**The Blood Proteome Landscape: Taxonomy, Data Resources, and Integration**

Asier Larrea 1,2,3,4 , Cesar Martin1,2,\* , Yasset Perez-Riverol 4,\**,*

1 Department of Biochemistry and Molecular Biology, Universidad del País Vasco UPV/EHU, 48080, Bilbao, Spain

2 Department of Molecular Biophysics, Biofisika Institute, University of Basque Country and Consejo Superior de Investigaciones Científicas (UPV/EHU, CSIC), 48940, Leioa, Spain

Fundación Biofisika Bizkaia, 48940, Leioa, Spain.

3 Fundación Biofisika Bizkaia, 48940, Leioa, Spain.

4 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

# **Abstract**

Blood, as the main conduit for metabolites, nutrients, cells, and signaling molecules, offers a uniquely accessible window into human health and disease. Its proteome—a complex mixture derived from plasma and cellular components—reflects both physiological and pathological states. Advances in proteomics, particularly mass spectrometry-based methods and affinity-based platforms like Olink and SomaScan, have dramatically improved our ability to analyse blood proteins with high sensitivity and throughput. However, challenges remain, including the vast dynamic range of protein concentrations and the dominance of high-abundance proteins. To support research in this field, a range of public databases and repositories such as ProteomeXchange, PeptideAtlas, Human Protein Atlas, PaxDB, and GPMDB catalog blood proteomics data, though no centralized repository exists for affinity proteomics. This review explores the cellular and molecular complexity of blood, surveys the latest technological advancements, and assesses the growing ecosystem of public datasets and databases. We highlight the ongoing challenges and transformative potential of blood proteomics in diagnostics, biomarker discovery, and personalized medicine.

# **1. Introduction**

Blood is the primary carrier of metabolites, nutrients, cells, and signalling molecules in the human body. It is one of the most useful and readily accessible biological fluids for assessing human health and disease [1](https://www.zotero.org/google-docs/?BDLOTH). Its proteome—all of the proteins found in blood at any moment—provides a unique insight into both normal physiological and pathological conditions. The complexity of blood arises from its diverse composition: plasma accounts for about 55% of its volume, while cellular elements—including erythrocytes (43%), leukocytes, and platelets (2%)—constitute the rest. In addition, the blood proteome contains various immune cell types, each with unique functions and specific proteomes reflective of their roles in immune response, inflammation, and regulation. Understanding the blood proteome in its entirety—from plasma proteins actively secreted by various tissues to those released due to cellular damage, as well as the specialized proteomes of blood cells—offers crucial insights into both normal physiology and disease mechanisms.

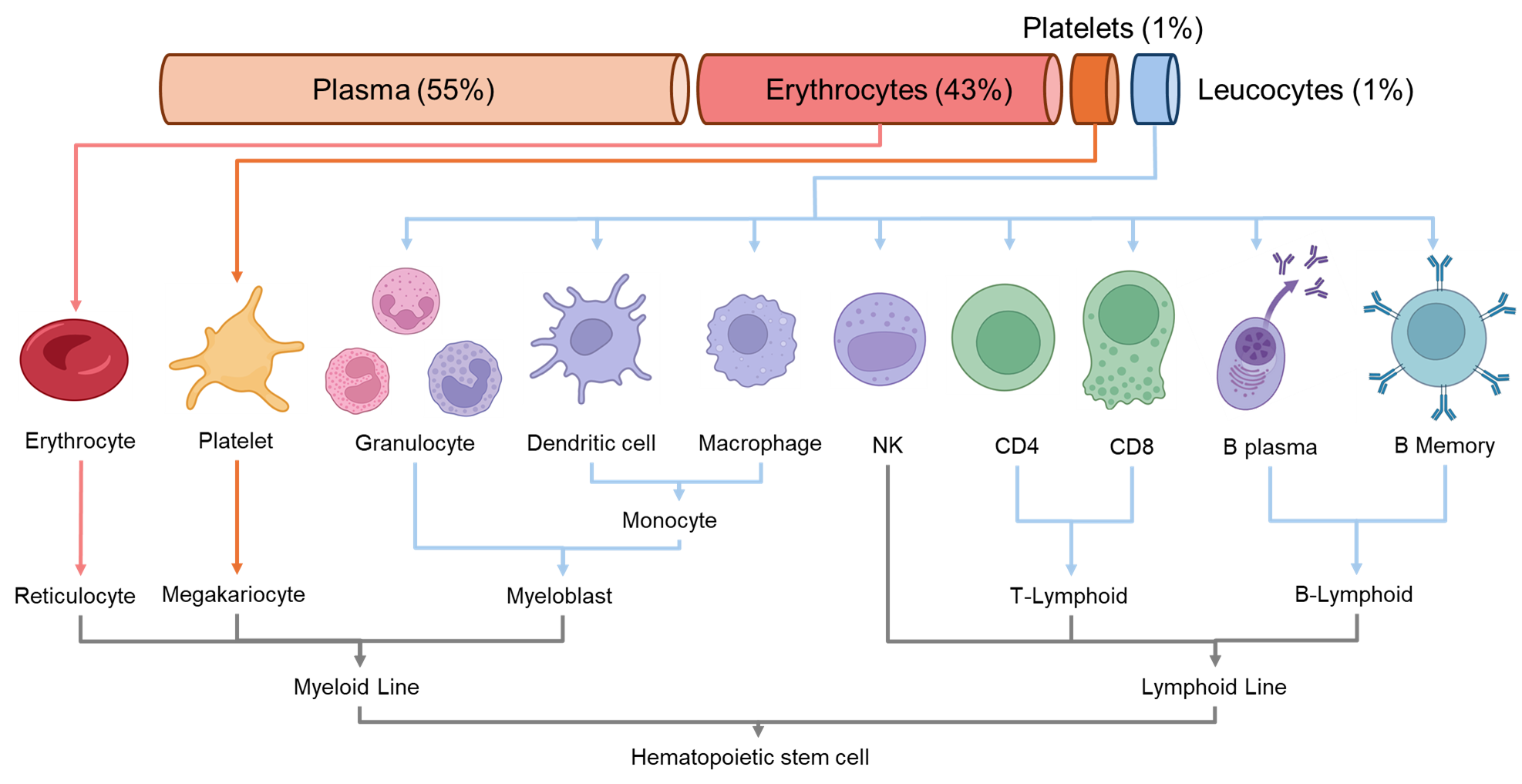
Recent advances in proteomics technologies including mass spectrometry (MS) proteomics, or affinity reagent-based technologies (ARBTs) proteomics methods such as Olink or SomaScan have dramatically expanded our ability to identify, quantify, characterize and catalogue blood proteins with unprecedented depth and precision [2](https://www.zotero.org/google-docs/?F8edSQ). MS proteomics enables the identification and quantification of thousands of proteins across studies with a moderate to large number of samples. The recent introduction of the Thermo Scientific Orbitrap Astral mass spectrometer has further advanced proteomic analysis by combining high-speed acquisition with enhanced sensitivity and dynamic range, enabling the quantification of five times more peptides per unit time than previous state-of-the-art Thermo Scientific Orbitrap instruments [3](https://www.zotero.org/google-docs/?6u6OWw). In contrast, ARBTs like Olink provide quantification of specific protein targets and can scale to much larger cohorts [4](https://www.zotero.org/google-docs/?AUAFHv). Despite these technological leaps, the dynamic range of protein concentrations in blood—spanning 10 to 12 orders of magnitude [5,6](https://www.zotero.org/google-docs/?20PfvP)—continues to pose significant analytical challenges, with highly abundant proteins, such as albumin often masking the detection of low-abundance regulatory proteins [6](https://www.zotero.org/google-docs/?qOOFiQ).

