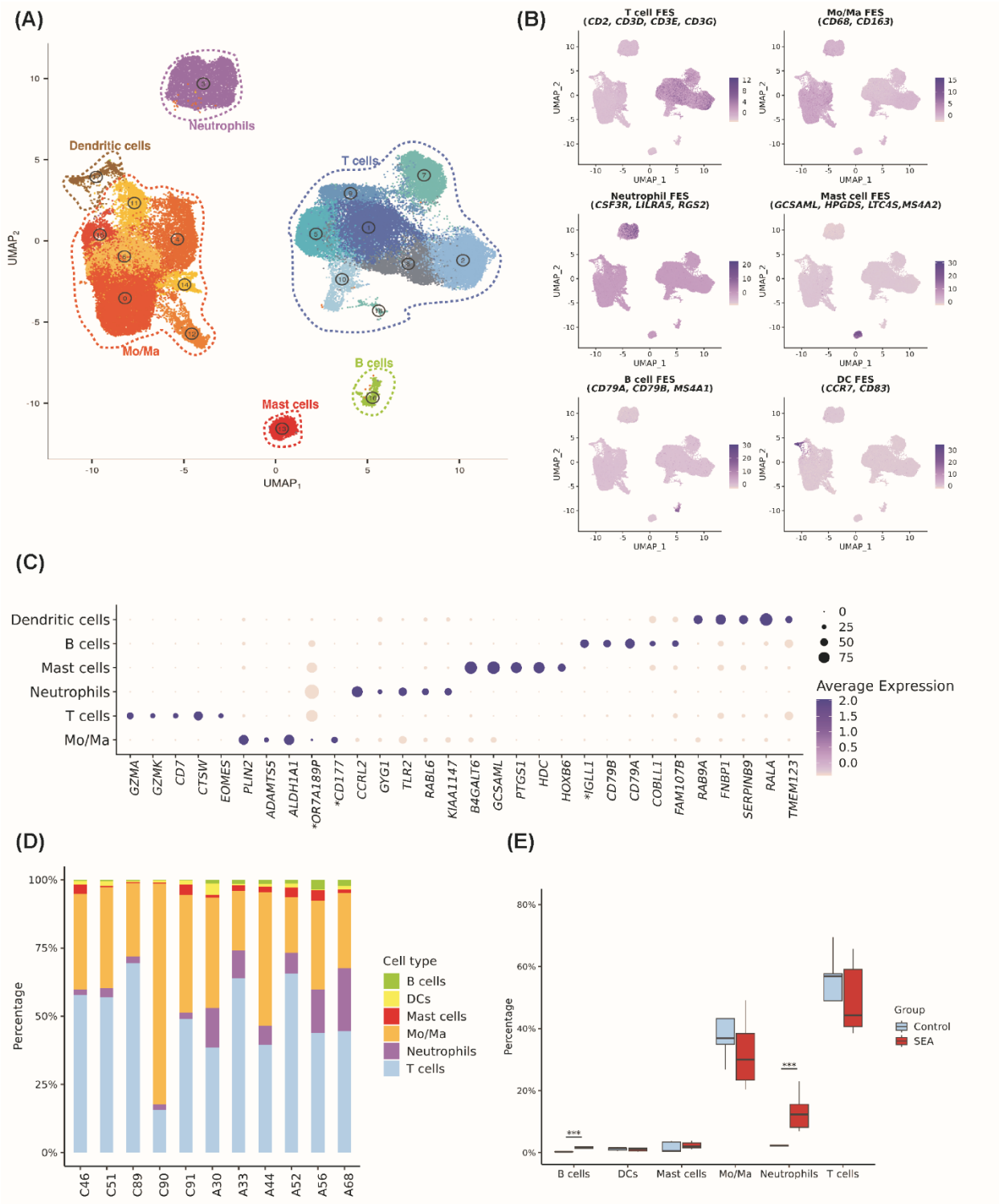


Figure 1: Major cell types identified in the BALF of asthmatic and control horses using scRNA-seq. (A) UMAP representation of the 19 clusters identified as six major cell types. *Mo/Ma, monocyte-macrophage.* **(B)** Gene expression patterns of cell type canonical markers. *DC, dendritic cell; FES, feature expression score.* **(C)** Top five differentially expressed genes per major cell type (one non-coding gene removed). *NCBI 103 annotations for *LOC100146200*: *OR7A189P*, *LOC100069985*: *CD177, LOC102147726*: *IGLL1*. **(D)** Distribution of the six major cell types in asthmatic and control

horses. **(E)** Number of cells from each major cell type in the asthmatic and control groups. *SEA, severe equine asthma*. ***, *P*-value < 0.001.

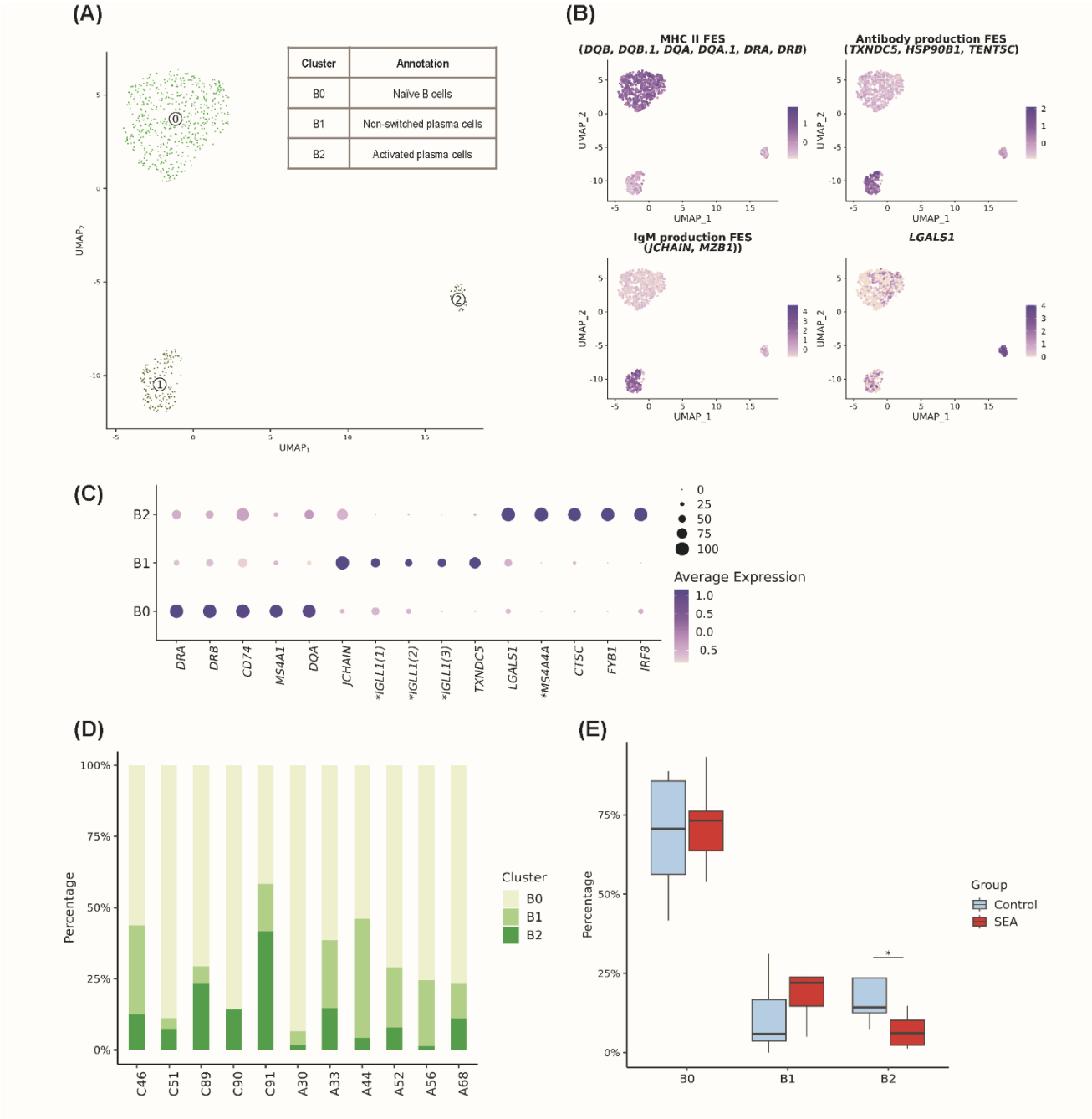


Figure 2: B cell subtypes identified in the BALF of asthmatic and control horses using scRNA-seq. (A) UMAP representation of the three clusters identified. **(B)** Gene expression patterns used for annotation. *FES, feature expression score.* **(C)** Top five differentially expressed genes per cluster. *NCBI 103 annotations for *LOC111774805*, *LOC100060608* and *LOC102147726*: *IGGL1*, for *LOC100061331*: *MS4A4A*. **(D)** Distribution of the clusters among asthmatic and control horses. **(E)**

Number of cells from each B cell cluster in the asthmatic and control groups. *SEA, severe equine asthma.* **P*-value < 0.05.

3.3 Gene expression profile of neutrophils indicates altered NETosis and migratory function in SEA

Using a mixed model approach, we compared gene expression between asthmatic and control horses within each cell type and subtype. Tables S9 – S13 provide the results of the DGE analysis, in which positive log fold changes indicate upregulation in the SEA group.

Neutrophils exhibited an "asthma signature" characterized by upregulation of *CHI3L1* and *MAPK13*, known markers of neutrophilic asthma in humans,[14–16] and downregulation of *SLC7A11*, an indicator of ferroptosis reduced in neutrophilic mice asthma.[17] Apoptotic neutrophils (Neu0) had upregulated *S100A9* and *RETN*, both involved in NETosis function.[18,19] Pro-inflammatory neutrophils (Neu1) had downregulated *KLF2*. Reduced KLF2 levels can promote neutrophil migration[20] and exacerbate NET-induced transfusion-related acute lung injury.[21] In the ISG^{high} neutrophils (Neu2), we observed upregulation of *ADGRE5*, also known as *CD97*, which may promote migration of ISG^{high} neutrophils to the lungs.[22]

Gene expression features with a potential protective effect on the lower airways were also identified. The antileukoproteinase gene *SLPI* was upregulated in asthmatic horses, which has an anti-inflammatory role by inhibiting the NFκB pathway and preventing excess NET formation.[23] *NFKB1* was indeed downregulated in neutrophils from asthmatic horses. Downregulation of *IL17RC* suggested a reduced capacity to respond to the Th17 cytokine IL17. The predominant contributor of protective features was the apoptotic neutrophil subtype, with upregulation of *SLPI* and downregulation of *CCL20* and *NR4A3*. The Th17-associated cytokine CCL20 is a potent chemotactic factor for lymphocytes and DCs. The downregulation of *CCL20* could thus have an overall anti-inflammatory effect, with reduced chemotaxis of immune cells and reduced Th17-signalling. *NR4A3* positively regulates neutrophil survival.[24] Hence, its downregulation may mitigate neutrophil persistence in the lungs in SEA.

In summary, neutrophils from asthmatic horses exhibited DGE patterns indicative of asthma, including known markers of human asthma. Moreover, these neutrophils displayed an expression profile consistent with increased migratory capacity and the potential for NET formation. The simultaneous expression of genes with a protective action suggests a dual pro- and anti-inflammatory role of neutrophils in SEA.

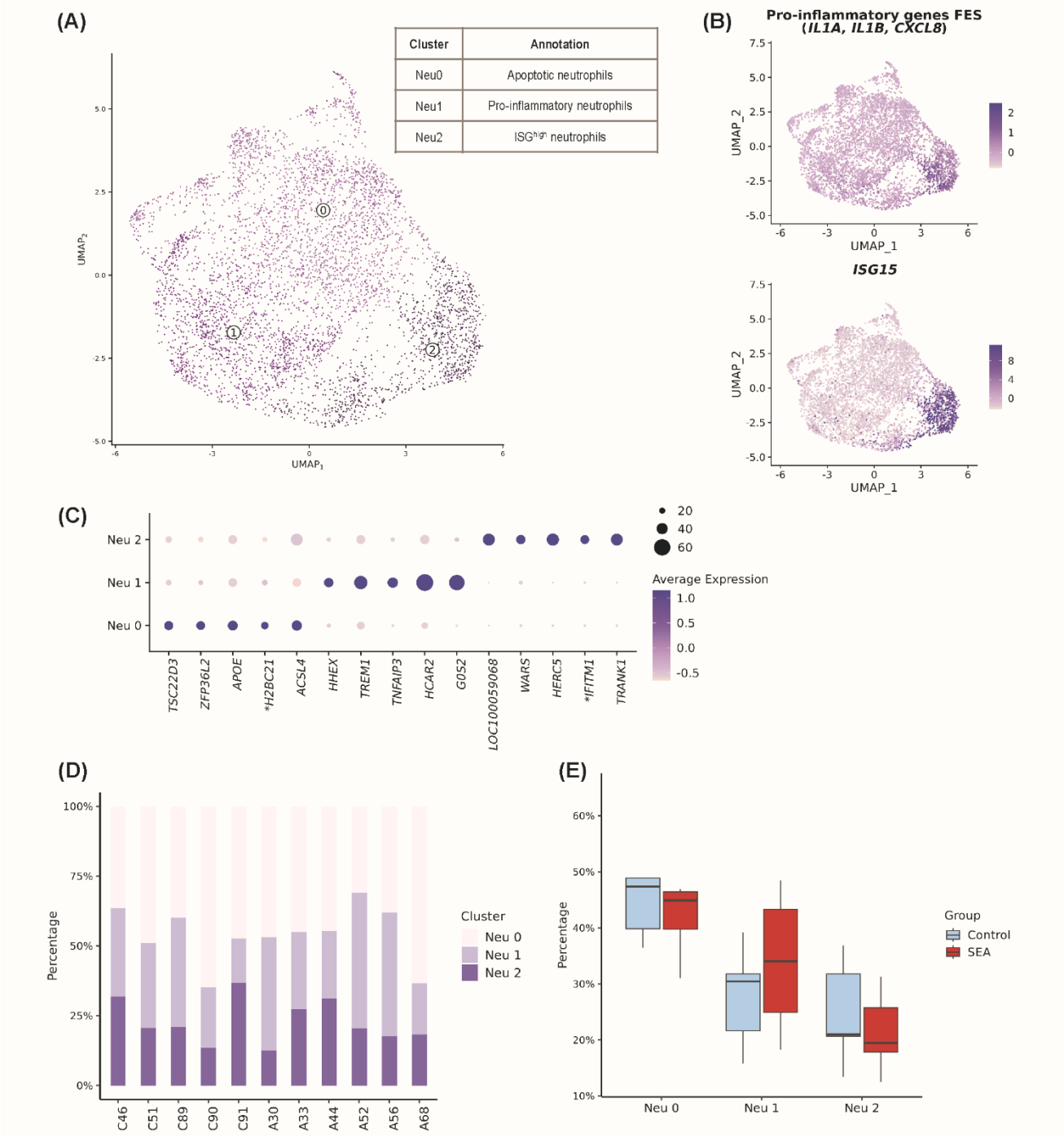


Figure 3: Neutrophils subtypes identified in the BALF of asthmatic and control horses using scRNA-seq. (A) UMAP representation of the three clusters identified. **(B)** Gene expression patterns used for annotation. *FES, feature expression score.* **(C)** Top five differentially expressed genes per cluster (one mitochondrial gene removed). NCBI 103 annotation for *LOC100059068*: *IFIT5-like*. *NCBI 103 annotation for *LOC111774805*: *H2BC21*, *LOC100050797*: *IFITM1*. **(D)** Distribution of clusters among asthmatic and control horses. **(E)** Number of cells from each neutrophil cluster in the asthmatic and control groups. *SEA, severe equine asthma.*

3.4 Gene expression patterns of T cells support a Th17-oriented immune response in SEA

In the T cells of asthmatic horses, two known markers of human asthma, *IL26*[25] and *OLFM4,*[26] were upregulated. As in neutrophils, the acute asthma marker *RETN*[27] was upregulated in cytotoxic T cells (T0). Moreover, T cells of asthmatic horses exhibited a robust Th17 signature, characterized by the upregulation of *IL17A*, *IL17F*, *IL21* and *CCL20*.