The study of the blood proteome has also been facilitated by the development of multiple repositories and databases that store, integrate and catalogue the public proteomics data, the proteomes of each blood component and the functions of those proteins. The ProteomeXchange (PX - <http://www.proteomexchange.org>) consortium was established in 2011 to streamline MS data dissemination, publication, and replication across various resources and databases, ensuring the long-term sustainability of public proteomics data [7](https://www.zotero.org/google-docs/?MKqsOW). Its members not only archive public proteomics datasets but also systematically reanalyse them to create catalogues of specific blood components such as the PeptideAtlas plasma builds [6](https://www.zotero.org/google-docs/?Bb8Woc) or the quantms plasma proteome [8](https://www.zotero.org/google-docs/?sk6EiQ). Additionally, multiple databases and resources include catalogues of protein identifications and quantifications from blood proteomics datasets; the most relevant ones are the Human Protein Atlas (HPA) [9](https://www.zotero.org/google-docs/?xmvP6Z), the Protein Abundance Across Organisms Database (PaxDB)[10](https://www.zotero.org/google-docs/?4E7bV0), and the Global Proteome Machine Database (GPMDB)[11](https://www.zotero.org/google-docs/?EZxSCq). These databases are essential resources for researchers, providing access to information on protein expression in blood and its associations with disease states and phenotypes. By May 2025, there are more than 42000 public datasets ProteomeXchange, 20208 coming from human studies. These infrastructures not only enhance data accessibility and reusability but also enable comparative analyses across different studies, techniques, and biological contexts. In contrast, there is no centralized repository dedicated to ARBT proteomics methods. A PubMed search for "SomaScan" yields 423 articles, while the SomaScan platform maintains its own curated repository of publications where its products have been used, comprising 1,262 entries. Similarly, a search for "Olink" in PubMed returns 1,085 articles, and the Olink platform provides information on 2,647 publications, detailing the specific products used in each study.

This review examines the taxonomic composition of blood and highlights the MS-based and ARBTs proteomics approaches that have driven advances in blood proteomics. We assess the current landscape of proteomics databases, the availability of publicly accessible blood proteome datasets, and the key challenges in their integration and analysis. Our aim is to provide a comprehensive overview of the databases, datasets, and technologies shaping blood proteomics, emphasizing its transformative potential in diagnostics, biomarker discovery, and personalized medicine.

# **2. A Taxonomy of blood components**

Blood is a complex biological matrix consisting of plasma and cellular components (**Figure 1**) —erythrocytes, leucocytes, and platelets—that play vital roles in nutrient transport, waste removal, immune defense, and coagulation. Its connection to most organs makes blood an invaluable resource for medical diagnostics and research, offering a comprehensive view of systemic health and disease states [1](https://www.zotero.org/google-docs/?IB1NsD). Blood analysis is minimally invasive, making it a convenient tool for routine health assessments and diagnosis and investigations into biomarkers and therapeutic strategies. Here, we present a taxonomy of the major blood components found in public databases and datasets, detailing their abundance in blood and their functions in the human body, which will be discussed throughout this manuscript.



**Figure 1: Blood taxonomy components**. Blood is composed of erythrocytes (43% of the volume), platelets (1%) and leukocytes (1%), suspended in plasma (55%). Blood cells derive from hematopoietic stem cells, and are divided into two lines: the Myeloid Line, the precursor of erythrocytes, platelets granulocytes, dendritic cells and macrophages and the Lymphoid Line, the precursor of NK, CD4 and CD8 T cells, B plasma and memory cells.

## **2.1. Plasma**

Plasma constitutes ~55% of blood volume, serving as a medium for transporting nutrients, waste, hormones, and proteins [1](https://www.zotero.org/google-docs/?teaIQj). Serum, derived from coagulated plasma (without anticoagulants), lacks fibrinogen, resulting in a slightly lower protein concentration [12](https://www.zotero.org/google-docs/?PFCpVG). Plasma is advantageous for faster preparation and minimizing coagulation-related variability, making it the preferred choice in most studies, despite serum sometimes being favored for specific proteomic analyses due to reduced interference from clotting factors [13](https://www.zotero.org/google-docs/?n3GmlF).

**2.1.1. Plasma proteomics in numbers**

Plasma contains an estimated 6,000 to 10,000 proteins based on the combined findings of multiple studies; however, no single study has reported this many proteins independently [14](https://www.zotero.org/google-docs/?eRg2Ox). The upper bound of this estimate is largely derived from bioinformatic predictions of proteins containing signal peptides or other secretion-associated features, such as those reported by tools like SignalP or SecretomeP [15](https://www.zotero.org/google-docs/?ppWDpO). While these methods cast a wide net—including both classically and non-classically secreted proteins—the actual number of actively secreted and plasma-resident proteins is likely lower, once factors like intracellular contamination, retention signals, and false positives are taken into account.

Despite these caveats, the plasma proteome still represents a substantial, though relatively small, subset of the full human proteome, which comprises approximately 20,000 to 22,000 protein-coding genes. In the context of proteomic analysis, it is important to distinguish between proteins that have merely been reported—i.e., detected in at least one study—and those that have been reliably quantified, meaning their concentrations have been consistently measured across multiple samples and studies. While up to 10,000 unique proteins have been reported in plasma, only around 4,000 to 7,000 have been quantified with confidence [16](https://www.zotero.org/google-docs/?3cJSYw). This discrepancy underscores the persistent challenge of measuring low-abundance proteins within plasma’s vast dynamic range, which spans over 12 orders of magnitude [6](https://www.zotero.org/google-docs/?QplG7H).

In addition to technical limitations, the analysis of the plasma proteome is further complicated by substantial biological variability. Protein composition and concentration in plasma can differ markedly between individuals due to factors such as age, sex, genetic background, lifestyle, and physiological state. Even within the same individual, the plasma proteome is dynamic and can fluctuate throughout the day in response to circadian rhythms, food intake, physical activity, and stress [17](https://www.zotero.org/google-docs/?N0xCra). These sources of variation can obscure true biological signals and complicate the identification of consistent patterns, particularly when studying low-abundance proteins or comparing datasets across cohorts. As a result, careful experimental design, appropriate normalization strategies, and inclusion of well-matched controls are essential for obtaining reproducible and biologically meaningful insights from plasma proteomics data.

Compared to the average human cell, which expresses 6,000–12,000 proteins [18](https://www.zotero.org/google-docs/?Rx3LU1), plasma has a more diverse and dynamic composition due to its role as a transport medium for proteins originating from various tissues. The dynamic range of plasma proteins spans over 10–12 orders of magnitude, with highly abundant proteins like albumin (40g/L, ~50–60% of total plasma protein mass), immunoglobulins (~35–40%), fibrinogen (∼4%) transferrin (∼1%) and lipoproteins (∼1%), overshadowing low-abundance regulatory proteins, including cytokines and growth factors such as Interleukin 1 beta (1.1 ng/L) [19](https://www.zotero.org/google-docs/?pHboSA). Functionally, most plasma proteins are involved in immune response, transport, hemostasis, and intercellular signaling, reflecting its role in systemic homeostasis [20](https://www.zotero.org/google-docs/?0Kn72w).

Given its complex and dynamic composition, the plasma proteome plays a pivotal role in clinical diagnostics. Numerous plasma proteins serve as established biomarkers for a wide range of diseases. For example, cardiac troponins are used to diagnose myocardial infarction [21](https://www.zotero.org/google-docs/?S7HfMO), while creatinine and cystatin C indicate kidney dysfunction [22](https://www.zotero.org/google-docs/?PCVtt7), and cancer-associated markers such as CEA, PSA, and CA-125 assist in cancer detection [23](https://www.zotero.org/google-docs/?NTa3Pr). These examples illustrate how the ability to detect and quantify specific proteins in plasma has become central to modern medicine.