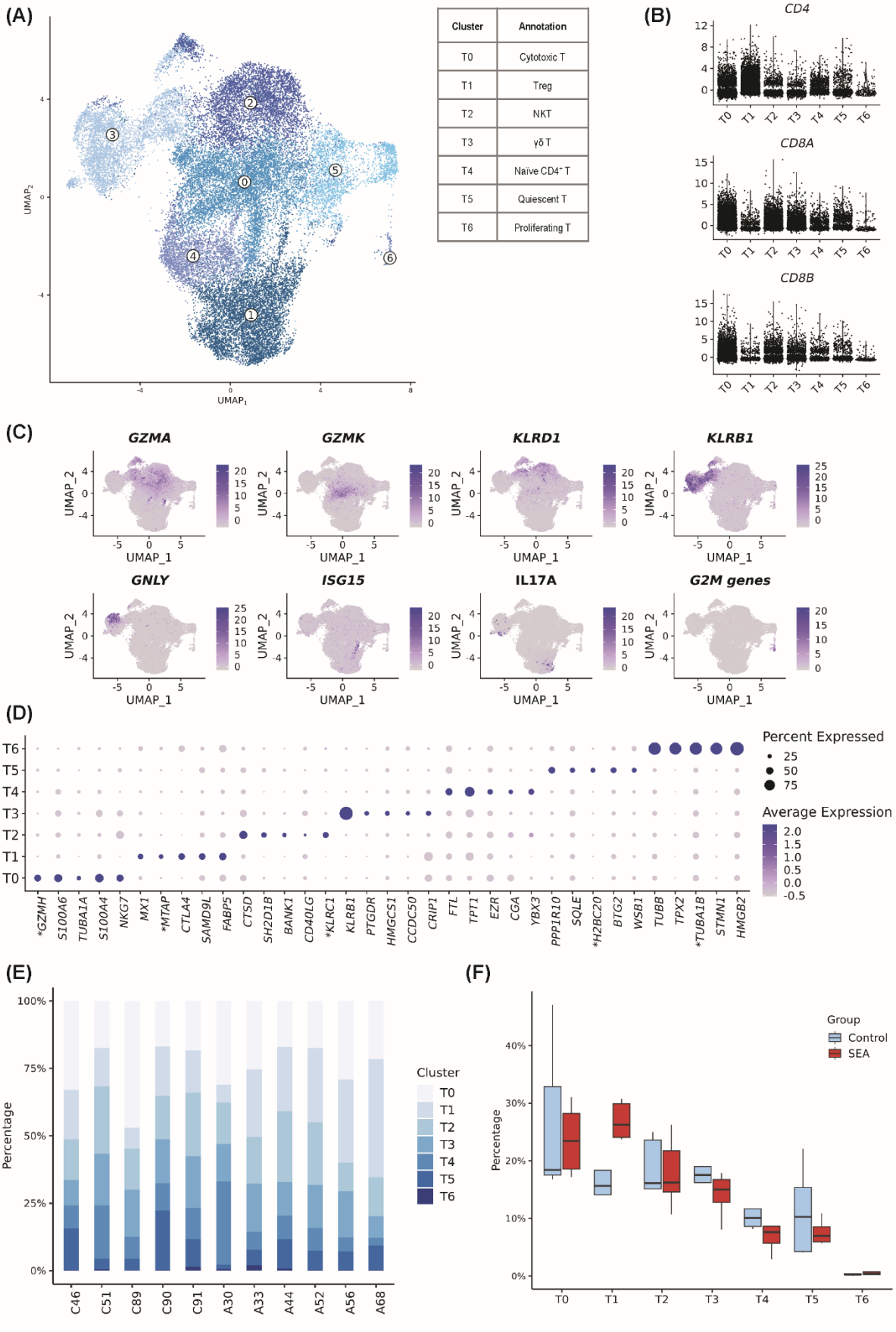
Naïve *CD4*$^+$ T cells (T4) showed a simultaneous upregulation of Th17-associated genes (*IL17A*, *IL1B*, *CCL20* and *NFKBID*) and *FOXP3*. This supported the hypothesis that naïve *CD4*$^+$ T cells adopt a Th17 pathway during differentiation in SEA, as *FOXP3* expression is transiently present during Th17 cell development.[28]

Furthermore, Treg cells (T1) displayed a Th17-oriented profile, with upregulation of *IL21* and *IL17A* and downregulation of *EOMES*, a known suppressor of Th17 differentiation in human Treg cells[29] (Figure S6).

The γδ T (T3) cells conjointly upregulated *IL17A* and *IL1R*, consistent with a γδ17 phenotype.[30] In mice, γδ T cells possess an intrinsic capacity for IL17 production, which is directly induced by IL23 and IL1.[31] Notably, γδ17 T cells are implicated in various human inflammatory diseases,[31] and increased *IL1R* expression has been associated with neutrophilic asthma and reduced pulmonary function in humans.[30]

3.5 Genes associated with T cell function are dysregulated in SEA

Several genes involved in T cell function were differentially expressed in asthmatic horses. Specifically, the marker of T cell exhaustion, *TOX2,*[32] was upregulated, along with *S1PR5*, whose expression is induced by antigen exposure.[33] Cytotoxic T cells (T0) downregulated *GZMB*, a gene associated with lymphocytic inflammation in the lungs.[34] The downregulation of *IL18R1* and *XCL1* supported Treg cell dysfunction. Indeed, downregulation of the IL18 receptor is associated with unresponsiveness of exhausted *CD8*$^+$ T cells.[35] Furthermore, dysfunctional Treg cells in individuals with allergic asthma have been shown to downregulate *XCL1*.[36] Among the T cell subtypes, NKT cells (T2) upregulated *NPY*, a gene associated with reduced NK function.[37]



Cluster	Annotation
T0	Cytotoxic T
T1	Treg
T2	NKT
T3	γδ T
T4	Naïve CD4$^+$ T
T5	Quiescent T
T6	Proliferating T

Figure 4: T cell subtypes identified in the BALF of asthmatic and control horses using scRNA-seq. (A) UMAP representation of the seven clusters identified. *NKT, Natural Killer T cell.* **(B)** Expression levels of the cell surface markers *CD4*, *CD8a* and *CD8b*. **(C)** Gene expression patterns used for annotation **(D)** Top five differentially expressed genes per cluster (snRNA, non-coding genes and ribosomal protein genes removed). *NCBI 103 annotations for *LOC100051986*: *GZMH*, *LOC100065392*: *MTAP*, *LOC100062823*: *KLRC1*, *LOC100053968*: *H2BC20*, *LOC100059091*: *TUBA1B*. **(E)** Distribution of the clusters among asthmatic and control horses. **(F)** Number of cells from each T cell cluster in the asthmatic and control groups. *SEA, severe equine asthma.*

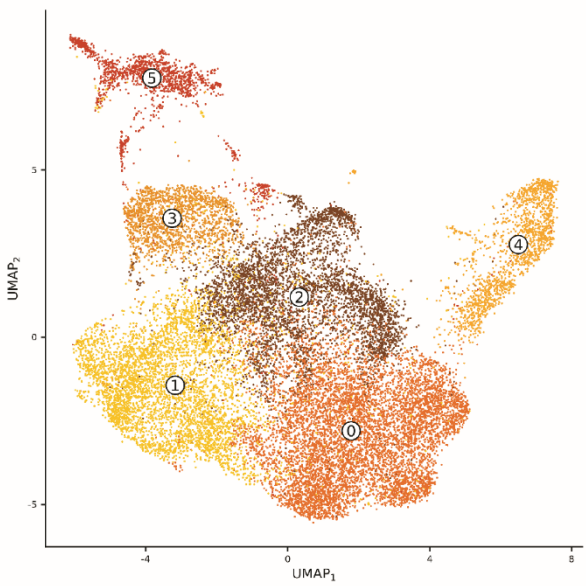
3.6 Monocytes and alveolar macrophages display a Th17 signature in SEA

Among the upregulated genes in Mo/Ma were *OLFM4*, associated with severe lung disease in humans,[38,39] and *CHI3L1,* a marker of neutrophilic asthma.[15] *S100A8*, known for its increased expression in individuals with steroid-resistant neutrophilic asthma,[40] was upregulated, and so was *TLR1*, recently identified as a potential therapeutic target for asthma in humans.[41]

In the *FCN1*^{high} AMs (Mo/Ma0), upregulated genes included *PGLYRP1*, *PTX3* and *CCL20*. In mice, *Pglyrp1* promotes pro-asthmatic Th2 and Th17 responses,[42] while *PTX3* is a marker of non-eosinophilic asthma in humans.[43] In horses, BALF *PTX3* expression increases in acute asthmatic crisis, particularly in dust-activated foamy macrophages.[44] The simultaneous upregulation of the Th17-associated cytokine *CCL20* and downregulation of the Th1-cytokine *CCL11* supported a Th17 polarization of *FCN1*^{high} AMs. Moreover, in the ISG^{high} AMs (Mo/Ma2), genes encoding *CCL20* and its receptor *CCR6* were upregulated, further advocating for a Th17 phenotype.

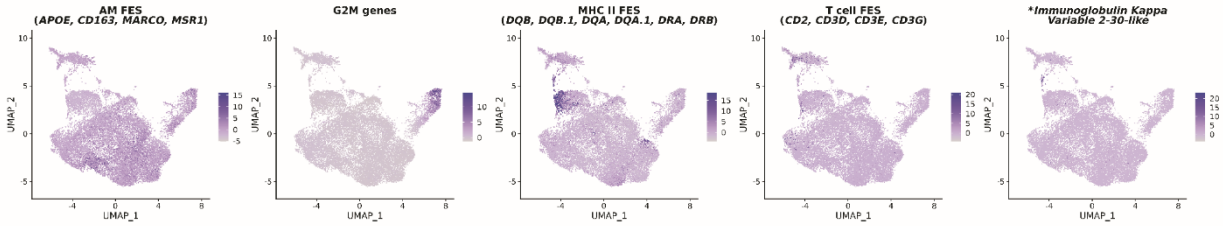
The B cell chemoattractant *CXCL13* was upregulated in intermediate monocytes (Mo/Ma3), and putative monocyte-lymphocyte complexes (Mo/Ma5). Furthermore, intermediate monocytes demonstrated upregulation of *S100A9*, *S100A12*, *CCL17* and *S1PR5*. *S100A9* and *S100A12* serve as biomarkers for neutrophilic asthma.[45,46] *CCL17* is associated with asthma and may contribute to airway remodeling through fibroblast activation via the CCR4-CCL17 axis.[47,48] *S1PR5* regulates monocyte trafficking,[49] suggesting intermediate monocytes from asthmatic horses may possess a higher migratory capacity.

(A)

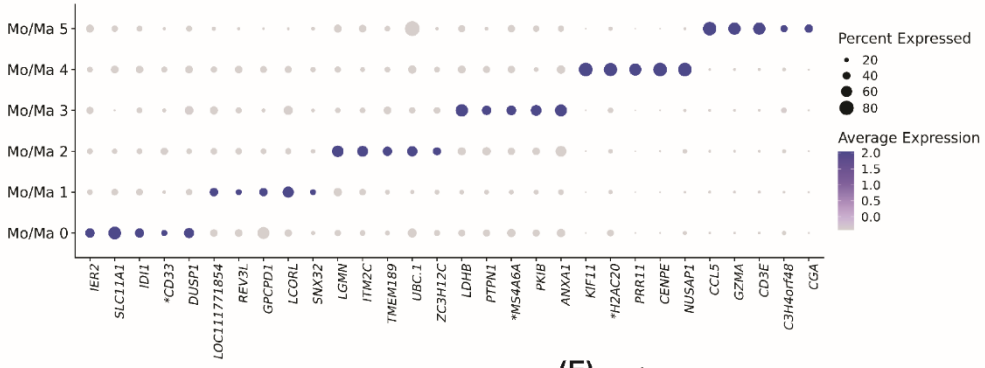


Cluster	Annotation
Mo/Ma0	*FCN1*^{high} alveolar macrophages
Mo/Ma1	Undetermined alveolar macrophages
Mo/Ma2	ISG^{high} alveolar macrophages
Mo/Ma3	Intermediate monocytes
Mo/Ma4	Proliferating alveolar macrophages
Mo/Ma5	Double positive monocyte-lymphocyte cells

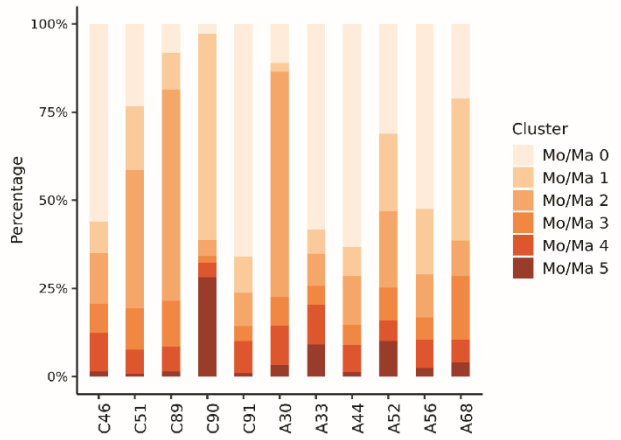
(B)



(C)



(D)



(E)

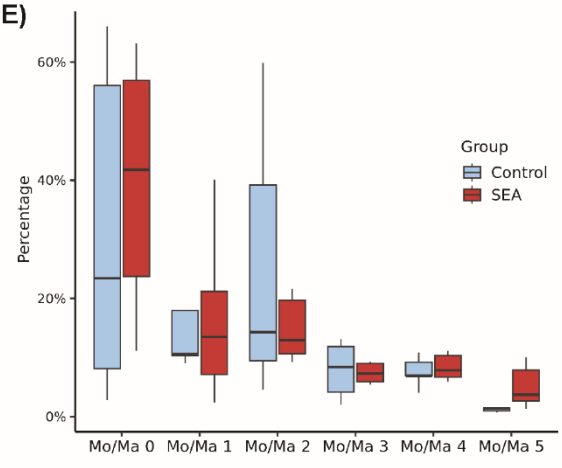


Figure 5: Monocytes-macrophages (Mo/Ma) subtypes identified in the BALF of asthmatic and control horses using scRNA-seq. (A) UMAP representation of the six clusters identified. **(B)** Gene expression patterns used for annotation. *FES, feature expression score.* *NCBI 103 annotations for *LOC100147522*: *GZMH.* **(C)** Top five differentially expressed genes per cluster (non-coding, mitochondrial and ribosomal protein genes removed). NCBI 103 annotation for *LOC111771854*: *oleosin-B6-like*. *NCBI 103 annotations for *LOC100066849*: *CD33*, *LOC100061154*: *MS4A6A, LOC100058587*: *H2AC20*. **(D)** Distribution of the clusters among asthmatic and control horses. **(E)** Number of cells from each Mo/Ma cluster in the asthmatic and control groups. *SEA, severe equine asthma.*

3.7 Th17 activation may result from a crosstalk between monocytes and lymphocytes

The presence of multiple cell types within the Mo/Ma5 cluster was supported by the high number of DEGs identified. This cluster exhibited simultaneous upregulation of *CXCL13* and *IL17A*, both associated with the Th17 pathway. Interestingly, while several T cell clusters in the dataset upregulated *IL17A*, none of the Mo/Ma clusters, except Mo/Ma5, showed this upregulation. Conversely, *CXCL13* upregulation was exclusive to Mo/Ma5 and not observed in any T cell clusters. This led us to conclude that the co-upregulation of *IL17A* and *CXCL13* originated from the dual nature of Mo/Ma5 as monocyte-lymphocyte complexes. Downregulation of the Th1-associated gene *CD27* and granzyme B-like genes further suggested a Th17 polarization within the cells composing the complexes.[50] Additionally, inflammasome-related genes (*SIGLEC14*, *KCNK13* and *PELI2*)[51] were upregulated.