However, not all clinically relevant proteins are measurable under normal physiological conditions. Many proteins associated with disease states are present in healthy individuals at concentrations below the detection limit [24](https://www.zotero.org/google-docs/?mOEBkE). These low-abundance proteins often become detectable only when their expression increases in response to pathological processes. Consequently, a comprehensive understanding of the plasma proteome requires studying both healthy and diseased individuals in order to distinguish between baseline expression and disease-induced changes. This dual approach is critical for identifying true biomarkers and improving diagnostic specificity and sensitivity.

**2.1.2. Plasma subgroups and terminology**

Building on the complexity and diversity of the plasma proteome, it is useful to classify plasma proteins based on their origin within the broader framework of the circulating proteome. The human circulating proteome—sometimes referred to as the plasma proteome—encompasses all proteins present in the bloodstream, including those actively secreted by cells and those released as a result of cellular turnover or damage [25](https://www.zotero.org/google-docs/?e6XmA5). While the term plasma proteome technically refers to the protein content of plasma (the cell-free component of blood after removal of clotting factors), the circulating proteome is often used more broadly to capture proteins in both plasma and serum. Within this landscape, a conceptual and functional division can be made between two major sources of proteins: the secretome and tissue leakage proteins [26](https://www.zotero.org/google-docs/?8W6Rky).

The secretome comprises proteins that are actively secreted by cells through classical (signal peptide-dependent) or non-classical pathways. These proteins play essential roles in cell signaling, immune responses, and maintaining physiological homeostasis. Due to their extracellular availability and disease-specific alterations, secreted proteins are prime candidates for biomarker discovery and therapeutic targeting. According to estimates by Uhlén et al., of the 1,709 genes encoding secreted proteins, 730 are predicted to be secreted into the bloodstream [20](https://www.zotero.org/google-docs/?Pu3Lo2). This subset includes diverse functional classes such as cytokines (up to 154 proteins), hormones (75), and growth factors (33). Interestingly, nearly 100 of these 730 proteins lack any functional annotation in UniProt, highlighting the extent of unexplored biology within the secretome. The origin of these secreted proteins also varies: 139 proteins are primarily secreted by the liver (e.g., albumin, apolipoproteins, and complement factors), 179 proteins originate from blood and lymphoid tissues (e.g., chemokines, granzymes), and 253 proteins are expressed across multiple tissues, reflecting their widespread physiological importance.

In contrast, tissue leakage proteins are not actively secreted but instead enter the circulation passively, often as a result of cellular injury, apoptosis, or necrosis. These proteins serve as indicators of tissue-specific damage and are crucial for real-time assessment of organ health and disease progression. Classic clinical examples include cardiac troponins—released during myocardial infarction [27](https://www.zotero.org/google-docs/?iZIZjT)—and liver enzymes like ALT and AST, which rise in response to hepatocyte injury [28](https://www.zotero.org/google-docs/?seJDYJ). While tissue leakage proteins may complicate the interpretation of proteomic data by increasing background noise, they also offer a valuable window into pathological processes and are increasingly integrated into diagnostic frameworks.

Another interesting subgroup of plasma proteome is the core plasma proteome. The core proteome comprises proteins consistently found in plasma under normal physiological conditions, including albumin, immunoglobulins, and transport proteins. This stable subset establishes a baseline reference, facilitating the detection of pathological changes and improving the accuracy of diagnostic assays. In a recent study, a meta-analysis of the results of targeted quantitative screening of human blood plasma was performed, reporting 53 proteins that were found in at least 70% of the samples and with a coefficient of variation below 40% [29](https://www.zotero.org/google-docs/?rIc6ww).

## **2.2 The Cellular Proteome of Blood**

The cellular proteome includes the proteins from various blood cell types, each with specialized functions like oxygen transport, immune defense, and coagulation. While blood proteomics often focuses on plasma, analysing the cellular proteome is essential for understanding blood cell roles in physiological processes and immune responses. Changes in the cellular proteome can also indicate conditions such as infections and immune disorders.

### 2.2.1. Erythrocytes

Erythrocytes, or red blood cells, make up approximately 43% of blood and are primarily responsible for transporting oxygen via hemoglobin [30](https://www.zotero.org/google-docs/?E09Z8d). As the most abundant cell type in the bloodstream, they play a critical role in various physiological and pathological processes. Despite lacking nuclei and most organelles, erythrocytes possess a highly specialized and dynamic proteome that is essential for maintaining cellular integrity and function. Alterations in the erythrocyte proteome have been linked to a range of diseases, including hemoglobinopathies, anemia, cardiovascular disorders, and even systemic conditions like diabetes and Alzheimer's disease [31](https://www.zotero.org/google-docs/?lSNPY9).

Additionally, distinct erythrocyte subtypes provide further insights into disease mechanisms. Reticulocytes, the immature precursors of mature erythrocytes, retain residual RNA and certain organelles, making them valuable for studying erythropoiesis and bone marrow function [32](https://www.zotero.org/google-docs/?phgqw3). Ghost cells, which are erythrocyte membranes devoid of cytoplasmic content, are essential for investigating membrane-associated proteins and their role in cellular signaling, ion transport, and immune response [33](https://www.zotero.org/google-docs/?JrQu4I).

### 2.2.2. Platelets

Platelets (<1% of blood volume) are essential for hemostasis and wound healing. Produced from megakaryocytes, they form clots by adhering to vessel walls and aggregating. Despite lacking nuclei, they retain organelles like mitochondria, enabling a 7–10-day functional lifespan [34](https://www.zotero.org/google-docs/?9R1zQu). Platelet proteomics has proven essential in understanding hemostasis, thrombosis, and immune responses. Proteomic analysis has helped identify biomarkers for thrombotic disorders like immune thrombocytopenic purpura (ITP) and thrombotic thrombocytopenic purpura (TTP). In ITP, changes in cytoskeletal and signaling proteins affect platelet [35](https://www.zotero.org/google-docs/?W7ZCNm), while in TTP, protein alterations related to the ADAMTS13 enzyme aid in diagnosis and treatment [36](https://www.zotero.org/google-docs/?tcV1OK).

### 2.2.3. T cells

Leukocytes (<1% of blood volume) are the body's primary defense against infections. They circulate in the bloodstream and can migrate to tissues to perform their functions. Divided into the lymphoid and myeloid lineages, leukocytes play roles in both innate (immediate, nonspecific) and adaptive (specific, long-lasting) immunity, making them crucial for maintaining health and combating disease [37](https://www.zotero.org/google-docs/?RMrMKR). Leukocytes have a dynamic proteome that shifts in response to environmental cues. In their steady state, they support immune surveillance and homeostasis. When activated by pathogens or inflammation, their proteome undergoes significant changes, including enhanced cytokine production, altered receptor expression, and metabolic adjustments that boost immune responses [38](https://www.zotero.org/google-docs/?KI8Jqi). Studying both states helps identify immune deficiencies and understand conditions like inflammation, autoimmunity, and cancer.

Among leukocytes, T cells are a prominent subtype with specialized functions. CD8+ T cells, known as cytotoxic T cells, are responsible for directly targeting and eliminating infected or malignant cells [39](https://www.zotero.org/google-docs/?YIWhwc). CD4+ T cells, or helper T cells, are further divided into subtypes such as Th1, Th2, Th17, and Treg cells, each playing distinct roles in coordinating immune responses and maintaining tolerance [40](https://www.zotero.org/google-docs/?YwFvbI). The proteomic profiling of T cells has been pivotal in uncovering biomarkers for various diseases. For instance, in multiple sclerosis, proteomic analysis has identified altered expression patterns in CD4+ T cells that correlate with disease severity and progression[41](https://www.zotero.org/google-docs/?dj3wZ0). In cancer immunotherapy, profiling the proteome of CD8+ T cells has revealed key exhaustion markers, enabling the development of strategies to enhance T cell persistence and efficacy in fighting tumors [42](https://www.zotero.org/google-docs/?eo3yOo).