3.8 Gene expression patterns of DCs suggest enhanced migratory capacity and non-Th2 response in SEA

The gene *MARCO* was downregulated in asthmatic DCs. In a murine OVA-asthma model, *Marco*-deficient mice showed increased eosinophilic airway inflammation and airway hyperresponsiveness, accompanied by enhanced migration of lung DCs to draining lymph nodes.[52] Consequently, reduced *MARCO* expression in equine lung DCs may enhance their migration to lymph nodes, leading to an amplified immune response against aeroallergens.

Further analysis of DC subtypes yielded significant results for DC0 (annotated as cDC2s), with the notable upregulation of *GLRX2* and downregulation of *CCL8*. Administration of GLRX2 has been shown to reduce airway inflammation in an OVA-asthma model,[53] indicating its potential protective function. CCL8 is responsible for the recruitment of basophils, eosinophils and mast cells in allergic processes and contributes to airway allergic inflammation by promoting a Th2 immune response.[54] Hence, *CCL8* downregulation in cDC2s argues against a Th2 response in SEA.

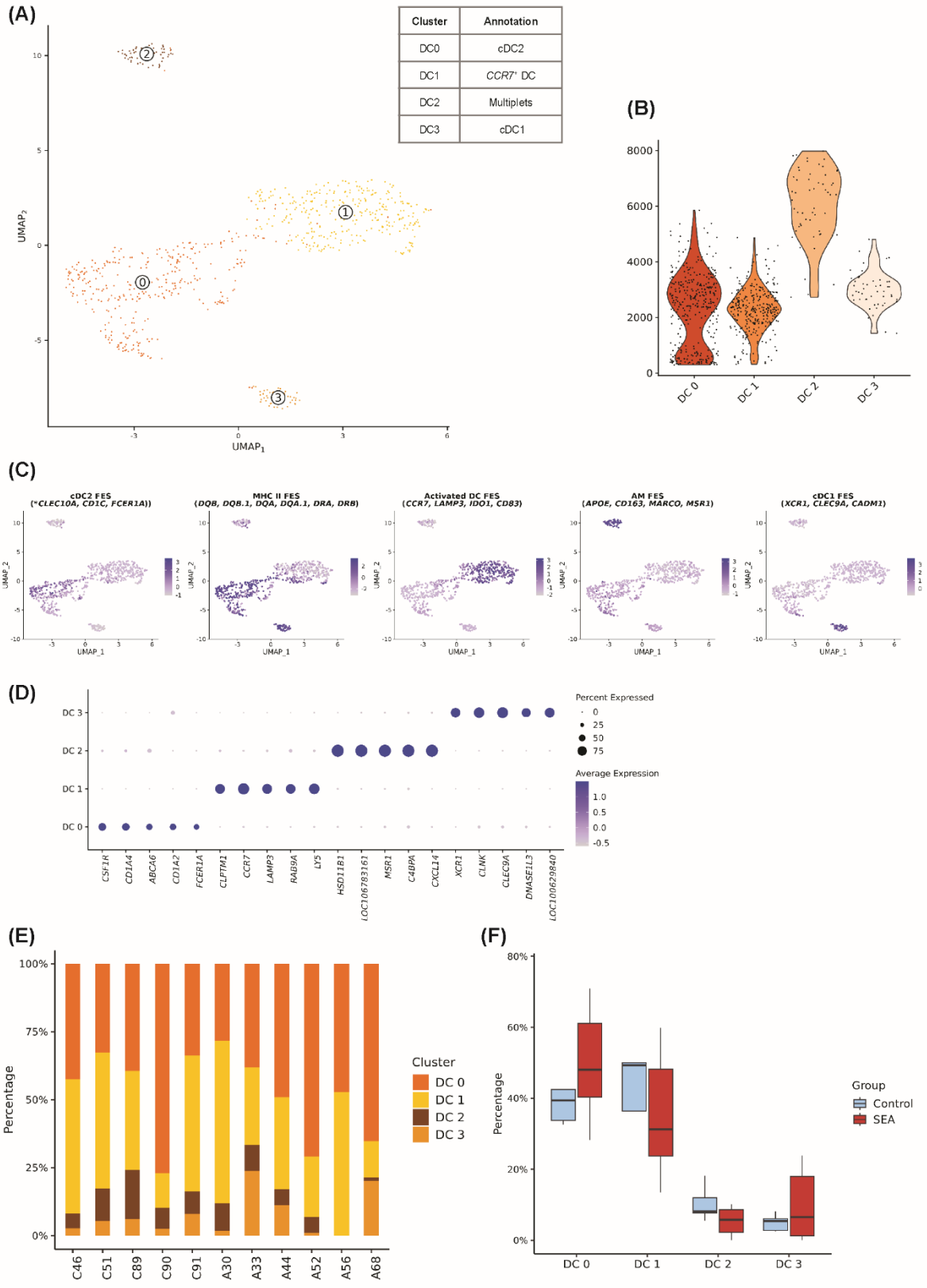


Figure 6: Dendritic cell (DC) subtypes identified in the BALF of asthmatic and control horses using scRNA-seq. (A) UMAP representation of the four clusters identified. *cDC, conventional dendritic cell.* **(B)** RNA feature count for each DC cluster. **(C)** Gene expression patterns used for

annotation. *FES, feature expression score*. *NCBI 103 annotations for *LOC100072936* and *LOC100072933*: *CLEC10*. **(D)** Top five differentially expressed genes per cluster (non-coding genes removed). NCBI 103 annotations for *LOC100629840*: *bone marrow proteoglycan-like*, *LOC106783161*: *apolipoprotein R*. **(E)** Distribution of the clusters among asthmatic and control horses. **(F)** Number of cells from each DC cluster in the asthmatic and control groups. *SEA, severe equine asthma.*

3.9 Gene expression patterns of B cells and mast cells of asthmatic horses points to airway remodeling

The B cells of asthmatic horses upregulated *POU2AF1*, whose elevated expression has been associated with interstitial pulmonary fibrosis[55] and chronic obstructive pulmonary disease[56] in humans. Its expression negatively correlates with lung function.[56] Additionally, *YBX3* was downregulated in mast cells of asthmatic horses. Reduced circulating *YBX3* mRNA is a sensitive predictor of idiopathic pulmonary fibrosis in humans.[57]

4 Discussion

Severe equine asthma is characterized by neutrophilic inflammation in the lower airways, resembling a subset of non-Th2 human asthma. We utilized scRNA-seq to investigate the immune mechanisms underlying SEA. Among the six major cell types identified, B cells and neutrophils were more abundant in asthmatic horses. Notably, the fraction of activated (switched) plasma cells was decreased, indicating a non-Th2 response. Both T cells and Mo/Ma displayed a strong Th17 signature, including upregulation of *CXCL13* by intermediate monocytes. Furthermore, a subset of cells exhibited an expression profile indicative of monocyte-lymphocyte complexes potentially contributing to Th17 activation. Neutrophils showed an increase in NETosis function and reduced capacity to respond to Th17 signals. These findings support a primary Th17-mediated immune response in neutrophilic SEA, probably initiated through monocyte-T cell crosstalk (Figure 7).

Similar Th17-associated responses have been observed in non-Th2 asthma in humans, including organic dust-induced asthma and a subset of non-Th2 asthma patients.[7,58] Although SEA is traditionally considered a Th2-mediated disease, an increasing body of evidence suggests the involvement of Th17 inflammation in the pathological process. Increased levels of *IL17* mRNA have been observed in the BALF of horses with SEA following antigen challenge.[9] Dysregulation of miRNA in the serum of asthmatic horses supports the existence of a mixed Th2/Th17 response.[10] Furthermore, a comprehensive miRNA-mRNA study in equine lung tissues suggests a predominant Th17 pathway, along with some indications of a parallel Th2-type response.[11] Transcriptomics, proteomics, and tissue staining analyses of mediastinal lymph nodes in horses further support a predominant Th17 response in SEA.[12] Recently, T cells isolated from the BALF of horses with SEA demonstrated a Th17 polarization, as evidenced by an elevated frequency of IL17A-secreting lymphocytes following in vitro stimulation.[59]

While studies on asthma have mainly focused on T cells,[58] our study demonstrated the involvement of both T cells and Mo/Ma populations in driving Th17 inflammation in SEA. Importantly, this resulted from alterations in gene expression patterns rather than expansion of these cell populations. The upregulation of key Th17 cytokines such as *IL17A*, *IL21*, and *CCL20* was observed in T cell clusters, suggesting their engagement in a Th17 differentiation pathway. Alveolar macrophages and intermediate monocytes also exhibited a strong Th17 signature, including *CXCL13* upregulation. CXCL13 cytokine levels are elevated in the serum and the BALF of asthmatic humans.[60,61] An anti-CXCL13 antibody reduced inflammation in an asthma mouse model, highlighting

CXCL13 as a promising therapeutic target.[60] This B cell chemoattractant predominantly produced by Mo/Ma and Th17-derived cells[62] is also upregulated in hay dust extract-stimulated PBMCs of asthmatic horses.[63] Because the latter study was performed on a cell mixture, the cellular origin of the increased *CXCL13* expression could not be ascertained. Our single-cell data indicated activated monocytes as the main source of *CXCL13,* most likely induced by IL17 released from T cells.[64] Activated monocytes could in turn induce Th17 differentiation of T cells.[65–67] Collectively, our results support a crosstalk between *IL17A*-producing T cells and *CXCL13*-producing monocytes in the context of a Th17-mediated immune response in SEA (Figure 7).

Of particular interest was the cluster Mo/Ma5 expressing both lymphocytes and monocytes markers, a transcriptomic profile previously observed in other equine BALF scRNA-seq studies.[13,68] The presence of monocyte-T cell interactions has been reported in human blood, with the frequency and phenotype of these cell-cell complexes varying depending on the immune response polarization.[69] Considering that the crosstalk between monocytes and T cells plays a key role in the development of various human inflammatory diseases,[65–67] the potential presence of *bona fide* monocyte-lymphocyte complexes in the lower airway compartment is particularly intriguing. The reciprocal activation of monocytes and lymphocytes may occur through direct cellular contact rather than solely through endocrine or paracrine mechanisms.

In contrast to previous reports (reviewed in [2,3,8]), we did not detect a Th2 or Th1 signature in the cells from asthmatic horses. Notably, we did not observe upregulation of characteristic Th2 and Th1 cytokines such as *IL4*, *IL13* or *IFNγ*, which aligns with the results of an equine BALF flow cytometric study.[59] Consistent with our findings, Th2 and Th17-associated gene expression seems to be regulated in opposite direction in the human airways.[70] In SEA-affected horses, downregulation of *IL4* correlates with increased IL17 staining intensity in the mediastinal lymph nodes.[12] Moreover, Th1- and Th2-associated genes are downregulated in antigen-challenged PBMCs from asthmatic horses.[63] The reduced fraction of activated plasma cells in our study population further argued against a Th2 response. Upon antigen stimulation, non-switched IgM-producing plasma cells become activated and produce immunoglobulins of other classes, a prerequisite for Th2 responses. While switched plasma cells were less frequent in asthmatic horses, the proportion of total B and plasma cells was significantly higher, likely due to *CXCL13* signaling.[64] Consequently, asthmatic horses have a larger pool of B cells, which can potentially differentiate into plasma cells and be activated. This could explain the increased susceptibility of asthmatic horses to certain Th2-associated diseases, such as insect bite hypersensitivity and urticaria.[71] Overall, our findings indicate that SEA is driven primarily by a Th17-mediated immune response characterized by an *IL17*-induced *CXCL13*-mediated recruitment of B cells into the lower airways (Figure 7), potentially predisposing asthmatic horses to secondary Th2-type responses.

The transcriptomic profile of T cells suggested alterations in T cell function, including T cell exhaustion, unresponsiveness of Treg cells, and reduced cytotoxicity in NKT cells. It remains unclear whether these dysregulations are associated with the Th17 polarization of the T cell population, or if they represent independent mechanisms. Nevertheless, these alterations in T cell function may potentiate the abnormal immune response observed in SEA.

Neutrophils are short-lived cells that persist in the lower airways of asthmatic horses due to IL17-induced influx and reduced apoptosis.[72] Neutrophil apoptosis is sometimes accompanied by the formation of NETs, which can trigger tissue damage and sustain chronic inflammation.[72] The observed dysregulation of NETosis-associated genes conforms to the previous observations of excessive

NETosis in the lungs of severely asthmatic horses.[73] Conversely, several DEGs indicated an anti-inflammatory phenotype, particularly in the apoptotic neutrophil subtype, potentially representing a protective mechanism against excessive inflammation. In summary, our findings confirm that BALF neutrophils from SEA-affected horses have a significant pro-inflammatory effect through increased neutrophil persistence and facilitated NET formation in the lungs. The concomitant anti-Th17 transcriptomic profile observed in apoptotic neutrophils suggests a parallel attempt to mitigate lung inflammation. Hence, neutrophils seem to act as effectors rather than primary instigators of asthmatic lung inflammation. Targeting treatment specifically towards the pro-inflammatory neutrophil subtype could disrupt the self-perpetuating inflammatory circle while preserving the antimicrobial functions of other neutrophil subtypes.

Employing scRNA-seq on equine BALF cells elucidated important underlying immune mechanisms of SEA, yet this study had limitations. One significant challenge when studying horses is the inadequate quality of the current reference annotation, necessitating manual annotation of the cell clusters, particularly for poorly defined cell subtypes. Nonetheless, the detection of previously identified cell types and subtypes in equine BALF[13,68] supports the reproducibility of our annotation. Some clusters, such as the "undetermined AMs" cluster, could not be confidently annotated. Further scRNA-seq studies and complementary techniques are required for conclusive insights.

ScRNA-seq is a relatively new technology that comes with computational challenges. One such challenge is the ability to detect and filter technical multiplets without removing biologically significant signals representing cell-cell complexes or new cell types with a dual lineage signature. In this study, we hypothesized that cluster Mo/Ma5 represented *bona fide* monocyte-lymphocyte complexes, supported by the presence of a similar transcriptomic signature in equine BALF cells[13,68] and human PBMCs.[69,74] Although the existence of cellular complexes was confirmed in human PBMCs using imaging flow cytometry,[69,74] validation in horses has yet to be performed. Another potential limitation associated with the 10X Genomics droplet-based technique is its low sensitivity for genes with a low average expression, which could explain the discrepancies with previous bulk RNA or proteomics studies, such as the absence of upregulated Th1 and Th2-associated cytokines.

As this study focused on neutrophilic SEA, results may not apply to other asthma subtypes. This is exemplified by a previous scRNA-seq study on BALF cells from horses with mastocytic asthma,[68] which exhibited a different transcriptomic signature. For example, *FKBP5* was significantly upregulated in mast cells, a gene that we did not detect in our dataset. Moreover, a recent flow cytometric analysis of BALF with functional assays on T cells provided further evidence that distinct mechanisms exist among the various forms of equine asthma.[59] Therefore, studying different endotypes separately is crucial to obtain meaningful results.

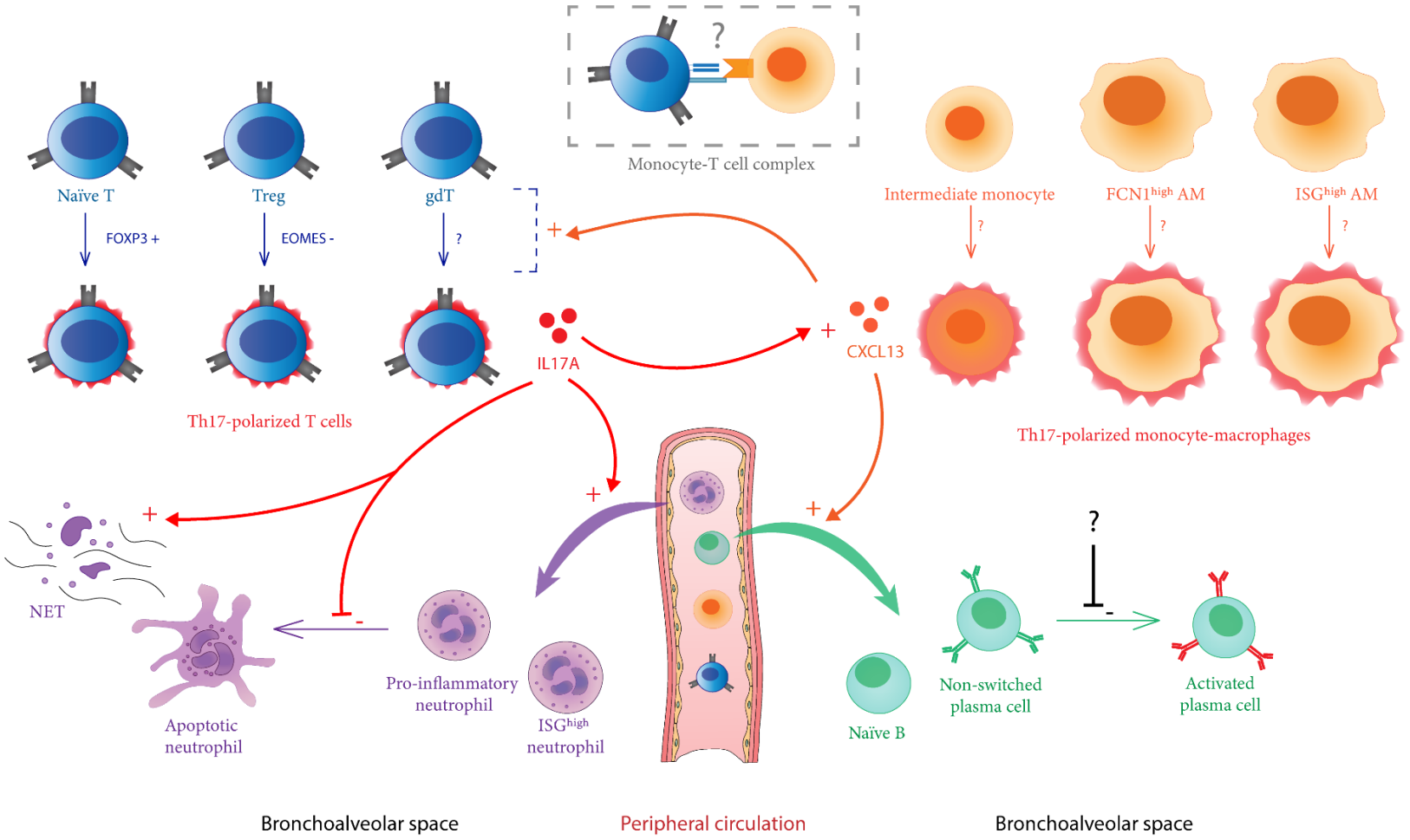


Figure 7: Th17-polarization of pulmonary immune response in severe equine asthma (SEA) - suggested pathways based on scRNA-seq analysis of bronchoalveolar cells. Th17-polarized T cells and monocyte-macrophages fuel an inflammation loop where activated monocytes release CXCL13, promoting Th17 polarization of T cells. IL17A, released by activated T cells, further induces CXCL13 release. Reciprocal activation of T cells and monocytes may also occur via direct cell-cell contact (monocyte-T cell complexes). IL17A and CXCL13 recruit B cells and neutrophils, respectively, from peripheral blood. IL17A influences neutrophils by decreasing apoptosis and enhancing their capacity for NETosis. In SEA, there is a reduced activation of non-switched plasma cells producing IgM antibodies, resulting in a decreased pool of activated plasma cells necessary for a Th2 response. *AM, alveolar macrophages; ISG, interferon interferon-stimulated genes; NET, neutrophil extracellular trap.*

5 Conclusion

The presented scRNA-seq analysis of equine bronchoalveolar cells provides insights into the major immune mechanisms underlying SEA. The use of scRNA-seq allowed us to overcome the influence of varying cell type distribution associated with the disease and to gain unprecedented resolution into the pathophysiology of SEA. This represents a significant breakthrough, challenging the prevailing perception of SEA as a Th2-associated disease. We identified the crucial role of monocytes in initiating the Th17 response in the lungs. The upregulation of *CXCL13* in lung and blood monocytes suggests its potential as a SEA biomarker and therapeutic target. Our findings indicate that monocyte activation may occur through direct cell-cell contact, a hypothesis that should be tested using imaging flow cytometry. This could reshape our understanding of immunotherapy approaches. Therapies targeting Th17-associated cytokines have proven ineffective in reducing symptoms in human asthma.[58] One possible approach could be to prevent monocyte activation by targeting monocyte-T cell synapses. Our results demonstrate several parallels with previous studies on non-Th2 neutrophilic asthma in humans, further validating the horse as a valuable model for studying human asthma.

6 Author Contributions

SS, TL and VG designed the study. Funding was acquired by VG. SS collected and processed the samples. SS and VJ performed the computational analysis. SS wrote the original draft. All authors reviewed, edited and approved the final version of the manuscript.

7 Acknowledgments

The authors would like to thank Dr. med. vet. Michelle Wyler and Dr. med. vet. Nicole Altermatt for their assistance in sample collection. We thank Dr. Pamela Nicholson and the laboratory technicians from the Next Generation Sequencing Platform of the University of Bern for performing the sequencing experiments. We are grateful for the Interfaculty Bioinformatics Unit of the University of Bern for providing the high-performance computing infrastructure. The authors furthermore acknowledge Dr. med. vet. Laureen Peters for conducting cytological analysis and the laboratory technicians of the Central Diagnostic Laboratory of the Vetsuisse Faculty for their assistance in sample processing.

8 Tables

Table 1: Study population characteristics

Variable	Control (N=5)[1]	Asthmatic (N=6)[1]	*P*-value[2]
Sex			0.6
Mare	3 (60%)	2 (33%)	
Gelding	2 (40%)	4 (67%)	
Age (years)	12 (11, 12)	12 (7, 19)	> 0.9
HOARSI* score	1 (1, 1)	4 (3, 4)	**< 0.01**
Weight (kg)	594 (582, 613)	578 (548, 609)	> 0.9
Clinical score (/23)	1 (0, 1)	6 (5, 7)	**< 0.01**
Tracheal mucous score (/5)	1 (1, 2)	3 (3, 4)	**< 0.05**
BALF yield (%)	48 (48, 48)	52 (45, 56)	0.7
BALF macrophages (%)	51 (48, 56)	51 (38, 55)	0.5
BALF lymphocytes (%)	40 (36, 47)	36 (22, 40)	0.4
BALF neutrophils (%)	4 (2, 7)	18 (12, 25)	**< 0.01**
BALF mast cells (%)	1 (0, 1)	1 (1, 2)	0.5
BALF eosinophils (%)	0 (0, 0)	0 (0, 0)	> 0.9

[1]Median (1st quartile, 3rd quartile)
[2]Fisher's exact test; Wilcoxon rank sum test
*HOARSI: Horse Owner Assessed Respiratory Signs

Table 2: Proportions of the major cell types determined with scRNA-seq and compared between asthmatic (SEA) and control (CTL) groups

Cell type	Number* (N=11)	Mean % (N=11)	Mean % (CTL, N=5)	Mean % (SEA, N=6)	Ratio SEA/CTL	*P*-value	FDR
B cells	756	1.3	0.3	1.9	5.9	**< 0.001**	**< 0.001**
Neutrophils	5,145	8.5	2.4	13	5.3	**< 0.001**	**< 0.001**
Mast cells	1,232	2.0	1.7	2.3	1.4	0.13	0.25
Mo/Ma	22,370	37.1	44.6	32.0	0.7	0.18	0.27
Dendritic cells	754	1.3	1.2	1.4	1.2	0.80	0.93
T cells	30,005	50.0	50.0	49.4	1.0	0.93	0.93

* Post quality filtering

Table 3: Proportions of the B cell subtypes identified with scRNA-seq and compared between asthmatic (SEA) and control (CTL) groups

B cell cluster	Mean % (N=11)	Mean % (CTL, N=5)	Mean % (SEA, N=6)	Ratio SEA/CTL	*P*-value	FDR
B0	72.2	68.6	71.9	1.0	0.62	0.62
B1	20.5	11.5	21.3	1.8	0.18	0.26
B2	7.3	19.9	6.8	0.3	**0.02**	**< 0.05**

9 References

1. Couëtil LLL, Cardwell JMM, Gerber V, Lavoie J -P. P, Léguillette R, Richard EAA. Inflammatory Airway Disease of Horses-Revised Consensus Statement. *J Vet Intern Med*. 2016;30(2):503-515. doi:10.1111/jvim.13824

2. Bond S, Léguillette R, Richard EA, et al. Equine asthma: Integrative biologic relevance of a recently proposed nomenclature. *J Vet Intern Med*. 2018;32(6):2088-2098. doi:10.1111/jvim.15302

3. Bullone M, Lavoie JP. Asthma "of horses and men"-How can equine heaves help us better understand human asthma immunopathology and its functional consequences? *Mol Immunol*. 2015;66(1):97-105. doi:10.1016/j.molimm.2014.12.005

4. Lange-Consiglio A, Stucchi L, Zucca E, Lavoie JP, Cremonesi F, Ferrucci F. Insights into animal models for cell-based therapies in translational studies of lung diseases: Is the horse with naturally occurring asthma the right choice? *Cytotherapy*. 2019;21(5):525-534. doi:10.1016/j.jcyt.2019.02.010

5. Kuruvilla ME, Lee FE-H, Lee GB. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. *Clin Rev Allergy Immunol*. 2019;56(2):219-233. doi:10.1007/s12016-018-8712-1

6. Tyler SR, Bunyavanich S. Leveraging -omics for asthma endotyping. *J Allergy Clin Immunol*. 2019;144(1):13-23. doi:10.1016/j.jaci.2019.05.015

7. Sheats MK, Davis KU, Poole JA. Comparative Review of Asthma in Farmers and Horses. *Curr Allergy Asthma Rep*. 2019;19(11):50. doi:10.1007/s11882-019-0882-2

8. Simões J, Batista M, Tilley P. The Immune Mechanisms of Severe Equine Asthma—Current Understanding and What Is Missing. *Animals*. 2022;12(6):744. doi:10.3390/ani12060744

9. Debrue M, Hamilton E, Joubert P, Lajoie-Kadoch S, Lavoie J-P. Chronic exacerbation of equine heaves is associated with an increased expression of interleukin-17 mRNA in bronchoalveolar lavage cells. *Vet Immunol Immunopathol*. 2005;105(1-2):25-31. doi:10.1016/j.vetimm.2004.12.013

10. Pacholewska A, Kraft MF, Gerber V, Jagannathan V. Differential expression of serum MicroRNAs supports CD4 + t cell differentiation into Th2/Th17 cells in severe equine asthma. *Genes (Basel)*. 2017;8(12). doi:10.3390/genes8120383

11. Hulliger MF, Pacholewska A, Vargas A, et al. An Integrative miRNA-mRNA Expression Analysis Reveals Striking Transcriptomic Similarities between Severe Equine Asthma and Specific Asthma Endotypes in Humans. *Genes (Basel)*. 2020;11(10):1143. doi:10.3390/genes11101143

12. Korn A, Miller D, Dong L, Buckles EL, Wagner B, Ainsworth DM. Differential Gene Expression Profiles and Selected Cytokine Protein Analysis of Mediastinal Lymph Nodes of Horses with Chronic Recurrent Airway Obstruction (RAO) Support an Interleukin-17 Immune Response. Allen IC, ed. *PLoS One*. 2015;10(11):e0142622. doi:10.1371/journal.pone.0142622

13. Sage SE, Nicholson P, Peters LM, Leeb T, Jagannathan V, Gerber V. Single-cell gene expression analysis of cryopreserved equine bronchoalveolar cells. *Front Immunol*. 2022;13(August):1-18. doi:10.3389/fimmu.2022.929922

14. Liu L, Zhang X, Liu Y, et al. Chitinase-like protein YKL-40 correlates with inflammatory phenotypes, anti-asthma responsiveness and future exacerbations. *Respir Res*. 2019;20(1):1-12. doi:10.1186/s12931-019-1051-9

15. Cremades-Jimeno L, de Pedro MÁ, López-Ramos M, et al. Prioritizing Molecular Biomarkers in Asthma and Respiratory Allergy Using Systems Biology. *Front Immunol*. 2021;12:1004. doi:10.3389/fimmu.2021.640791

16. Alevy YG, Patel AC, Romero AG, et al. IL-13–induced airway mucus production is attenuated by MAPK13 inhibition. *J Clin Invest*. 2012;122(12):4555-4568. doi:10.1172/JCI64896

17. Bao C, Liu C, Liu Q, et al. Liproxstatin-1 alleviates LPS/IL-13-induced bronchial epithelial cell injury and neutrophilic asthma in mice by inhibiting ferroptosis. *Int Immunopharmacol*. 2022;109:108770. doi:10.1016/j.intimp.2022.108770

18. Sprenkeler EGG, Zandstra J, van Kleef ND, et al. S100A8/A9 Is a Marker for the Release of Neutrophil Extracellular Traps and Induces Neutrophil Activation. *Cells*. 2022;11(2):236. doi:10.3390/cells11020236

19. Jiang S, Park DW, Tadie J-M, et al. Human Resistin Promotes Neutrophil Proinflammatory Activation and Neutrophil Extracellular Trap Formation and Increases Severity of Acute Lung Injury. *J Immunol*. 2014;192(10):4795-4803. doi:10.4049/jimmunol.1302764

20. Zhu L, Zeng D, Lei X, et al. KLF2 regulates neutrophil migration by modulating CXCR1 and CXCR2 in asthma. *Biochim Biophys Acta - Mol Basis Dis*. 2020;1866(12):165920. doi:10.1016/j.bbadis.2020.165920

21. Le A, Wu Y, Liu W, et al. MiR-144-induced KLF2 inhibition and NF-kappaB/CXCR1 activation promote neutrophil extracellular trap–induced transfusion-related acute lung injury. *J Cell Mol Med*. 2021;25(14):6511-6523. doi:10.1111/jcmm.16650

22. Leemans JC, te Velde AA, Florquin S, et al. The Epidermal Growth Factor-Seven Transmembrane (EGF-TM7) Receptor CD97 Is Required for Neutrophil Migration and Host Defense. *J Immunol*. 2004;172(2):1125-1131. doi:10.4049/jimmunol.172.2.1125

23. Majchrzak-Gorecka M, Majewski P, Grygier B, Murzyn K, Cichy J. Secretory leukocyte protease inhibitor (SLPI), a multifunctional protein in the host defense response. *Cytokine Growth Factor Rev*. 2016;28:79-93. doi:10.1016/j.cytogfr.2015.12.001

24. Prince LR, Prosseda SD, Higgins K, et al. NR4A orphan nuclear receptor family members, NR4A2 and NR4A3, regulate neutrophil number and survival. *Blood*. 2017;130(8):1014-1025. doi:10.1182/blood-2017-03-770164

25. Cardenas EI, Che KF, Konradsen JR, Bao A, Lindén A. IL-26 in asthma and COPD. *Expert Rev Respir Med*. 2022;16(3):293-301. doi:10.1080/17476348.2022.2045197

26. Chen X, Khalid K, Chen D, Qiu C. Serum levels of olfactomedin 4: a biomarker for asthma control state in asthmatics. *Ann Transl Med*. 2020;8(7):494-494. doi:10.21037/atm.2020.03.213

27. Al Mutairi SS, Mojiminiyi OA, Shihab-Eldeen A, Al Rammah T, Abdella N. Putative roles of circulating resistin in patients with asthma, COPD and cigarette smokers. *Dis Markers*. 2011;31(1):1-7. doi:10.3233/DMA-2011-0793

28. Mailer RKW, Joly A-L, Liu S, Elias S, Tegner J, Andersson J. IL-1β promotes Th17 differentiation by inducing alternative splicing of FOXP3. *Sci Rep*. 2015;5(1):14674. doi:10.1038/srep14674

29. Gruarin P, Maglie S, Simone M, et al. Eomesodermin controls a unique differentiation program in human IL-10 and IFN-γ coproducing regulatory T cells. *Eur J Immunol*. 2019;49(1):96-111. doi:10.1002/eji.201847722