### 2.2.4. B cells

B cells are a crucial component of the adaptive immune system, responsible for antibody production, antigen presentation, and cytokine secretion. Their proteome is highly dynamic, reflecting their diverse roles in immune surveillance and response. B cells exist in several subtypes, including naïve B cells, memory B cells, and plasma cells, each with distinct functions and proteomic profiles [43](https://www.zotero.org/google-docs/?Fxj14F). Proteomic analysis of B cells has provided valuable insights into various pathological conditions, including autoimmune diseases, cancers, and infectious disorders.

B cell proteomics has revealed key molecular changes in autoimmune diseases and cancer. In multiple sclerosis, altered protein expression in memory B cells promotes T cell activation and worsens disease, guiding B cell-targeted therapies [44](https://www.zotero.org/google-docs/?JV6vIA). In cancer, tumor-infiltrating B cells show diverse proteomic profiles that can either support or suppress immune responses, offering insights into prognosis and potential treatment targets [45](https://www.zotero.org/google-docs/?Yi87Q6).

### 2.2.5. NK

NK cells are a vital component of the innate immune system, playing crucial roles in immune surveillance, tumor control, and viral defense. NK cells are traditionally categorized into two main subtypes based on their surface marker expression: CD56 bright and CD56 dim cells, each with distinct functions and proteomic profiles [46](https://www.zotero.org/google-docs/?xChrjt). CD56 bright NK cells are predominantly found in lymphoid tissues and are known for their potent cytokine production, aiding in immune regulation and supporting adaptive immunity. Conversely, CD56 dim NK cells are more cytotoxic and are primarily located in the peripheral blood, where they are responsible for directly targeting infected or malignant cells. Proteomic analysis has revealed unique protein expression patterns between these subtypes, shedding light on their specialized functions in immunity.

NK cell proteomics has uncovered expression changes linked to immune dysregulation in autoimmune diseases and cancer. In rheumatoid arthritis, CD56 bright NK cells show altered cytokine pathways that drive inflammation [47](https://www.zotero.org/google-docs/?gbOrxk). In cancer, tumor-infiltrating CD56⁺ NK cells display signs of exhaustion and dysfunction [48](https://www.zotero.org/google-docs/?Co8YDZ). These findings have guided immunotherapy strategies aimed at boosting NK cell function through cytokine stimulation and checkpoint blockade.

### 2.2.6. Granulocytes

Granulocytes are a vital component of the innate immune system, contributing to pathogen defense, inflammation, and tissue repair. They comprise three major subtypes: neutrophils, eosinophils, and basophils, each exhibiting distinct proteomic profiles that reflect their specialized immune functions. Neutrophils are the most abundant granulocytes and are key players in acute inflammation and microbial defense [49](https://www.zotero.org/google-docs/?fy6yjl). Proteomic analysis has revealed key markers of neutrophil activation, such as myeloperoxidase, neutrophil elastase, and S100 proteins, which are elevated in inflammatory disorders [50](https://www.zotero.org/google-docs/?Lv9qFa). Eosinophils, known for their role in allergic responses and parasitic infections, have proteomic signatures enriched in granule proteins like major basic protein (MBP) and eosinophil peroxidase (EPO). Basophils, although less abundant, are crucial in allergic inflammation and histamine release. Proteomic analysis has found unique markers such as carboxypeptidase A3 and basogranulin linked to basophil hyperreactivity in conditions like chronic urticaria asthma [51](https://www.zotero.org/google-docs/?fxTGxj).

### 2.2.7. Monocytes

Monocytes are an essential component of the innate immune system, acting as precursors to dendritic cells and macrophages. Monocytes can differentiate into two major subtypes: classical monocytes (CD14^++ CD16^−) and non-classical monocytes (CD14^+ CD16^+), each with distinct functions and proteomic profiles [52](https://www.zotero.org/google-docs/?7Gt19D). In autoimmune diseases, monocyte proteomics identifies molecular changes tied to disease progression. For instance, in rheumatoid arthritis (RA), proteomic studies show altered protein expression related to inflammation, including those linked to phagocytosis and reactive oxygen species production in monocytes and macrophages [53](https://www.zotero.org/google-docs/?7Wiyxa). These changes contribute to the chronic inflammation and tissue damage observed in RA.

In the context of cancer, proteomic profiling of monocytes and their differentiated counterparts, dendritic cells and macrophages, has provided insights into their roles in the tumor microenvironment. Tumor-associated macrophages, which arise from monocytes, exhibit distinct proteomic signatures that reflect both immune-suppressive and immune-supportive roles, depending on the tumor type and stage [54](https://www.zotero.org/google-docs/?Yifnk9). Moreover, dendritic cell proteomics has revealed markers of antigen presentation and T cell activation, crucial for the development of cancer immunotherapies [55](https://www.zotero.org/google-docs/?eNIYN9). These proteomic alterations offer potential prognostic markers and therapeutic targets for modulating immune responses in cancer.

# **3. The Dynamic Complexity of the Proteome: Insights from Proteomics**

The proteome encompasses the entire set of proteins expressed by an organism, tissue, cell, or biological fluid at a specific time and under specific conditions. While the genome provides a relatively stable blueprint, the proteome is highly dynamic, reflecting changes in developmental stage, environment, disease state, and physiological conditions [56](https://www.zotero.org/google-docs/?GgErNG). Proteomics, the study of the proteome, delves into cataloguing and analysing proteins, their functions, interactions, and roles within biological pathways, offering insights into normal physiological processes and disease mechanisms [57](https://www.zotero.org/google-docs/?gWNKxd).

The human proteome, despite being derived from around 20,000 protein-coding genes [58](https://www.zotero.org/google-docs/?6bE2Pn), is far more complex due to factors such as alternative splicing, post-translational modifications (PTMs), and protein complex formation [59](https://www.zotero.org/google-docs/?CZ3j2e). This complexity results in millions of proteoforms variants, which act as functional indicators of biological states. Advances in proteomics technologies, particularly MS, have driven the field forward, enabling the analysis of complex protein mixtures and the quantification of protein abundance [60](https://www.zotero.org/google-docs/?VeR62L). Specialized areas within proteomics—such as structural, functional, expression, and interaction proteomics—explore various aspects of protein behavior, including 3D structure determination, biological roles, expression profiling, and interaction networks [61](https://www.zotero.org/google-docs/?JJJs1b).

Quantitative proteomics specifically investigates the identification and quantification of proteins to track changes under different physiological or pathological conditions. Unlike genomic studies, which examine gene expression, proteomics analyses functional molecules directly involved in cellular processes, reflecting more immediate biological responses. However, challenges such as the wide dynamic range of protein concentrations demands sophisticated analytical techniques [6](https://www.zotero.org/google-docs/?kqSLHi). These methods help address issues of depth, accuracy, and sensitivity in protein expression studies, making proteomics a vital tool for understanding diseases and advancing precision medicine.

## **3.1. MS-based proteomics and the study of blood**

MS is a powerful analytical technique used extensively in proteomics for its ability to identify and quantify proteins with high precision. MS operates by ionizing protein fragments (typically peptides derived from enzymatic digestion) and analysing their mass-to-charge ratios [62](https://www.zotero.org/google-docs/?XIeoSX). Modern MS systems, often paired with liquid chromatography (LC-MS) [63](https://www.zotero.org/google-docs/?PXdjWT), allow for high-resolution separation and identification of up to 7000 proteins in complex biological samples like plasma [16](https://www.zotero.org/google-docs/?9ZPzH1) or up to 10000 in some cell types. Key advancements such as data-dependent acquisition (DDA) and data-independent acquisition (DIA) have enhanced its sensitivity and throughput, making it a cornerstone of discovery-based proteomics [64](https://www.zotero.org/google-docs/?HVlFBR).

A significant strength of MS lies in its versatility. It is the gold standard for identifying PTMs due to its ability to analyse peptide-specific modifications like phosphorylation, acetylation, or glycosylation [65](https://www.zotero.org/google-docs/?BpzJsu). MS is also well-suited for identifying novel proteins, offering unparalleled discovery potential when analysing complex samples [66](https://www.zotero.org/google-docs/?gNFfuB). Selected Reaction Monitoring (SRM) and Multiple Reaction Monitoring (MRM), which are targeted MS approaches, further enhance the technique’s utility by enabling the precise and sensitive quantification of specific proteins or peptides. By using predefined precursor and fragment ions, SRM and MRM ensure high specificity and reproducibility, making them indispensable for quantifying low-abundance proteins and validating biomarkers, especially in clinical and high-throughput applications [67](https://www.zotero.org/google-docs/?hrh4N9).

To address the inherent complexity and vast dynamic range of the plasma proteome, several pre-analytical strategies have been developed to enhance MS performance. One widely used approach is depletion of high-abundance proteins, such as albumin, immunoglobulins, and fibrinogen, which together account for over 90% of total plasma protein mass [68](https://www.zotero.org/google-docs/?dsuCMG). Removing these dominant proteins reduces sample complexity and facilitates the detection of low-abundance species. Alternatively, fractionation techniques, including strong cation exchange, reverse-phase liquid chromatography, or isoelectric focusing, are employed to separate peptides or proteins based on specific physicochemical properties, allowing more focused MS analysis of sample subsets [69](https://www.zotero.org/google-docs/?zAefYh). Additionally, enrichment strategies—such as immunoaffinity purification or lectin-based capture—are used to selectively isolate PTM peptides or rare proteins, increasing sensitivity for specific targets [70](https://www.zotero.org/google-docs/?6aqRH1). These pre-processing steps, combined with advances in instrument resolution and acquisition modes, are critical for mitigating the signal suppression effects caused by highly abundant proteins and for improving the overall depth and reproducibility of plasma proteomics. An emerging and complementary enrichment strategy involves the use of engineered nanoparticles, which exploit their tunable surface chemistries to form protein coronas that selectively adsorb subsets of plasma proteins [71](https://www.zotero.org/google-docs/?OuvbWA). These nanoparticles can capture a wide range of proteins by utilising differences in size, charge, hydrophobicity, and binding affinities, enriching low-abundance species without immunodepletion or extensive fractionation. A notable example is Seer Inc. 's Proteograph platform, which uses chemically distinct nanoparticles to interact with thousands of proteins simultaneously [72](https://www.zotero.org/google-docs/?ptRzHI). Each nanoparticle type forms a unique protein corona, enabling unbiased and multiplexed sampling across a wide dynamic range. This strategy has demonstrated improved proteomic depth, reproducibility, and scalability, making it particularly valuable for biomarker discovery and clinical applications.

Beyond plasma analysis, MS is also widely applied to profile the proteome of specific cell types, including immune cells, epithelial cells, and stem cells. To achieve this, cells are typically isolated using methods such as fluorescence-activated cell sorting FACS [73](https://www.zotero.org/google-docs/?Moc7T0) or density gradient centrifugation [74](https://www.zotero.org/google-docs/?Dbi9BY) to ensure high purity. Single-cell proteomics is an emerging application, where innovations in nanoLC-MS and sample processing miniaturization now allow proteomic profiling at the single-cell level [75](https://www.zotero.org/google-docs/?mew5jv). These advances have enabled researchers to study dynamic proteome changes during cell differentiation, activation, and in response to environmental or pathological stimuli, offering a deeper understanding of cell-specific functions within complex tissues. While highly specific, MS remains expensive and requires skilled operators and computational resources for data analysis. Additionally, its moderate throughput and intricate workflows may limit its routine use for large-scale or clinical applications (**Table 1**).

## **3.2. Affinity Reagent-based Technique Proteomic**

ARBTs are indispensable tools in modern proteomics, offering highly specific and sensitive detection of proteins through the use of affinity reagents such as antibodies or aptamers. Methods including immunoassays, protein microarrays, and Oligonucleotide-Assisted Affinity Techniques (OAATs) exemplify the versatility of these approaches, especially for targeted proteomic analysis of blood samples. Among ARBTs, OAATs have emerged as powerful platforms for multiplexed protein quantification [76](https://www.zotero.org/google-docs/?gTDnrV). These techniques combine the molecular recognition capacity of affinity reagents with nucleic acid-based detection, enabling high sensitivity and scalability. Notable examples include the Proximity Extension Assay (PEA) and SOMAmer-based assays. PEA (championed by Olink Proteomics) employs oligonucleotide-conjugated antibody pairs that generate amplifiable DNA sequences upon dual binding to target proteins, enabling the quantification of thousands of plasma proteins, depending on the panel design [76](https://www.zotero.org/google-docs/?cd4XDy). SOMAmer assays, developed by SomaLogic, use chemically modified DNA aptamers in place of antibodies to achieve similar sensitivity and throughput [77](https://www.zotero.org/google-docs/?IIzXVl). Advancements in this space continue with technologies like NULISA (Nucleic acid Linked Immuno-Sandwich Assay), which combine in-solution antibody binding with oligonucleotide and biotin tagging to improve specificity and signal detection [78](https://www.zotero.org/google-docs/?C7vNm3).

Traditional immunoassays, such as ELISA and multiplex bead-based platforms like Luminex’s xMAP [79](https://www.zotero.org/google-docs/?qLrrnI), remain foundational in clinical and research settings. These assays detect proteins via antigen–antibody binding, with readouts based on enzymatic, fluorescent, or chemiluminescent signals. Though historically low-throughput, modern immunoassays can now quantify dozens of proteins simultaneously, making them well-suited for studying low-abundance targets in blood [79](https://www.zotero.org/google-docs/?ODPXiu).

Protein microarrays further expand throughput by immobilizing thousands of capture reagents—typically antibodies or aptamers—on solid surfaces for large-scale screening applications such as biomarker discovery and immune profiling [80](https://www.zotero.org/google-docs/?zRkIj8). While highly scalable, they may suffer from lower sensitivity than proximity-based methods like PEA and require amplification strategies for reliable detection. Their specificity also hinges on the quality and validation of the immobilized reagents.

To address the challenges posed by the high dynamic range and complexity of the plasma proteome, ARBTs leverage dual-recognition mechanisms and signal amplification via nucleic acid tags, allowing for femtomolar-level detection in minimally processed samples. This makes them especially advantageous for high-throughput and clinical applications, where consistency, ease of use, and sensitivity are paramount.

Nonetheless, the success of ARBTs relies heavily on the quality of the affinity reagents. Assays such as PEA and SOMAmer depend on predefined antibodies or aptamers, which may suffer from off-target binding or insufficient specificity, particularly in highly multiplexed formats [81](https://www.zotero.org/google-docs/?pjqZm5). As noted in the literature, each capture reagent must be individually validated to assess its analytical specificity and resistance to cross-reactivity, particularly in complex biological matrices where proteoforms and PTMs may influence binding [77](https://www.zotero.org/google-docs/?npSrk7). Without thorough validation, the accuracy of protein quantification remains uncertain.

Another key limitation is that most ARBTs provide relative rather than absolute quantification due to the absence of standardized calibration curves or universal internal protein concentration references. Despite these constraints, the field benefits from an extensive and growing catalog of affinity reagents, estimated at over five million, covering roughly 95% of protein-coding genes, as indexed by platforms like Antibodypedia (<https://www.antibodypedia.com>).