30. Evans MD, Esnault S, Denlinger LC, Jarjour NN. Sputum cell IL-1 receptor expression level is a marker of airway neutrophilia and airflow obstruction in asthmatic patients. *J Allergy Clin Immunol*. 2018;142(2):415-423. doi:10.1016/j.jaci.2017.09.035

31. Akitsu A, Iwakura Y. Interleukin-17-producing γδ T (γδ17) cells in inflammatory diseases. *Immunology*. 2018;155(4):418-426. doi:10.1111/imm.12993

32. Seo H, Chen J, González-Avalos E, et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8 + T cell exhaustion. *Proc Natl Acad Sci*. 2019;116(25):12410-12415. doi:10.1073/pnas.1905675116

33. Evrard M, Wynne-Jones E, Peng C, et al. Sphingosine 1-phosphate receptor 5 (S1PR5) regulates the peripheral retention of tissue-resident lymphocytes. *J Exp Med*. 2022;219(1). doi:10.1084/jem.20210116

34. Hirota JA, Hiebert PR, Gold M, et al. Granzyme B Deficiency Exacerbates Lung Inflammation in Mice after Acute Lung Injury. *Am J Respir Cell Mol Biol*. 2013;49(3):453-462. doi:10.1165/rcmb.2012-0512OC

35. Ingram JT, Yi JS, Zajac AJ. Exhausted CD8 T Cells Downregulate the IL-18 Receptor and Become Unresponsive to Inflammatory Cytokines and Bacterial Co-infections. Wherry EJ, ed. *PLoS Pathog*. 2011;7(9):e1002273. doi:10.1371/journal.ppat.1002273

36. Nguyen KD, Fohner A, Booker JD, Dong C, Krensky AM, Nadeau KC. XCL1 Enhances Regulatory Activities of CD4+CD25highCD127low/− T Cells in Human Allergic Asthma. *J Immunol*. 2008;181(8):5386-5395. doi:10.4049/jimmunol.181.8.5386

37. Nair MPN, Schwartz SA, Wu K, Kronfol Z. Effect of neuropeptide Y on natural killer activity of normal human lymphocytes. *Brain Behav Immun*. 1993;7(1):70-78. doi:10.1006/brbi.1993.1007

38. Brand HK, Ahout IML, de Ridder D, et al. Olfactomedin 4 Serves as a Marker for Disease Severity in Pediatric Respiratory Syncytial Virus (RSV) Infection. Schildgen O, ed. *PLoS One*. 2015;10(7):e0131927. doi:10.1371/journal.pone.0131927

39. Gong F, Li R, Zheng X, et al. OLFM4 Regulates Lung Epithelial Cell Function in Sepsis-Associated ARDS/ALI via LDHA-Mediated NF-κB Signaling. *J Inflamm Res*. 2021;Volume 14(December):7035-7051. doi:10.2147/JIR.S335915

40. Bozinovski S, Cross M, Vlahos R, et al. S100A8 Chemotactic Protein Is Abundantly Increased, but Only a Minor Contributor to LPS-Induced, Steroid Resistant Neutrophilic Lung Inflammation in Vivo. *J Proteome Res*. 2005;4(1):136-145. doi:10.1021/pr049829t

41. Wjst M. Exome variants associated with asthma and allergy. *Sci Rep*. 2022;12(1):21028. doi:10.1038/s41598-022-24960-6

42. Park SY, Jing X, Gupta D, Dziarski R. Peptidoglycan Recognition Protein 1 Enhances Experimental Asthma by Promoting Th2 and Th17 and Limiting Regulatory T Cell and Plasmacytoid Dendritic Cell Responses. *J Immunol*. 2013;190(7):3480-3492. doi:10.4049/jimmunol.1202675

43. Gao P, Tang K, Wang M, et al. Pentraxin levels in non-eosinophilic versus eosinophilic asthma. *Clin Exp Allergy*. 2018;48(8):981-989. doi:10.1111/cea.13168

44. Ramery E, Fievez L, Fraipont A, Bureau F, Lekeux P. Characterization of pentraxin 3 in the horse and its expression in airways. *Vet Res*. 2010;41(2):18. doi:10.1051/vetres/2009066

45. Kotsiou OS, Papagiannis D, Papadopoulou R, Gourgoulianis KI. Calprotectin in Lung Diseases. *Int J Mol Sci*. 2021;22(4):1706. doi:10.3390/ijms22041706

46. Lin Q, Ni H, Zhong J, Zheng Z, Nie H. Identification of hub genes and potential biomarkers of neutrophilic asthma: evidence from a bioinformatics analysis. *J Asthma*. 2023;60(2):348-359. doi:10.1080/02770903.2022.2051544

47. Ai X, Shen H, Wang Y, et al. Developing a Diagnostic Model to Predict the Risk of Asthma Based on Ten Macrophage-Related Gene Signatures. Qin S, ed. *Biomed Res Int*. 2022;2022:1-14. doi:10.1155/2022/3439010

48. Wang Q, Liu S, Min J, et al. CCL17 drives fibroblast activation in the progression of pulmonary fibrosis by enhancing the TGF-β/Smad signaling. *Biochem Pharmacol*. 2023;210:115475. doi:10.1016/j.bcp.2023.115475

49. Debien E, Mayol K, Biajoux V, et al. S1PR5 is pivotal for the homeostasis of patrolling monocytes. *Eur J Immunol*. 2013;43(6):1667-1675. doi:10.1002/eji.201343312

50. Hoek KL, Greer MJ, McClanahan KG, et al. Granzyme B prevents aberrant IL-17 production and intestinal pathogenicity in CD4+ T cells. *Mucosal Immunol*. 2021;14(5):1088-1099. doi:10.1038/s41385-021-00427-1

51. Deng J, Yu X-Q, Wang P-H. Inflammasome activation and Th17 responses. *Mol Immunol*. 2019;107:142-164. doi:10.1016/j.molimm.2018.12.024

52. Arredouani MS, Franco F, Imrich A, et al. Scavenger Receptors SR-AI/II and MARCO Limit Pulmonary Dendritic Cell Migration and Allergic Airway Inflammation. *J Immunol*. 2007;178(9):5912-5920. doi:10.4049/jimmunol.178.9.5912

53. Hanschmann E-M, Berndt C, Hecker C, et al. Glutaredoxin 2 Reduces Asthma-Like Acute Airway Inflammation in Mice. *Front Immunol*. 2020;11(November):1-11. doi:10.3389/fimmu.2020.561724

54. Branchett WJ, Cook J, Oliver RA, et al. Airway macrophage-intrinsic TGF-β1 regulates pulmonary immunity during early-life allergen exposure. *J Allergy Clin Immunol*. 2021;147(5):1892-1906. doi:10.1016/j.jaci.2021.01.026

55. McDonough JE, Ahangari F, Li Q, et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight*. 2019;4(22). doi:10.1172/jci.insight.131597

56. Zhang D-W, Ye J-J, Sun Y, et al. CD19 and POU2AF1 are Potential Immune-Related Biomarkers Involved in the Emphysema of COPD: On Multiple Microarray Analysis. *J Inflamm Res*. 2022;Volume 15(April):2491-2507. doi:10.2147/JIR.S355764

57. Di Mauro S, Scamporrino A, Fruciano M, et al. Circulating Coding and Long Non-Coding RNAs as Potential Biomarkers of Idiopathic Pulmonary Fibrosis. *Int J Mol Sci*. 2020;21(22):8812. doi:10.3390/ijms21228812

58. Xie Y, Abel PW, Casale TB, Tu Y. TH17 cells and corticosteroid insensitivity in severe asthma. *J Allergy Clin Immunol*. 2022;149(2):467-479. doi:10.1016/J.JACI.2021.12.769

59. Gressler AE, Lübke S, Wagner B, Arnold C, Lohmann KL, Schnabel CL. Comprehensive Flow Cytometric Characterization of Bronchoalveolar Lavage Cells Indicates Comparable Phenotypes Between Asthmatic and Healthy Horses But Functional Lymphocyte Differences. *Front Immunol*. 2022;13(July):1-17. doi:10.3389/fimmu.2022.896255

60. Baay-Guzman GJ, Huerta-Yepez S, Vega MI, et al. Role of CXCL13 in asthma: Novel therapeutic target. *Chest*. 2012;141(4):886-894. doi:10.1378/chest.11-0633

61. Alturaiki W. Elevated Plasma Levels of CXCL13 Chemokine in Saudi Patients With Asthma Exacerbation. *Cureus*. 2022;14(1):1-7. doi:10.7759/cureus.21142

62. Takagi R, Higashi T, Hashimoto K, et al. B Cell Chemoattractant CXCL13 Is Preferentially Expressed by Human Th17 Cell Clones. *J Immunol*. 2008;181(1):186-189. doi:10.4049/jimmunol.181.1.186

63. Pacholewska A, Jagannathan V, Drögemüller M, et al. Impaired Cell Cycle Regulation in a Natural Equine Model of Asthma. Wade C, ed. *PLoS One*. 2015;10(8):e0136103. doi:10.1371/journal.pone.0136103

64. Al-Kufaidy R, Vazquez-Tello A, BaHammam AS, Al-Muhsen S, Hamid Q, Halwani R. IL-17 enhances the migration of B cells during asthma by inducing CXCL13 chemokine production in structural lung cells. *J Allergy Clin Immunol*. 2017;139(2):696-699.e5. doi:10.1016/j.jaci.2016.07.037

65. Roberts CA, Dickinson AK, Taams LS. The Interplay Between Monocytes/Macrophages and CD4+ T Cell Subsets in Rheumatoid Arthritis. *Front Immunol*. 2015;6(NOV):571. doi:10.3389/fimmu.2015.00571

66. Pang Y, Du X, Xu X, Wang M, Li Z. Monocyte activation and inflammation can exacerbate Treg/Th17 imbalance in infants with neonatal necrotizing enterocolitis. *Int Immunopharmacol*. 2018;59:354-360. doi:10.1016/j.intimp.2018.04.026

67. Evans HG, Gullick NJ, Kelly S, et al. In vivo activated monocytes from the site of inflammation in humans specifically promote Th17 responses. *Proc Natl Acad Sci*. 2009;106(15):6232-6237. doi:10.1073/pnas.0808144106

68. Riihimäki M, Fegraeus K, Nordlund J, et al. Single cell transcriptomics delineates the immune-cell landscape in equine lower airways and reveals upregulation of the FKBP5 gene in horses with asthma, , 07 April 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs. *Res Sq*. 2023;(Version 1):1-29. doi:10.21203/rs.3.rs-2768703/v1

69. Burel JG, Pomaznoy M, Lindestam Arlehamn CS, et al. Circulating T cell-monocyte complexes are markers of immune perturbations. *Elife*. 2019;8:1-21. doi:10.7554/eLife.46045

70. Choy DF, Hart KM, Borthwick LA, et al. TH2 and TH17 inflammatory pathways are reciprocally regulated in asthma. *Sci Transl Med*. 2015;7(301). doi:10.1126/scitranslmed.aab3142

71. Kehrli D, Jandova V, Fey K, Jahn P, Gerber V. Multiple Hypersensitivities Including Recurrent Airway Obstruction, Insect Bite Hypersensitivity, and Urticaria in 2 Warmblood Horse Populations. *J Vet Intern Med*. 2015;29(1):320-326. doi:10.1111/jvim.12473

72. Davis KU, Sheats MK. The Role of Neutrophils in the Pathophysiology of Asthma in Humans and Horses. *Inflammation*. 2021;44(2):450-465. doi:10.1007/s10753-020-01362-2

73. Janssen P, Tosi I, Hego A, Maréchal P, Marichal T, Radermecker C. Neutrophil Extracellular Traps Are Found in Bronchoalveolar Lavage Fluids of Horses With Severe Asthma and Correlate With Asthma Severity. *Front Immunol*. 2022;13(July):1-18. doi:10.3389/fimmu.2022.921077

74. Burel JG, Pomaznoy M, Lindestam Arlehamn CS, et al. The Challenge of Distinguishing Cell–Cell Complexes from Singlet Cells in Non-Imaging Flow Cytometry and Single-Cell Sorting. *Cytom Part A*. 2020;97(11):1127-1135. doi:10.1002/cyto.a.24027

10 Supplementary Material

Supplementary figures: S1-S6

Supplementary tables

Supplementary table 1. Summary metrics of the detected cells for the 12 samples sequenced
Supplementary table 2. Marker genes for the 19 clusters (complete dataset)
Supplementary table 3. Marker genes for the six major cell types (complete dataset)
Supplementary table 4. Marker genes for the neutrophil clusters (independent reanalysis)
Supplementary table 5. Marker genes for the B cell clusters (independent reanalysis)
Supplementary table 6. Marker genes for the T cell clusters (independent reanalysis)

Supplementary table 7. Marker genes for the monocyte-macrophage clusters (independent reanalysis)
Supplementary table 8. Marker genes for the dendritic cell clusters (independent reanalysis)
Supplementary table 9. Differential gene expression analysis – Major cell types
Supplementary table 10. Differential gene expression analysis – Neutrophils
Supplementary table 11. Differential gene expression analysis – T cells
Supplementary table 12. Differential gene expression analysis – Monocytes-Macrophages
Supplementary table 13. Differential gene expression analysis – Dendritic cells
Supplementary table 14. Characteristics of the cell suspensions subject to scRNA-seq

Supplementary materials and methods

Supplementary materials and methods 1 - Detailed materials and methods
Supplementary materials and methods 2 - Experimental protocol for equine bronchoalveolar cells cryopreservation

Discussion

In this thesis, we successfully demonstrated the applicability of scRNA-seq on cryopreserved equine bronchoalveolar cells. As an attempt to improve the mapping to transcriptome of our scRNA-seq reads and improve the overall quality of data, we constructed a custom reference annotation using long-read sequencing of BALF cells. Although we detected novel isoforms, mapping of our scRNA-seq transcripts did not improve significantly. Using an alternative reference annotation, we obtained a satisfactory read mapping, resulting in high quality scRNA-seq data. As anticipated, we were able to detect the major leukocyte populations present in BALF using scRNA-seq. While we could not identify the classic Th1 and Th2 subpopulations, we described novel monocyte-macrophages and T cell subtypes. We applied the knowledge gained in this first pilot scRNA-seq study to study the pathophysiology of SEA. Analysis of the single-cell transcriptomes identified a Th17 signature in SEA, several other interesting gene expression patterns as well as changes in B cell composition.

1. Optimizing experimental workflow and data analysis in scRNA-seq

The first scRNA-seq experiment was rich in lessons. It allowed us to identify potential challenges in the experimental workflow and data analysis, leading to optimizations in the wet lab protocol and pipeline analysis for future studies. For example, the target cell recovery in Study 1 was set at 5,000 cells per sample, but the scRNA-seq analysis yielded 2,000 or fewer cells per sample, indicating a cell capture efficiency of approximately 40%. Armed with this knowledge, we successfully adjusted the loading concentration of our cell suspension in Study 3, resulting in the retrieval of more than 5,000 cells per sample.

Another issue arose during the quality control analysis of our dataset, where we observed a low mapping rate to the transcriptome (mean of 33.5%). Several factors could contribute to this, such as read multimapping, incomplete 3’-UTR annotation, or inadequate transcriptome annotation. Read multimapping occurs when a transcript cannot be definitively mapped to a unique genomic region and is subsequently filtered out. This is often a consequence of poor annotation of the 3'-end of the transcript in the reference annotation. Similarly, if the 3’-UTRs including their poly(A) sites are not well annotated in the reference annotation, transcripts may fail to align and are subsequently filtered out (87). Additionally, the absence of BALF-specific isoforms in the reference equine transcriptome could have led to the exclusion of such transcripts from our scRNA-seq dataset.