While ARBTs are highly effective in plasma and other biofluids, their application to cell proteomics is more limited. This is primarily due to the need for cell lysis and protein extraction, which can disrupt protein conformations and hinder epitope accessibility, reducing binding efficiency [82](https://www.zotero.org/google-docs/?XvTKQq). Additionally, the diversity of intracellular environments, along with the vast array of proteoforms and PTM, complicates antibody or aptamer design and validation. It is worth mentioning that SomaLogic also provides a distinct technology called CyTOF (Cytometry by Time-Of-Flight), a mass cytometry platform used to depict the heterogeneity of immune single cells in different contexts [83](https://www.zotero.org/google-docs/?vxZsqK). This technique relies on antibody-metal isotope conjugates, with protein presence measured based on metal signals detected by the MS. However, its major limitation is that it can only analyse 40-50 proteins per cell, as it is dependent on pre-validated antibodies and limited by the number of available metal tags.

In summary, ARBTs offer powerful, sensitive, and targeted approaches to blood proteomics. Each platform brings unique strengths: OAATs provide femtomolar sensitivity and robust multiplexing; immunoassays are cost-effective and clinically established; protein microarrays offer scalability for discovery. However, all share the common challenge of reagent dependency, which limits exploratory and PTM-centric studies. When appropriately validated and applied, ARBTs serve as a critical complement to MS and other proteomic technologies.

## **3.3. Comparative Overview of MS and ARBT in Plasma Proteomics**

Comparing different plasma proteomics techniques has become increasingly necessary and timely, as methodological differences can lead to divergent biological insights and impact biomarker discovery. In one of the largest proteomic comparisons performed thus far, Eldjarn et al [24](https://www.zotero.org/google-docs/?gL1j7P) systematically compared Olink and SomaScan platforms across thousands of plasma samples. Despite both platforms targeting over a thousand proteins, the overlap between their measurements was limited: only about 36% of proteins showed strong correlation between the two, similar results obtained in other studies [81,84,85](https://www.zotero.org/google-docs/?xS7mGh). Genetic analyses further revealed that around 60% of the genetic influences on protein expression were platform-specific, indicating that the two technologies often capture different aspects of the proteome. This highlights the limited interchangeability between platforms and underscores the need for cross-validation when interpreting biomarker findings.

In a very recent article, Beimers et al compared a benchmark of six MS-based platforms (including nanoparticle enrichment approaches like Seer) and Olink revealed substantial differences in protein coverage, reproducibility, and biological relevance. Depending on the MS approach, between 600 and 4,500 proteins were detected per platform, with a cumulative total of 6,393 unique protein groups identified across all techniques. However, overlap between platforms was modest: only ~25% of the proteins measured by MS and Olink overlapped, and each method enriched for different Gene Ontology categories. Technical reproducibility ranged from 8.7% to 26%, generally improving with protein abundance. While Olink typically detected more proteins than most MS platforms, the Seer platform stood out, identifying ~7,000 proteins with remarkable completeness—6,604 of them were consistently found in all samples. Strikingly, in differential expression analyses comparing cancer and healthy samples, only a single biomarker was shared across platforms, highlighting the divergent biological signals captured by each. Additionally, niche assessments—such as mixing human and chicken plasma—demonstrated MS superior species discrimination, where Olink failed. The study concludes that choosing a proteomics platform should be driven not only by protein depth but also by the biological question, as platform performance varies drastically depending on target pathways, detection thresholds, and analytical reproducibility. Overall, the general consensus is that each approach measures different properties and also has different strengths and weaknesses [86](https://www.zotero.org/google-docs/?W5R1wE)(**Table 1**).

## **Table 1: Comparison of Proteomic Techniques in Blood Analysis**

| **Criteria** | **Mass Spectrometry** | **PEA/SOMA Logic** | **Immunoassay** | **Protein Microarray** |
| --- | --- | --- | --- | --- |
| **Protein Discovery (New Proteins)** | **Best choice**  Unbiased, no prior knowledge required | **Not suitable**  Relies on known antibodies | **Not suitable** Detects known targets | **Not suitable** Requires known probes |
| **Differential or relative quantification** | **Absolute and relative** | **Relative** | **Relative** | **Relative** |
| **Post-Translational Modifications** | **Excellent**  Can detect a wide range of PTMs | **Limited**  Mostly focused on specific interactions | **Limited**  Depends on antibody specificity | **Moderate**  Can detect some PTMs if included |
| **Report of Very Low-Concentration Proteins** | **High Sensitivity**  Good for low-abundance proteins | **Very High Sensitivity**  If the target is known | **High Sensitivity**  If the target is known | **Moderate Sensitivity**  Not ideal for very low abundance |
| **Detection Range** | **Very Broad ~7–8** orders of magnitude | **Narrower ~3–4** orders of magnitude | **Narrower ~3–4** orders of magnitude | **Moderate ~3** orders of magnitude |
| **High Throughput Analysis** | **Moderate** Requires complex sample prep | **High**  Relatively high throughput | **High**  Especially for multiplexing | **Very High**  It can analyse many proteins at once |
| **Specificity (Detecting Known Targets)** | **High**  With proper database matching | **Moderate**  Antibody-based detection | **High**  Antibody-based specificity | **Moderate**  Limited by probe specificity |
| **Cost** | **High**  Expensive equipment and reagents | **Moderate**  Cost of reagents | **Low**  Affordable for basic assays | **Moderate to High**  Depends on array complexity |
| **Ease of Use** | **Low**  Complex and requires expertise | **Moderate** Relatively easy with kits | **High**  Simple protocols and equipment | **Moderate**  Array preparation and data analysis |

# **4. The Blood proteome in datasets and databases**

A critical component of human proteome research is the availability of public data—spanning datasets, databases, and resources—that provide access to raw mass spectra, peptide identifications, PTMs, and protein expression profiles. The blood proteome is no exception [87](https://www.zotero.org/google-docs/?QeqdN3). To ensure that these datasets are accessible, reusable, and properly curated, proteomics research depends on two key types of resources: data repositories and databases [87](https://www.zotero.org/google-docs/?MGPJnn).

Proteomics data resources can be broadly categorized into: (i) data repositories and archives, which store raw and processed experimental results to support transparency and reproducibility; (ii) protein expression databases, which provide standardized, analysed proteomics data including protein identifications, quantifications; PTMs and variants; and (iii) knowledgebase resources, which integrate proteomics data with broader biological information such as protein function, structure, and genomic context. The typical data flow among these resources begins with researchers depositing their datasets—including raw data, metadata, and experimental descriptions—into public archives (ProteomeXchange [7](https://www.zotero.org/google-docs/?avGCmw), PRIDE [88](https://www.zotero.org/google-docs/?uLaCbR) or MassIVE [89](https://www.zotero.org/google-docs/?s24iUY)) . Protein expression databases then reanalyse and standardize these data using uniform workflows, making expression information accessible via APIs, downloadable files, or web interfaces (PeptideAtlas [90](https://www.zotero.org/google-docs/?qnzPBQ) or HPA [91](https://www.zotero.org/google-docs/?j4bFaF)). Finally, broader knowledge base resources—such as UniProt, Ensembl, and Reactome—integrate this information into a comprehensive biological context, combining it with additional data sources to annotate proteins and genes with details on function, localization, and structure.