Consequently, we decided to construct a custom BALF-specific transcriptome. We leveraged the long-read sequencing technique, resulting in the identification of 14,234 full-length transcript isoforms from 7,017 unique genes, including 3,428 novel isoforms. These included transcripts absent from existing genome annotations, transcripts mapping to putative novel genes and fusion transcripts combining exons from multiple genes. However, when we used this custom BALF-specific transcriptome as a reference for scRNA-seq analysis, the improvement in read mapping was not significant. Instead, we manually extended the annotated 3'-UTRs of the NCBI Equus caballus annotation release 103 by 2 kb, which increased the read mapping rate to 53.4%. This corresponds to the mapping rates found in other species and indicates that the initial low read mapping was primarily due to inadequate annotation of the 3'-ends, rather than the absence of BALF-specific isoforms. However, because our custom annotation was based on long-read sequencing, it likely contains improved 3' end annotations, which should have enhanced read mapping. The absence of a significant improvement in read mapping is likely attributed to the fact that our custom transcriptome covered only 60% of the 21,129 coding transcripts from the EquCab3.0 NCBI annotation release 103. Many of our scRNA-seq transcripts were not represented in our custom transcriptome. Given the stochastic nature of gene expression, performing long-read sequencing on a larger number of samples may have helped capture a more comprehensive picture of the BALF transcriptome. Although the use of a custom equine BALF long-read transcriptome as a reference did not improve the quality of scRNA-seq data analysis, our BALF-specific transcriptome holds promise for future equine bulk or single-cell RNA-seq studies that investigate isoform-level transcripts. We propose integrating it into the ongoing construction of the atlas of equine tissue-specific transcriptomes led by the FAANG consortium.

2. Annotating cell clusters and interpreting scRNA-seq data: lessons learned and challenges ahead

The interpretation of scRNA-seq data begins with the annotation of cell clusters. In Study 1, we initially attempted automated annotation, expecting it to be a fast and reliable approach. To incorporate cell labels from lower airways, we used human lung cell and ferret BALF cell scRNA-seq datasets as references. However, the automated annotation did not yield meaningful results. Consequently, we resorted to manual annotation of the major cell groups and cell subtypes. When we used Study 1 as a reference dataset for automated annotation in Study 3, we achieved accurate labeling for almost all cells (99.6%). This suggests that gene markers for cell types in scRNA-seq analysis are specific to each species.

However, automated annotation did not perform well for cell subtypes. This is partly because not all cell subtypes present in Study 3 had been detected in Study 1. The differences in horse phenotypes between the two studies, as well as the increased cell numbers per sample in Study 3, may have also contributed to this discrepancy. Notably, the presence of BALF neutrophilia, a characteristic of SEA, enabled the detection of several neutrophil subtypes in Study 3. To build a comprehensive panel of markers for all cell subtypes present in BALF, it will be necessary to perform scRNA-seq on BALF samples from horses with various types of diseases.

Interestingly, we observed differences in gene expression patterns between Study 1 and Study 3. For instance, the expression levels of CD4 and CD8 markers clearly differentiated T helper cells from cytotoxic cells in Study 1, whereas this distinction was less evident in Study 3. Similarly, the differentiation between monocytes and alveolar macrophages (AMs) based on the expression of AM canonical markers was more pronounced in Study 1 than in Study 3. We suspect that the higher biological variability in Study 3, resulting from the inclusion of both asthmatic and healthy horses, contributed to this observation. The disease-associated alterations in gene expression may have overshadowed the cell population-specific signals.

Despite these differences, we were able to identify similar cell types and subtypes in both studies. Furthermore, our cell populations largely matched the clusters identified in another recent scRNA-seq study on equine BALF (88), indicating the reproducibility of our annotation. For example, we identified natural killer T (NKT) cells, γδ T cells, FCN1^{high} AMs, and intermediate monocytes in both our scRNA-seq datasets. However, although these clusters exhibited a similar signature, the gene markers for each cluster were not identical. This underscores the fact that a single marker cannot define a cell population. It is crucial to utilize a curated list of markers, with their expression levels taken into consideration when deciphering the identity of a population. In the future, developing algorithms capable of recognizing specific patterns within a cell population could facilitate annotation.

An important takeaway is that the cell type markers in scRNA-seq analysis differ significantly from those used in antibody-based techniques like flow cytometry. A striking example is the low expression of CD14 in our datasets, a surface marker commonly used to identify monocytes in horses. Additionally, we were unable to identify typical Th1 and Th2 cells based on the expression of their associated cytokines. This discrepancy could be attributed to data sparsity, as cytokine genes may be expressed at lower levels compared to other genes, making them more challenging to capture with scRNA-seq. This highlights the need to establish

scRNA-seq-specific marker sets for cell types. It is also worth considering that the categorization of Th1 and Th2 cells may not accurately reflect the biological reality, as T helper cells may exist as a continuum of cell states transitioning from one to the other.

In our scRNA-seq studies, we could not always confidently identify the cell clusters. We exercised caution in labeling to avoid carrying uncertain annotations forward to future studies. For instance, one AM cell cluster was labeled as "undetermined AM" in Study 3. We also observed unexpected gene expression signals, such as immunoglobulins in T cells and monocyte-macrophages, which we could not definitively explain. Given their relatively low expression levels, these signals may be spurious events. However, without a clear understanding of what constitutes a normal level of gene expression in scRNA-seq data, definitive conclusions are difficult to draw. Through scRNA-seq, we may come to realize that some genes previously thought to be cell-specific, such as immunoglobulins in B cells, may be more widely expressed than initially believed.

To validate our annotation, we considered employing the CITE-seq solution from 10X Genomics. CITE-seq combines classic scRNA-seq with the quantitative determination of surface epitopes, enabling direct linkage of genetic expression to cell surface markers. This would have allowed us to establish connections between our results and previous studies conducted using flow cytometry. However, the conjugation of DNA barcodes with equine-specific antibodies, a prerequisite for CITE-seq, proved to be cost-prohibitive. In the future, collaborative efforts to create a horse-specific library of preconjugated DNA barcodes would be valuable, enabling the combined single-cell immunophenotyping and transcriptome sequencing of equine samples.

3. Discrepancies in cell composition analysis: comparing scRNA-seq and cytology methods

In Study 1, we observed a marked discrepancy in the lymphocyte/macrophage ratio depending on the method used for cell counting. The ratio was five times higher with scRNA-seq compared to cytology. Due to the small size of our study population (three horses), we were unable to statistically analyze this difference. A similar difference in the lymphocyte/macrophage ratio between scRNA-seq and cytological analysis was also present in the BALF of four healthy dogs in a study conducted by Fastrès and colleagues (89). We suspected that this inconsistency arose from the underestimation of lymphocytes or overestimation of macrophages during manual cytological counting, as previous research has

demonstrated when compared to flow cytometric counting (90,91). Another possible explanation could be the underestimation of macrophages using scRNA-seq due to their higher vulnerability to RNA degradation.

However, in Study 3, the statistical analysis did not reveal a significant difference in the composition of the five major leukocyte types between cytology and scRNA-seq. Nonetheless, there was a clear tendency towards an increased lymphocyte/macrophage ratio with scRNA-seq, except for two horses (one asthmatic and one control) that exhibited an inverse trend. The reason for this inverse pattern in the two horses remains unclear, but it may have limited the power of our analysis, resulting in a non-significant outcome. To reach a definitive conclusion, a larger sample population should be used to compare the differential cell count of the five major leukocyte types between cytology and scRNA-seq.

4. Unraveling the Th17-polarized immune response in SEA

The application of scRNA-seq analysis has provided unparalleled insights into the molecular pathways of SEA at the cellular level. In our thesis project, we focused on analyzing BALF rather than blood samples, as BALF was assumed to reflect more accurately the local immune response. It is worth noting that the impact of SEA, both in terms of cell populations and function, differs between BALF and blood (16), with blood likely to be more representative of the systemic inflammatory response and less specific for pulmonary immune mechanisms.

Apart from neutrophils, which were found to be elevated in SEA horses based on our inclusion criteria, only the proportion of B cells was affected by the disease, which is a novel finding in horses. This finding contradicts a recent flow cytometric study (16), where B cells were identified in BALF but were not found to be influenced by EA. However, in that study, B cells could not be confidently identified due to the overall low expression of the CD79A surface marker, highlighting the advantages of utilizing the scRNA-seq technique.

We did not observe any significant differences in the distribution of cell subtypes between SEA and control horses, except for B cells. This suggests that gene dysregulation plays a more substantial role in the pathophysiology of SEA than the expansion of specific disease-associated cell populations. Therefore, to accurately investigate a disease, it is recommended to supplement flow cytometry assays, which quantify cell populations in mixed samples, with transcriptomics experiments or functional in vitro assays.

In Study 3, we observed evidence of a Th17 (or type 3) polarized immune response. Previous studies in horses had hinted at a potential Th17 response (11,31–33), sometimes accompanied by a Th2 response (33). However, in Study 3, we only identified upregulation of two Th2-related genes, *PRB1* (92) and *NMUR1* (93), in the monocyte-macrophage cell group. It is important to note that these genes are not widely accepted as definitive Th2 markers. Their association with a Th2 response is derived from only one study each, which involved asthmatic patients without distinguishing between different endotypes. Furthermore, the lower proportion of switched B cells, which are capable of producing IgG and IgE and initiating an effective Th2 response, supports the notion of a non-Th2 response in SEA. A recent flow cytometric study in equine BALF similarly found no increase in Th1 or Th2-associated cytokines in relation to EA (16).

The Th17-biased response observed in Study 3 was driven by T cells, monocytes and AMs. Among T cells, there was a notable Th17 signature in *CD4*$^{+}$ naïve T cells, γδ T cells and Treg cells. Interestingly, when T cells sorted from the BALF of horses with SEA are stimulated in vitro, they exhibit a Th17 polarization characterized by an increased frequency of IL17A-secreting lymphocytes (16). The Th17 bias and dysregulation of genes associated with T function in Treg cells is of particular interest. It has been previously demonstrated that the proportion of Treg cells increase in horses with SEA (94), especially following glucocorticoid therapy (95). Our findings provide additional support to the role of Treg cells in the pathophysiology of asthma, indicating that they may contribute by undergoing a transition to a Th17 phenotype.

We detected a Th17 polarization in monocytes-macrophages, cells that are often overlooked in the study of SEA. These populations are not typically considered among the conventional inflammatory leukocyte types such as neutrophils, mast cells, and eosinophils. However, there is a growing recognition of their potential role in the development of EA (90,96–98). One particularly exciting finding was the upregulation of *CXCL13* in intermediate monocytes, which aligns with a hypothesis put forth by our research group (32). *CXCL13* is a chemokine that attracts B cells and is produced by Th17 cells and activated monocytes. In a previous study using RNA sequencing, our group demonstrated the upregulation of *CXCL13* in PBMCs of asthmatic horses following in vitro stimulation (32). This upregulation was primarily attributed to activated monocytes, although other possibilities included upregulation in Th17 cells or expansion of either of the CXCL13-producing cell populations. We were able to resolve this dilemma through the use of scRNA-seq. Since *CXCL13* is upregulated in both blood and

BALF, it emerges as a particularly interesting candidate gene for further investigation as a biomarker for SEA.

In humans, scRNA-seq studies have predominantly focused on type-2 diseases, primarily utilizing murine models of allergic asthma (68). However, one study combined microarray and scRNA-seq to explore the association between combustion-derived particulate matter (PM) and neutrophilic asthma. Using a murine model, they demonstrated increased IL17 signaling in respiratory epithelial cells, activating a pathway that leads to the induction of pathogenic Th17 cells formation and subsequent airway neutrophilia (99). It would be interesting to conduct a similar investigation on structural pulmonary cells of horses with SEA to determine whether the Th17-specific inflammatory response is initiated in epithelial cells, which are the first to encounter inhaled antigens.

Prior investigations have demonstrated that SEA is linked to increased persistence and reduced apoptosis of neutrophils, as well as enhanced NET formation (42,43,100). Our scRNA-seq analysis revealed similar characteristics, allowing us to further distinguish the unique features of each identified neutrophil subtype. Apoptotic neutrophils exhibited elevated NETosis activity but also exhibited increased expression of anti-inflammatory genes, diminished migratory capacity, and reduced survival potential. These findings suggest that this particular neutrophil population may play a protective role in asthma. On the other hand, the pro-inflammatory neutrophils and ISG^{high} neutrophils displayed enhanced migratory capacity, indicating their potential involvement in airway neutrophilia. These findings have important implications for the development of novel therapies, as targeting solely the pro-inflammatory neutrophils could preserve the anti-inflammatory and anti-microbial functions of the other neutrophil populations.

Based on the findings of this thesis project, we propose that SEA is primarily driven by a Th17 immune response. It is important to note that these findings specifically pertain to neutrophilic SEA, as a separate scRNA-seq study on horses with predominantly mastocytic EA revealed distinct cell populations and DEGs (88). Further support for distinct mechanisms in MEA and SEA comes from flow cytometric analysis of BALF and functional assays on T cells (16). Therefore, it is crucial to consider and investigate MEA and SEA as separate endotypes with their own unique characteristics.

5. Identification and significance of monocyte-lymphocyte complexes in SEA

An intriguing discovery in our scRNA-seq experiments was the identification of cells expressing markers from both T cell and monocyte lineages. In previous scRNA-seq studies, such cells were considered artifacts and disregarded (101–103). However, recent research in humans has shed light on the importance of these double positive cell populations. Burel et al. detected cells expressing conflicting lineage markers CD3 and CD14 in human blood using flow cytometry and scRNA-seq. Through imaging flow cytometry, they demonstrated that these cells were indeed monocyte-T cell complexes, exhibiting LFA-1/ICAM-1 polarization at their point of contact. These cell-cell complexes were consistently present in fresh PBMCs and whole blood, regardless of sample preservation or manipulation, suggesting they were not a result of ex vivo handling (104).

We suspect that the identified monocyte-T cell and monocyte-B cell clusters in Study 1 and Study 3 represent similar interacting cell populations and are not simply a random co-encapsulation artifact from the scRNA-seq microfluidics platform. Several factors support this hypothesis. Firstly, these clusters were consistently identified in both studies and in another recent scRNA-seq study on equine BALF (88). Secondly, if co-encapsulation were random, the cell types found together in the droplets would vary, but we consistently observed monocytes paired with lymphocytes. The upregulation of genes associated with lymphocyte activation within these clusters further supports our hypothesis. However, to validate this hypothesis, complementary experimental techniques such as imaging flow cytometry or microscopy imaging following cell sorting should be used (75).

The presence of monocyte-lymphocyte complexes in the BALF could have significant clinical implications. In humans, their frequency has been shown to change depending on the type of immune perturbations (104). The Th17 polarization and upregulation of inflammasome-associated genes in the double positive monocyte-lymphocyte cluster in Study 3 similarly suggest these cell-cell complexes may have a role in the pathogenesis of SEA. They could represent a novel type of biomarker or prove to be relevant therapeutic targets.

6. Limitations and considerations in scRNA-seq analysis of equine BALF: implications for study design and data interpretation

Although this thesis project marked a significant advancement in the application of scRNA-seq to equine BALF and provided insights into the mechanisms of SEA at the single-cell level, we must acknowledge several limitations. Firstly, the sample size across all three

studies was small. The initial study served as a pilot to demonstrate the feasibility of scRNA-seq in equine BALF and develop a methodology for subsequent investigations. While including only three horses may seem limited, it proved sufficient to identify the major immune cell types in equine BALF. Given the proof-of-concept nature of the study, it could have been more appropriate to utilize healthy horses without respiratory disease. We used the three horses available on-site, for which we had a permission for animal use. These horses exhibited HOARSI scores and BALF cytology consistent with MEA, and they were considered to be in remission, as indicated by their low clinical scores at the time of sample collection. Our preliminary data enabled us to obtain animal use permission for client-owned horses, which allowed us to incorporate horses with SEA and control subjects into the study (Study 3). Ultimately, this approach demonstrated that similar cell types and subtypes could be identified across three distinct populations with different phenotypes: MEA, SEA, and control groups. In study 2, the inclusion of only two horses was driven by the need to manage sequencing costs. By sequencing BALF samples from a healthy horse and a horse with SEA, we aimed to capture transcriptomes representing both phenotypes. However, expanding the sample size could have enhanced transcript coverage and improved the detection of alternate isoforms. In Study 3, we included 12 horses based on a power analysis. Unfortunately, the scRNA-seq experiment failed for one of the control samples. Nevertheless, we sampled more than ten times the number of cells/sample than originally planned in the power analysis (500 cells/sample), which increases the likelihood of detecting a biological signal. However, to ensure the robustness of our findings, it is imperative to validate the reproducibility of our results by conducting scRNA-seq on a larger and more diverse study population.

The horses included in Study 1 were a convenience sample of teaching horses available on-site. For Study 2 and Study 3, a combination of criteria was utilized to classify the horses as either healthy or asthmatic. In previous studies, the focus had primarily been on BALF cytology to categorize horses with MEA and SEA, with SEA defined as BALF neutrophils exceeding 25% (4). In line with other research groups, we employed the less strict criteria of 10% BALF neutrophilia (16) and incorporated historical and clinical parameters to characterize our case and control groups (15,16). This approach to case selection is being increasingly adopted in equine respiratory research, as it better reflects the diversity of clinical presentations observed. Consequently, findings obtained in a research setting may have greater relevance to clinical practice. Ideally, adequate phenotyping would have been confirmed using lung function testing. Esophageal balloon catheterization is considered the gold standard for accurately characterizing

horses with SEA and those without respiratory disease (9). However, this approach was not feasible in client-owned horses due to perceived invasiveness.

All experiments conducted in this project utilized fresh frozen equine BALF cells. However, we did not specifically investigate the impact of freezing on cell viability and gene expression. A significant portion of this project involved optimizing the protocol for cell cryopreservation. We took great care to minimize cell manipulation and employed gentle techniques to avoid cell bursting or stress-related changes in gene expression. Although small clusters of dead cells were observed in the scRNA-seq studies, these cells were easily identifiable and filtered out to ensure they did not affect the DGE analysis. Several studies have indicated that the impact of sample preservation is minimal in scRNA-seq (70), and cryopreservation is increasingly considered the standard approach. Nevertheless, it would be ideal to conduct a dedicated study to specifically examine the impact of cryopreservation on the transcriptome of equine BALF cells.

Our interpretation of the DEGs identified with scRNA-seq was based on human datasets, primarily the HPA (61). Due to the limited availability of transcriptomic and proteomic data in horses compared to humans and conventional animal models, there is a lack of equine-specific databases linking genes to their functionality. Therefore, we had to rely on the assumption that gene function is relatively conserved across species. Notably, in horses, certain genes that are typically associated with asthma exhibit higher homology to humans compared to mice (25), supporting the validity of our assumption. Nevertheless, it is highly desirable to have functional mapping of the equine genome, and efforts in this direction are gradually progressing, thanks to initiatives such as the FAANG consortium (81).

Outlook

In this thesis project, we generated three major hypotheses:

1) SEA is a Th17 driven disease
2) CXCL13 produced by activated monocytes is a key signaling factor in SEA
3) Monocyte-lymphocyte complexes are present in equine BALF

We plan to challenge these hypotheses through the experiments described below.

1. **Bulk RNA-seq deconvolution of equine BALF cells' bulk transcriptomes**

Deconvolution is a computational technique that enables the retrieval of single-cell transcriptome information from a bulk transcriptome when the different cell types present in the sample are known (105). Building upon the insights gained from Study 1 and Study 3, we aim to apply this methodology to bulk transcriptome data obtained from equine BALF. The application of deconvolution analysis has the potential to access the single-cell immunological landscape at a reduced cost. In the future, we envision a scenario where BALF collected during routine respiratory work-up in asthmatic horses could undergo RNA-seq, enabling the detection of the specific immune pathways involved. This approach would set the stage for precision medicine, where each horse could receive a treatment tailored to its asthma endotype.

2. **Potential of CXCL13 as a biomarker of SEA**

We propose to investigate the diagnostic potential of CXCL13 in SEA by measuring its expression levels in the blood and BALF of horses presented with respiratory disease. By studying its association with SEA, we aim to determine whether CXCL13 could serve as a biomarker for this condition.

3. **Imaging flow cytometry of equine BALF cells**

To challenge our hypothesis, we plan to utilize imaging flow cytometry or imaging microscopy following cell sorting on equine BALF samples. This approach will provide visual confirmation and detailed analysis of the identified monocyte-lymphocyte complexes in equine BALF, contributing to a better understanding of their role in the disease.

In the years to come, researchers should leverage the potential of scRNA-seq to investigate the impact of specific antigen exposure in interventional studies, such as hay challenge or corticosteroid therapy. Furthermore, efforts should be expanded to explore other endotypes of equine asthma. Recent work by Riihimäki and colleagues (88) focused on scRNA-seq analysis of horses with mastocytic asthma, which appears to be the predominant form of EA in Sweden. It is essential to investigate horse populations with diverse genetic backgrounds, varied environmental exposures, and distinct clinical presentations to gain a comprehensive understanding of all the endotypes of equine asthma.The advent of scRNA-seq has opened new avenues for studying disease mechanisms, which will significantly expedite the discovery of biomarkers and therapeutic targets. Ultimately, we hope that these advancements will lead to substantial improvements in diagnosing and managing equine asthma.

References

1. Network GA. The Global Asthma Report 2022. *Int J Tuberc Lung Dis* (2022) **26**:1–104. doi:10.5588/ijtld.22.1010

2. Hotchkiss JW, Reid SWJ, Christley RM. A survey of horse owners in Great Britain regarding horses in their care. Part 2: Risk factors for recurrent airway obstruction. *Equine Vet J* (2007) **39**:301–308. doi:10.2746/042516407X180129

3. Ivester KM, Couëtil LL, Moore GE. An observational study of environmental exposures, airway cytology, and performance in racing thoroughbreds. *J Vet Intern Med* (2018) **32**:1754–1762. doi:10.1111/jvim.15226

4. Couëtil LLL, Cardwell JMM, Gerber V, Lavoie J -P. P, Léguillette R, Richard EAA. Inflammatory Airway Disease of Horses-Revised Consensus Statement. *J Vet Intern Med* (2016) **30**:503–515. doi:10.1111/jvim.13824

5. Bond S, Léguillette R, Richard EA, Couetil L, Lavoie JP, Martin JG, Pirie RS. Equine asthma: Integrative biologic relevance of a recently proposed nomenclature. *J Vet Intern Med* (2018) **32**:2088–2098. doi:10.1111/jvim.15302

6. Bessonnat A, Hélie P, Grimes C, Lavoie J. Airway remodeling in horses with mild and moderate asthma. *J Vet Intern Med* (2022) **36**:285–291. doi:10.1111/jvim.16333

7. Laumen E, Doherr MG, Gerber V. Relationship of horse owner assessed respiratory signs index to characteristics of recurrent airway obstruction in two Warmblood families. *Equine Vet J* (2010) **42**:142–148. doi:10.2746/042516409X479586

8. Ramseyer A, Gaillard C, Burger D, Straub R, Jost U, Boog C, Marti E, Gerber V. Effects of Genetic and Environmental Factors on Chronic Lower Airway Disease in Horses. *J Vet Intern Med* (2007) **21**:149–156. doi:10.1111/j.1939-1676.2007.tb02941.x

9. Couetil L, Cardwell JM, Leguillette R, Mazan M, Richard E, Bienzle D, Bullone M, Gerber V, Ivester K, Lavoie J-PP, et al. Equine Asthma: Current Understanding and Future Directions. *Front Vet Sci* (2020) **7**:1–21. doi:10.3389/fvets.2020.00450

10. Holcombe SJ, Jackson C, Gerber V, Jefcoat A, Berney C, Eberhardt S, Robinson NE. Stabling is associated with airway inflammation in young Arabian horses. *Equine Vet J* (2010) **33**:244–249. doi:10.2746/042516401776249606

11. Debrue M, Hamilton E, Joubert P, Lajoie-Kadoch S, Lavoie J-P. Chronic exacerbation of equine heaves is associated with an increased expression of interleukin-17 mRNA in bronchoalveolar lavage cells. *Vet Immunol Immunopathol* (2005) **105**:25–31. doi:10.1016/j.vetimm.2004.12.013

12. Gerber V, Straub R, Marti E, Hauptman J, Herholz C, King M, Imhof A, Tahon L, Robinson NE. Endoscopic scoring of mucus quantity and quality: Observer and horse variance and relationship to inflammation, mucus viscoelasticity and volume. *Equine Vet J* (2004) **36**:576–582. doi:10.2746/0425164044864525

13. Couëtil LL, Rosenthal FS, Simpson CM. Forced expiration: a test for airflow obstruction in horses. *J Appl Physiol* (2000) **88**:1870–1879. doi:10.1152/jappl.2000.88.5.1870

14. Bullone M, Lavoie J-P. Science-in-brief: Equine asthma diagnosis: Beyond bronchoalveolar lavage cytology. *Equine Vet J* (2017) **49**:263–265. doi:10.1111/evj.12679

15. Mahalingam-Dhingra A, Bedenice D, Mazan MR. Bronchoalveolar lavage hemosiderosis in lightly active or sedentary horses. *J Vet Intern Med* (2023) doi:10.1111/jvim.16692

16. Gressler AE, Lübke S, Wagner B, Arnold C, Lohmann KL, Schnabel CL. Comprehensive Flow Cytometric Characterization of Bronchoalveolar Lavage Cells Indicates Comparable Phenotypes Between Asthmatic and Healthy Horses But Functional Lymphocyte Differences. *Front Immunol* (2022) **13**:1–17. doi:10.3389/fimmu.2022.896255

17. Leclere M, Lavoie-Lamoureux A, Joubert P, Relave F, Setlakwe EL, Beauchamp G, Couture C, Martin JG, Lavoie J-P. Corticosteroids and Antigen Avoidance Decrease Airway Smooth Muscle Mass in an Equine Asthma Model. *Am J Respir Cell Mol Biol* (2012) **47**:589–596. doi:10.1165/rcmb.2011-0363OC

18. Mazan MR, Deveney EF, DeWitt S, Bedenice D, Hoffman A. Energetic cost of breathing, body composition, and pulmonary function in horses with recurrent airway obstruction. *J Appl Physiol* (2004) **97**:91–97. doi:10.1152/japplphysiol.00629.2003

19. Baumgartner M, Boisson T, Erhard MH, Zeitler-Feicht MH. Common Feeding Practices Pose A Risk to the Welfare of Horses When Kept on Non-Edible Bedding. *Animals*

(2020) **10**:411. doi:10.3390/ani10030411

20. Rendle D. Steroids and laminitis - the value of evidence over anecdote. *Vet Rec* (2019) **185**:79–81. doi:10.1136/vr.l4730

21. Sleeper MM, Kearns CF, Mckeever KH. Chronic clenbuterol administration negatively alters cardiac function. *Med Sci Sport Exerc* (2002) **34**:643–650.

22. Jeffcott LB, Rossdale PD, Freestone J, Frank CJ, Towers-Clark PF. An assessment of wastage in Thoroughbred racing from conception to 4 years of age. *Equine Vet J* (1982) **14**:185–198. doi:10.1111/j.2042-3306.1982.tb02389.x

23. Rossdale P, Hopes R, Digby N, Offord K. Epidemiological study of wastage among racehorses 1982 and 1983. *Vet Rec* (1985) **116**:66–69. doi:10.1136/vr.116.3.66

24. Woodrow JS, Sheats MK, Cooper B, Bayless R. Asthma: The Use of Animal Models and Their Translational Utility. *Cells* (2023) **12**:1091. doi:10.3390/cells12071091

25. Lange-Consiglio A, Stucchi L, Zucca E, Lavoie JP, Cremonesi F, Ferrucci F. Insights into animal models for cell-based therapies in translational studies of lung diseases: Is the horse with naturally occurring asthma the right choice? *Cytotherapy* (2019) **21**:525–534. doi:10.1016/j.jcyt.2019.02.010

26. Annunziato F, Romagnani C, Romagnani S. The 3 major types of innate and adaptive cell-mediated effector immunity. *J Allergy Clin Immunol* (2015) **135**:626–635. doi:10.1016/j.jaci.2014.11.001

27. Kuruvilla ME, Lee FE-H, Lee GB. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. *Clin Rev Allergy Immunol* (2019) **56**:219–233. doi:10.1007/s12016-018-8712-1

28. Simões J, Batista M, Tilley P. The Immune Mechanisms of Severe Equine Asthma—Current Understanding and What Is Missing. *Animals* (2022) **12**:744. doi:10.3390/ani12060744

29. Kehrli D, Jandova V, Fey K, Jahn P, Gerber V. Multiple Hypersensitivities Including Recurrent Airway Obstruction, Insect Bite Hypersensitivity, and Urticaria in 2 Warmblood Horse Populations. *J Vet Intern Med* (2015) **29**:320–326. doi:10.1111/jvim.12473

30. Lanz S, Brunner A, Graubner C, Marti E, Gerber V. Insect Bite Hypersensitivity in Horses is Associated with Airway Hyperreactivity. *J Vet Intern Med* (2017) **31**:1877–1883. doi:10.1111/jvim.14817

31. Korn A, Miller D, Dong L, Buckles EL, Wagner B, Ainsworth DM. Differential Gene Expression Profiles and Selected Cytokine Protein Analysis of Mediastinal Lymph Nodes of Horses with Chronic Recurrent Airway Obstruction (RAO) Support an Interleukin-17 Immune Response. *PLoS One* (2015) **10**:e0142622. doi:10.1371/journal.pone.0142622

32. Pacholewska A, Kraft MF, Gerber V, Jagannathan V. Differential expression of serum MicroRNAs supports CD4 + t cell differentiation into Th2/Th17 cells in severe equine asthma. *Genes (Basel)* (2017) **8**: doi:10.3390/genes8120383

33. Hulliger MF, Pacholewska A, Vargas A, Lavoie J, Leeb T, Gerber V, Jagannathan V. An Integrative miRNA-mRNA Expression Analysis Reveals Striking Transcriptomic Similarities between Severe Equine Asthma and Specific Asthma Endotypes in Humans. *Genes (Basel)* (2020) **11**:1143. doi:10.3390/genes11101143

34. Pacholewska A, Jagannathan V, Drögemüller M, Klukowska-Rötzler J, Lanz S, Hamza E, Dermitzakis ET, Marti E, Leeb T, Gerber V. Impaired Cell Cycle Regulation in a Natural Equine Model of Asthma. *PLoS One* (2015) **10**:e0136103. doi:10.1371/journal.pone.0136103

35. Cordeau M-E, Joubert P, Dewachi O, Hamid Q, Lavoie J-P. IL-4, IL-5 and IFN-γ mRNA expression in pulmonary lymphocytes in equine heaves. *Vet Immunol Immunopathol* (2004) **97**:87–96. doi:10.1016/j.vetimm.2003.08.013

36. Kleiber C, Grunig G, Jungi T, Schmucker N, Gerber H, Davis WC, Straub R. Phenotypic Analysis of Bronchoalveolar Lavage Fluid Lymphocytes in Horses with Chronic Pulmonary Disease. *J Vet Med Ser A* (1999) **46**:177–184. doi:10.1046/j.1439-0442.1999.00210.x

37. Giguère S, Viel L, Lee E, MacKay RJ, Hernandez J, Franchini M. Cytokine induction in pulmonary airways of horses with heaves and effect of therapy with inhaled fluticasone propionate. *Vet Immunol Immunopathol* (2002) **85**:147–158. doi:10.1016/S0165-2427(01)00420-2

38. Horohov DW, Beadle RE, Mouch S, Pourciau SS. Temporal regulation of cytokine mRNA expression in equine recurrent airway obstruction. *Vet Immunol Immunopathol* (2005) **108**:237–245. doi:10.1016/j.vetimm.2005.07.013

39. Padoan E, Ferraresso S, Pegolo S, Castagnaro M, Barnini C, Bargelloni L. Real time RT-PCR analysis of inflammatory mediator expression in recurrent airway obstruction-affected horses. *Vet Immunol Immunopathol* (2013) **156**:190–199. doi:10.1016/j.vetimm.2013.09.020