## **4.1 ProteomeXchange**

To standardize the sharing of MS-based proteomics data, the ProteomeXchange Consortium (<https://www.proteomexchange.org/>) was established as a global framework for the dissemination of MS datasets [7](https://www.zotero.org/google-docs/?vlxIXL). ProteomeXchange acts as a centralized hub that integrates submissions from multiple partner repositories, ensuring data consistency and accessibility across different platforms. Researchers submitting their datasets to ProteomeXchange receive a unique identifier (PXD accession, e.g., PXD000001), making it easier to track and reference datasets in publications. By 2025, ProteomeXchange is integrated by 6 major partners: the Proteomics Identification Database (PRIDE - <https://www.ebi.ac.uk/pride>**)** hosted at the European Bioinformatics Institute (EMBL-EBI), UK[88](https://www.zotero.org/google-docs/?s1rTfu); the Mass Spectrometry Interactive Virtual Environment (MassIVE - <https://massive.ucsd.edu/>) hosted at the Center for Computational Mass Spectrometry, University of California, USA[92](https://www.zotero.org/google-docs/?m19p6z); the Japan Proteome Standard Repository/Database(jPOST - <https://globe.jpostdb.org>**)** hosted at National Bioscience Database Center, Japan [93](https://www.zotero.org/google-docs/?Cjwbty); the Integrated Proteome Resources **(**iProX - <https://www.iprox.cn/>**)** hosted at Beijing Proteome Research Center, China [94](https://www.zotero.org/google-docs/?dWj1ol); Panorama Public (<https://panoramaweb.org/>) hosted at Department of Genome Sciences at the University of Washington, USA[95](https://www.zotero.org/google-docs/?jRoher); PeptideAtlas and PASSEL hosted at the Institute for System Biology, USA[90](https://www.zotero.org/google-docs/?QTNVS8). By 2025, the ProteomeXchange consortium stores more than 40000 MS-based datasets, including 18000 human studies. Nearly 6000 of the available studies were uploaded in 2024, showing an increasing trend, and the top 3 instruments used to collect the data were Q Exactive (4269), Orbitrap Fusion Lumos (2602) and Q Exactive HF (2260).

As of now, 826 human plasma datasets have been uploaded to ProteomeXchange, with up to 75% originating from the PRIDE repository (**Table 2**). Plasma is one of the most studied tissues, with numerous datasets reflecting strong research interest and supporting reliable proteomic analyses. In contrast, cellular components of plasma, such as platelets (125 datasets) and NK cells (19 datasets), are less represented. The limited number and diversity of cell-type studies hinder efforts to build comprehensive proteomes for individual cell types. Multiple reviews describe how these general archives work and how to submit data to ProteomeXchange partners [87,96](https://www.zotero.org/google-docs/?JmF5Ct).

**Table 2: Distribution of the dataset by cell types, blood components in ProteomeXchange, and corresponding partners**. The absence of a dataset does not indicate that the archive lacks datasets of that type (-); rather, it suggests that they cannot be easily filtered or searched.

| **Database** | **Plasma** | **Erythrocytes** | **Platelet** | **Monocytes** | **Dendritic cells** | **NK** | **Neutrophil** | **T cell** | **B cell** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ProteomeXchange** | 826 | 57 | 125 | 82 | 32 | 19 | 96 | 26 | 96 |
| **PRIDE** | 624 | 51 | 103 | 76 | 29 | 18 | 84 | 22 | 83 |
| **MassIVE** | 41 | 4 | 5 | 2 | 2 | 1 | 3 | 1 | 5 |
| **Panorama Web** | 19 | 1 | - | - | - | - | - | - | - |
| **iProX** | 92 | 1 | 14 | - | 2 | - | 8 | 3 | 7 |
| **jPOST** | 30 | - | 1 | 1 | - | - | 1 | - | 2 |

**4.2. Protein expression databases for blood proteomes**

Data repositories store raw and processed proteomics data, while curated databases organize, annotate, and integrate this information. By applying quality control and bioinformatics processing, they enhance usability and enable comparative analyses and hypothesis generation across biological contexts. In the context of blood proteomics, databases such as PeptideAtlas [90](https://www.zotero.org/google-docs/?3Pe6nt), HPA [9](https://www.zotero.org/google-docs/?eC33Ok), and PaxDb [10](https://www.zotero.org/google-docs/?iq0Pan) provide valuable insights into protein abundance, tissue-specific expression, and functional characteristics, supporting both fundamental research and clinical applications. The following sections highlight key databases that contribute to the study of the blood proteome.

### 4.2.1. PeptideAtlas

PeptideAtlas (<https://peptideatlas.org/>) was initiated in 2004 as part of the Human Proteome Organization (HUPO) Proteomics Standards Initiative [90](https://www.zotero.org/google-docs/?jXTG7C). It was developed to provide a resource for mapping the human proteome based on high-quality peptide identifications. The database compiles data from publicly available MS experiments, submitted by researchers worldwide or retrieved from repositories such as PRIDE [88](https://www.zotero.org/google-docs/?ROGyDq) and MassIVE [92](https://www.zotero.org/google-docs/?IbL4Cn) via ProteomeXchange[7](https://www.zotero.org/google-docs/?47stAP). While PeptideAtlas focuses on providing evidence for all detectable peptides across various proteomes, it also offers information on reported PTMs and the distribution of peptide counts or observations for each reported protein across different tissues and conditions.

As of recent updates (2023-2024) [6,58](https://www.zotero.org/google-docs/?V9HR6n), PeptideAtlas includes 113 datasets and 4608 canonical proteins quantified in plasma and serum, excluding isoforms and immunoglobulins. Canonical proteins are considered when at least two non-nested, uniquely mapping peptides larger than 9 amino acids are identified that together extend at least 18 amino acids, as described in the HUPO Human Proteome Project (HPP) MS data interpretation guidelines [97](https://www.zotero.org/google-docs/?4nxmUS). The average number of proteins identified in each dataset is 318, although there are only 12 datasets above 800 proteins and only 4 above 1000 proteins (PXD027573, PXD030476, PXD023650, PXD007884). However, PeptideAtlas does not specifically curate erythrocyte or other blood component data except for extracellular vesicles, with 33 datasets and 4985 canonical proteins identified thus far.

### 4.2.2. Human Protein Atlas

HPA (<https://www.proteinatlas.org/>) was launched in 2003 to map all human proteins in tissues and cells using antibody-based methods [9](https://www.zotero.org/google-docs/?ICD4gT). Data in the HPA are primarily generated externally and in-house using immunohistochemistry, RNA sequencing, and antibody-based proteomics, with some supplementation from MS studies. HPA offers data about 8 different resources, The Human Blood Atlas, The Human Cell Lines and The Human Cell Proteomes, among others[98](https://www.zotero.org/google-docs/?Qrgy23). The Human Blood Atlas determines the blood protein levels in both health and disease based on MS and antibodies[20](https://www.zotero.org/google-docs/?KZJOgX). For the health proteome, HPA shows data of three different techniques: proteins detected by immunoassay, MS and PEA. For the immunoassay approach, HPA focuses on the blood secretome, which encompasses 785 proteins, according to their bibliographic studies. Plasma reference concentrations for 453 proteins actively secreted into the blood were collected from published research articles from 1983 to 2021 that used immunoassays to measure the target proteins. The concentration of the selected proteins ranged from 40 mg/ml (albumin) to 4 pg/ml (interleukin 5).

The MS approach is based on the 2023-24 build of PeptideAtlas[58](https://www.zotero.org/google-docs/?BuTw4w). This dataset contains the concentration of 4294 proteins, including secreted and leakage proteins. Since the most abundant proteins are usually depleted in MS assays, albumin is not present in the database, and the most abundant protein (ceruloplasmin) reaches 370 mg/L, while the less concentrated protein is the ring finger protein 213 with 1.6 ng/L. In the case of the PEA approach, it is based on a 2021 longitudinal wellness study where 76 healthy patients were analysed for 2 years[99](https://www.zotero.org/google-docs/?HbLyQE). The database contains inter- and intra- individual data for 1463 proteins, focusing on the coefficient of variation. HPA also offers data about the plasma proteome changes across 59 different diseases and 6,121 patients, including 1162 proteins analysed in a pan-disease study by PEA and 146 proteins by MS.

### 4.2.3. PaxDb

PaxDb (<https://pax-db.org/>) is a comprehensive resource launched in 2012 that aggregates and normalizes protein abundance data across multiple organisms and tissues and resources (PeptideAtlas, PRIDE, iProX, and jPOST), focusing on creating a consensus view on normal/healthy proteomes [10](https://www.zotero.org/google-docs/?8uvmgJ). Unlike databases that focus solely on peptide identification, PaxDb emphasizes quantitative proteomics based on spectral counting, providing absolute protein abundance values derived from diverse experimental datasets, such as MS-based proteomics and literature-curated studies. The most recent release of PaxDb (version 5.0, 2023)[100](https://www.zotero.org/google-docs/?taAsWQ) integrates over 800 datasets across 170 species, with a focus on parts per million (ppm) units to facilitate the direct comparison of protein abundance across samples. For human blood, PaxDb includes data from plasma, serum, and PBMCs, offering a broad view of the circulatory proteome.