40. Ainsworth DM, Grünig G, Matychak MB, Young J, Wagner B, Erb HN, Antczak DF. Recurrent airway obstruction (RAO) in horses is characterized by IFN-γ and IL-8 production in bronchoalveolar lavage cells. *Vet Immunol Immunopathol* (2003) **96**:83–91. doi:10.1016/S0165-2427(03)00142-9

41. Ainsworth DM, Wagner B, Franchini M, Grünig G, Erb HN, Tan J-Y. Time-dependent alterations in gene expression of interleukin-8 in the bronchial epithelium of horses with recurrent airway obstruction. *Am J Vet Res* (2006) **67**:669–677. doi:10.2460/ajvr.67.4.669

42. Janssen P, Tosi I, Hego A, Maréchal P, Marichal T, Radermecker C. Neutrophil Extracellular Traps Are Found in Bronchoalveolar Lavage Fluids of Horses With Severe Asthma and Correlate With Asthma Severity. *Front Immunol* (2022) **13**:1–18. doi:10.3389/fimmu.2022.921077

43. Vargas A, Boivin R, Cano P, Murcia Y, Bazin I, Lavoie JP. Neutrophil extracellular traps are downregulated by glucocorticosteroids in lungs in an equine model of asthma. *Respir Res* (2017) **18**:1–11. doi:10.1186/s12931-017-0689-4

44. Schnider D, Rieder S, Leeb T, Gerber V, Neuditschko M. A genome-wide association study for equine recurrent airway obstruction in European Warmblood horses reveals a suggestive new quantitative trait locus on chromosome 13. *Anim Genet* (2017) **48**:691–693. doi:10.1111/age.12583

45. Pacholewska A, Drögemüller M, Klukowska-Rötzler J, Lanz S, Hamza E, Dermitzakis ET, Marti E, Gerber V, Leeb T, Jagannathan V. The transcriptome of equine peripheral blood mononuclear cells. *PLoS One* (2015) **10**:1–16. doi:10.1371/journal.pone.0122011

46. Sheats MK, Davis KU, Poole JA. Comparative Review of Asthma in Farmers and

Horses. *Curr Allergy Asthma Rep* (2019) **19**:50. doi:10.1007/s11882-019-0882-2

47. Couetil LL, Thompson CA. Airway Diagnostics: Bronchoalveolar Lavage, Tracheal Wash, and Pleural Fluid. *Vet Clin North Am - Equine Pract* (2020) **36**:87–103. doi:10.1016/j.cveq.2019.12.006

48. Tomlinson JE, Wagner B, Felippe MJB, Van de Walle GR. Multispectral fluorescence-activated cell sorting of B and T cell subpopulations from equine peripheral blood. *Vet Immunol Immunopathol* (2018) **199**:22–31. doi:10.1016/j.vetimm.2018.03.010

49. Davis EG, Wilkerson MJ, Rush BR. Flow cytometry: Clinical applications in equine medicine. *J Vet Intern Med* (2002) **16**:404–410. doi:10.1111/j.1939-1676.2002.tb01257.x

50. Lavoie JP, Cesarini C, Lavoie-Lamoureux A, Moran K, Lutz S, Picandet V, Jean D, Marcoux M. Bronchoalveolar Lavage Fluid Cytology and Cytokine Messenger Ribonucleic Acid Expression of Racehorses with Exercise Intolerance and Lower Airway Inflammation. *J Vet Intern Med* (2011) **25**:322–329. doi:10.1111/j.1939-1676.2010.0664.x

51. Hughes KJ, Nicolson L, Da Costa N, Franklin SH, Allen KJ, Dunham SP. Evaluation of cytokine mRNA expression in bronchoalveolar lavage cells from horses with inflammatory airway disease. *Vet Immunol Immunopathol* (2011) **140**:82–89. doi:10.1016/j.vetimm.2010.11.018

52. Beekman L, Tohver T, Léguillette R. Comparison of Cytokine mRNA Expression in the Bronchoalveolar Lavage Fluid of Horses with Inflammatory Airway Disease and Bronchoalveolar Lavage Mastocytosis or Neutrophilia Using REST Software Analysis. *J Vet Intern Med* (2012) **26**:153–161. doi:10.1111/j.1939-1676.2011.00847.x

53. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* (2009) **6**:377–382. doi:10.1038/nmeth.1315

54. Nayak R, Hasija Y. A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. *Genomics* (2021) **113**:606–619. doi:10.1016/j.ygeno.2021.01.007

55. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* (2017) **9**:1–

12. doi:10.1186/s13073-017-0467-4

56. Lafzi A, Moutinho C, Picelli S, Heyn H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc* (2018) **13**:2742–2757. doi:10.1038/s41596-018-0073-y

57. Salomon R, Kaczorowski D, Valdes-Mora F, Nordon RE, Neild A, Farbehi N, Bartonicek N, Gallego-Ortega D. Droplet-based single cell RNAseq tools: A practical guide. *Lab Chip* (2019) **19**:1706–1727. doi:10.1039/c8lc01239c

58. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* (2019) **15**: doi:10.15252/msb.20188746

59. Kharchenko P V. The triumphs and limitations of computational methods for scRNA-seq. *Nat Methods* (2021) **8**: doi:10.1038/s41592-021-01171-x

60. Clarke ZA, Andrews TS, Atif J, Pouyabahar D, Innes BT, MacParland SA, Bader GD. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat Protoc* (2021) **16**:2749–2764. doi:10.1038/s41596-021-00534-0

61. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, Oksvold P, Edfors F, Limiszewska A, Hikmet F, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science (80-)* (2020) **367**: doi:10.1126/science.aay5947

62. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* (2016) **34**:1145–1160. doi:10.1038/nbt.3711

63. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res* (2015) **25**:1491–1498. doi:10.1101/gr.190595.115

64. Alexander MJ, Scott Budinger GR, Reyfman PA. Breathing fresh air into respiratory research with single-cell rna sequencing. *Eur Respir Rev* (2020) **29**:1–14. doi:10.1183/16000617.0060-2020

65. Chambers DC, Carew AM, Lukowski SW, Powell JE. Transcriptomics and single-cell RNA-sequencing. *Respirology* (2019) **24**:29–36. doi:10.1111/resp.13412

66. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit R V., Chang S, Conley SD, Mori Y, Seita J, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* (2020) **587**:619–625. doi:10.1038/s41586-020-2922-4

67. Vieira Braga FA, Kar G, Berg M, Carpaij OA, Polanski K, Simon LM, Brouwer S, Gomes T, Hesse L, Jiang J, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* (2019) **25**:1153–1163. doi:10.1038/s41591-019-0468-5

68. Tang W, Li M, Teng F, Cui J, Dong J, Wang W. Single-cell RNA-sequencing in asthma research. *Front Immunol* (2022) **13**: doi:10.3389/fimmu.2022.988573

69. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. Power analysis of single-cell rnA-sequencing experiments. *Nat Methods* (2017) **14**:381–387. doi:10.1038/nmeth.4220

70. Vieira Braga FA, Teichmann SA, Stubbington MJT. Are cells from a snowman realistic? Cryopreserved tissues as a source for single-cell RNA-sequencing experiments. *Genome Biol* (2017) **18**:54. doi:10.1186/s13059-017-1192-4

71. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, Harding P, Polanski K, Huang N, Nowicki-Osuch K, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol* (2020) **21**:1. doi:10.1186/s13059-019-1906-x

72. Guillaumet-Adkins A, Rodríguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, Vidal A, Martinez-Marti A, Felip E, Vivancos A, et al. Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol* (2017) **18**:1–15. doi:10.1186/s13059-017-1171-9

73. Cain MP, Hernandez- BJ, Chen J. Quantitative single-cell interactomes in normal and virus-infected mouse lungs. *Dis Model Mech* (2020) **13**: doi:10.1242/dmm.044404

74. Burel JG, Chawla A, Greenbaum JA, Peters B. Distinguishing cell–cell complexes from dual lineage cells using single-cell transcriptomics is not trivial. *Cytom Part A* (2022) doi:10.1002/cyto.a.24656

75. Burel JG, Pomaznoy M, Lindestam Arlehamn CS, Seumois G, Vijayanand P, Sette A, Peters B. The Challenge of Distinguishing Cell–Cell Complexes from Singlet Cells in Non-Imaging Flow Cytometry and Single-Cell Sorting. *Cytom Part A* (2020) **97**:1127–1135. doi:10.1002/cyto.a.24027

76. Germain P-L, Lun A, Garcia Meixide C, Macnair W, Robinson MD. Doublet

identification in single-cell sequencing data using scDblFinder. *F1000Research* (2022) **10**:979. doi:10.12688/f1000research.73600.2

77. Baran-Gale J, Chandra T, Kirschner K. Experimental design for single-cell RNA sequencing. *Brief Funct Genomics* (2018) **17**:233–239. doi:10.1093/bfgp/elx035

78. Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* (2018) **19**:562–578. doi:10.1093/biostatistics/kxx053

79. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A, et al. Eleven grand challenges in single-cell data science. *Genome Biol* (2020) **21**:31. doi:10.1186/s13059-020-1926-6

80. Vieth B, Parekh S, Ziegenhain C, Enard W, Hellmann I. A Systematic Evaluation of Single Cell RNA-Seq Analysis Pipelines. *bioRxiv* (2019)583013. doi:10.1101/583013

81. Peng S, Dahlgren AR, Donnelly CG, Hales EN, Petersen JL, Bellone RR, Kalbfleisch T, Finno CJ. Functional annotation of the animal genomes: An integrated annotation resource for the horse. *PLOS Genet* (2023) **19**:e1010468. doi:10.1371/journal.pgen.1010468

82. Kaelin CB, McGowan KA, Barsh GS. Developmental genetics of color pattern establishment in cats. *Nat Commun* (2021) **12**:5127. doi:10.1038/s41467-021-25348-2

83. Kraft F, Kurth I. Long-read sequencing to understand genome biology and cell function. *Int J Biochem Cell Biol* (2020) **126**:105799. doi:10.1016/j.biocel.2020.105799

84. Fichot EB, Norman RS. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome* (2013) **1**:10. doi:10.1186/2049-2618-1-10

85. Shafer MER. Cross-Species Analysis of Single-Cell Transcriptomic Data. *Front Cell Dev Biol* (2019) **7**:1–9. doi:10.3389/fcell.2019.00175

86. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* (2017) **14**:865–868. doi:10.1038/nmeth.4380

87. Deschamps-Francoeur G, Simoneau J, Scott MS. Handling multi-mapped reads in RNA-seq. *Comput Struct Biotechnol J* (2020) **18**:1569–1576. doi:10.1016/j.csbj.2020.06.014

88. Riihimäki M, Fegraeus K, Nordlund J, Waern I, Wernersson S, Akula S, Hellman L, Raine A. Single cell transcriptomics delineates the immune-cell landscape in equine lower airways and reveals upregulation of the FKBP5 gene in horses with asthma, , 07 April 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs.1–29. doi:10.21203/rs.3.rs-2768703/v1

89. Fastrès A, Pirottin D, Fievez L, Marichal T, Desmet CJ, Bureau F, Clercx C. Characterization of the Bronchoalveolar Lavage Fluid by Single Cell Gene Expression Analysis in Healthy Dogs: A Promising Technique. *Front Immunol* (2020) **11**:1–14. doi:10.3389/fimmu.2020.01707

90. Kang H, Bienzle D, Lee GKC, Piché É, Viel L, Odemuyiwa SO, Beeler-Marfisi J. Flow cytometric analysis of equine bronchoalveolar lavage fluid cells in horses with and without severe equine asthma. *Vet Pathol* (2022) **59**:91–99. doi:10.1177/03009858211042588

91. Hodge SJ, Hodge GL, Holmes M, Reynolds PN. Flow cytometric characterization of cell populations in bronchoalveolar lavage and bronchial brushings from patients with chronic obstructive pulmonary disease. *Cytom Part B - Clin Cytom* (2004) **61**:27–34. doi:10.1002/cyto.b.20020

92. Chen F, Liang Y, Zeng Z, Du L, Xu C, Guo Y, Xie C. Association of increased basic salivary proline-rich protein 1 levels in induced sputum with type 2-high asthma. *Immunity, Inflamm Dis* (2022) **10**:1–10. doi:10.1002/iid3.602

93. Ye Y, Luo J, Zeng N, Jiang S, Chen W, Hoyle RD, Klenerman P, Pavord ID, Xue L. Neuromedin U promotes human type 2 immune responses. *Mucosal Immunol* (2022) **15**:990–999. doi:10.1038/s41385-022-00543-6

94. Henríquez C, Perez B, Morales N, Sarmiento J, Carrasco C, Morán G, Folch H. Participation of T regulatory cells in equine recurrent airway obstruction. *Vet Immunol Immunopathol* (2014) **158**:128–134. doi:10.1016/j.vetimm.2013.12.005

95. Boivin R, Vargas A, Cano P, Lavoie JP. Glucocorticosteroids administration is associated with increased regulatory T cells in equine asthmatic lungs. *Vet Immunol Immunopathol* (2018) **201**:67–71. doi:10.1016/j.vetimm.2018.05.010

96. Laan TTJM, Bull S, Pirie R, Fink-Gremmels J. The Role of Alveolar Macrophages in

the Pathogenesis of Recurrent Airway Obstruction in Horses. *J Vet Intern Med* (2006) **20**:167–174. doi:10.1111/j.1939-1676.2006.tb02837.x

97. Wilson ME, McCandless EE, Olszewski MA, Robinson NE. Alveolar macrophage phenotypes in severe equine asthma. *Vet J* (2020) **256**:105436. doi:10.1016/j.tvjl.2020.105436

98. Karagianni AE, Summers KM, Couroucé A, Depecker M, McGorum BC, Hume DA, Pirie RS. The Effect of Race Training on the Basal Gene Expression of Alveolar Macrophages Derived From Standardbred Racehorses. *J Equine Vet Sci* (2019) **75**:48–54. doi:10.1016/j.jevs.2019.01.010

99. Harding JN, Gross M, Patel V, Potter S, Cormier SA. Association between particulate matter containing EPFRs and neutrophilic asthma through AhR and Th17. *Respir Res* (2021) **22**:275. doi:10.1186/s12931-021-01867-w

100. Davis KU, Sheats MK. The Role of Neutrophils in the Pathophysiology of Asthma in Humans and Horses. *Inflammation* (2021) **44**:450–465. doi:10.1007/s10753-020-01362-2

101. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* (2020) **26**:842–844. doi:10.1038/s41591-020-0901-9

102. Evren E, Ringqvist E, Tripathi KP, Sleiers N, Rives IC, Alisjahbana A, Gao Y, Sarhan D, Halle T, Sorini C, et al. Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* (2021) **54**:259-275.e7. doi:10.1016/j.immuni.2020.12.003

103. Fastrès A, Pirottin D, Fievez L, Tutunaru AC, Bolen G, Merveille AC, Marichal T, Desmet CJ, Bureau F, Clercx C. Identification of Pro-Fibrotic Macrophage Populations by Single-Cell Transcriptomic Analysis in West Highland White Terriers Affected With Canine Idiopathic Pulmonary Fibrosis. *Front Immunol* (2020) **11**: doi:10.3389/fimmu.2020.611749

104. Burel JG, Pomaznoy M, Lindestam Arlehamn CS, Weiskopf D, da Silva Antunes R, Jung Y, Babor M, Schulten V, Seumois G, Greenbaum JA, et al. Circulating T cell-monocyte complexes are markers of immune perturbations. *Elife* (2019) **8**:1–21.

doi:10.7554/eLife.46045

105. Im Y, Kim Y. A Comprehensive Overview of RNA Deconvolution Methods and Their Application. *Mol Cells* (2023) **46**:99–105. doi:10.14348/molcells.2023.2178

Declaration of Originality

Last name, first name:

Matriculation number:

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations.
All data, tables, figures and text citations which have been reproduced from any other source, including the internet, have been explicitly acknowledged as such.
I am aware that in case of non-compliance, the Senate is entitled to withdraw the doctorate degree awarded to me on the basis of the present thesis, in accordance with the "Statut der Universität Bern (Universitätsstatut; UniSt)", Art. 69, of 7 June 2011.

Place, date

Signature