Each dataset in PaxDb is accompanied by two key quality metrics: Coverage and the Interaction Consistency Score [10](https://www.zotero.org/google-docs/?djRGyJ). Coverage indicates the proportion of the proteome detected in a dataset, with higher values reflecting greater completeness. The Interaction Consistency Score assesses quality by measuring how well abundance ratios of known interacting proteins match expected patterns, as such proteins typically show similar levels. PaxDb accounts for 7 plasma proteome builds, including different versions of PeptideAtlas, with the most recent one (2021) containing data for 6538 proteins and several independent studies (PXD010899, PXD0004242). PaxDb also integrates all builds in one, with a total number of quantified plasma proteins of 7328. Notably, the 2021 PeptideAtlas build has a higher Interaction Consistency Score than the integrated build. For the serum proteome, PaxDb includes only the 2021 PeptideAtlas build, with up to 873 quantified proteins. PaxDb also contains proteomic data about different blood cells, obtained using single cell techniques, such as monocytes (5237 quantified proteins), b cells (9315), NK cells (8874), CD4 (7685) or CD8 (11016), most of them collected from a single study (PXD000561).

### 4.2.4. GPMDB

The GPMDB (<https://www.thegpm.org>) was launched to facilitate the analysis and interpretation of MS data in proteomics research [11](https://www.zotero.org/google-docs/?7dWpaL). It was developed as an open-access platform to support protein identification and annotation, with a focus on the integration of peptide and protein data from tandem MS (MS/MS) experiments. GPMDB provides several specialized builds, including the complete human proteome categorized by chromosome, as well as detailed datasets on acetylation and phosphorylation sites. Additionally, GPMDB features a build that classifies human proteins by tissue type, covering plasma, serum, erythrocytes, platelets, and more. The total number of proteins quantified in these categories is 5,154 for plasma, 7,945 for serum, 617 for erythrocytes, and 7,281 for platelets.

### 4.2.5. Other MS databases

Immpres (<http://immpres.co.uk/>) is a comprehensive online resource dedicated to immune cell proteomics profiles [101](https://www.zotero.org/google-docs/?z4Dmls). It provides curated data from various immune cell types, primarily focusing on mouse models, providing insights into gene regulation, signaling pathways, and immune cell differentiation. While its primary focus is on mouse data, ImmPRes also includes limited information on human immune cell types, specifically CD4 and CD8 T cells (PXD056475) and neutrophils. Researchers can access detailed expression data, facilitating the discovery of immune-specific biomarkers and functional genomics studies. Immpres is particularly valuable for exploring transcriptional networks within the immune system and supporting immunology-related research. However, they offer a limited amount of information about human immune cell types, focusing only on CD4 and CD8 T cells.

Additionally, some valuable proteomics databases, such as the Plasma Proteome Database (PPD) [102](https://www.zotero.org/google-docs/?M7Ja7y) and the Human Body Fluid Proteome Database (HBFPD) [103](https://www.zotero.org/google-docs/?W2wuNv), are no longer available. The PPD was a comprehensive resource dedicated to cataloging proteins identified in human plasma, providing insights into biomarkers, disease-related proteins, and plasma proteomics research. It included data on protein abundance, modifications, and clinical relevance. The HBFPD, on the other hand, focused on the proteomic profiles of various body fluids, such as cerebrospinal fluid, saliva, and urine, offering a broader perspective on proteins involved in physiological and pathological processes across different fluids. Together, these databases supported the study of protein dynamics and their roles in health and disease.

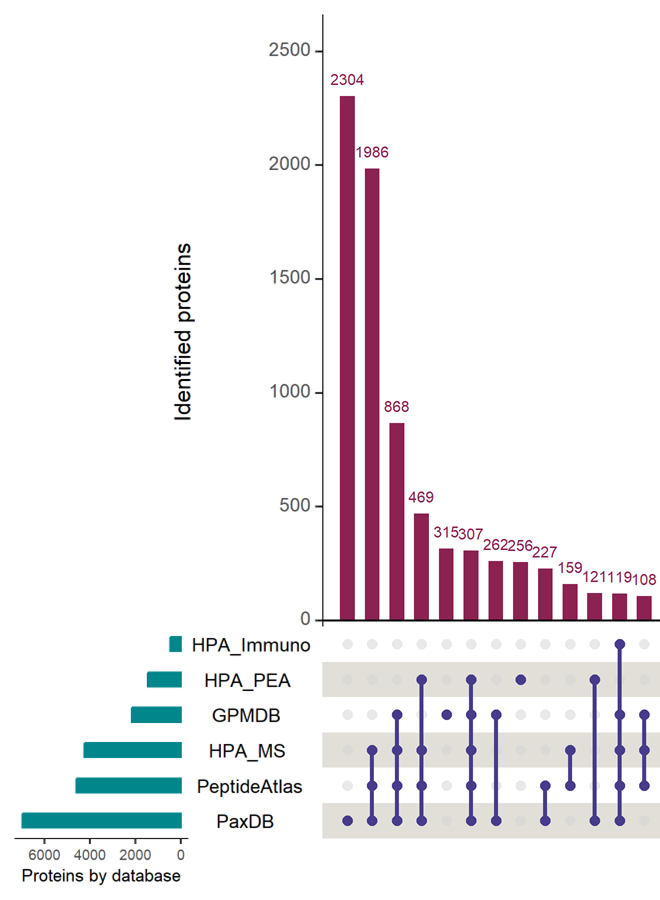
## **4.3. Non-MS repositories and databases**

Despite the diversity of proteomics techniques, the vast majority of publicly available repositories and databases are centred around MS-based proteomics. This is primarily because MS generates large, complex, high-throughput datasets requiring specialised storage and analysis infrastructure. In contrast, antibody-based techniques like PEA and immunoassays typically produce targeted, lower-dimensional datasets, which are often managed internally by research groups or industry partners rather than being deposited in public repositories. Additionally, some antibody-based data sources are proprietary or commercially controlled, limiting their availability in open-access databases.

There are, however, some repositories that do collect ARBT data, such as the EGA (European Genome-phenome Archive - <https://ega-archive.org/>) [104](https://www.zotero.org/google-docs/?cax6lL) or ImmPort (Immunology Database and Analysis Portal - <https://www.immport.org/home>) [105](https://www.zotero.org/google-docs/?wmyvz1). For instance, there are 7 Olink and 6 SOMAscan datasets uploaded in EGA, and 19 Olink and 8 SOMAscan datasets in ImmPort. A search for "SomaScan" on PubMed yields 423 articles, while the SomaScan platform has its curated repository containing 1,262 publications that utilize its products. In comparison, a search for "Olink" on PubMed returns 1,085 articles, and the Olink platform provides details on 2,647 publications, specifying the particular products used in each study.

## **4.4. Plasma protein coverage across: HPA, PeptideAtlas, PaxDB, GPMDB**

Proteomic databases analyse raw data using different techniques and sources, leading to similar but not identical results. As shown in **Figure 2**, the number of identified plasma proteins varies across PeptideAtlas, three HPA resources, PaxDb, and GPMDB. When combined, these sources report 8,038 plasma proteins, exceeding any single study to date but still below the estimated 10,000 in the full plasma proteome [26,102](https://www.zotero.org/google-docs/?sHjy9T). Notably, only 66 proteins are shared among all databases. However, when excluding the HPA Immunoassay database—which contains only 453 proteins—the number of shared proteins among the remaining databases increases to 285.

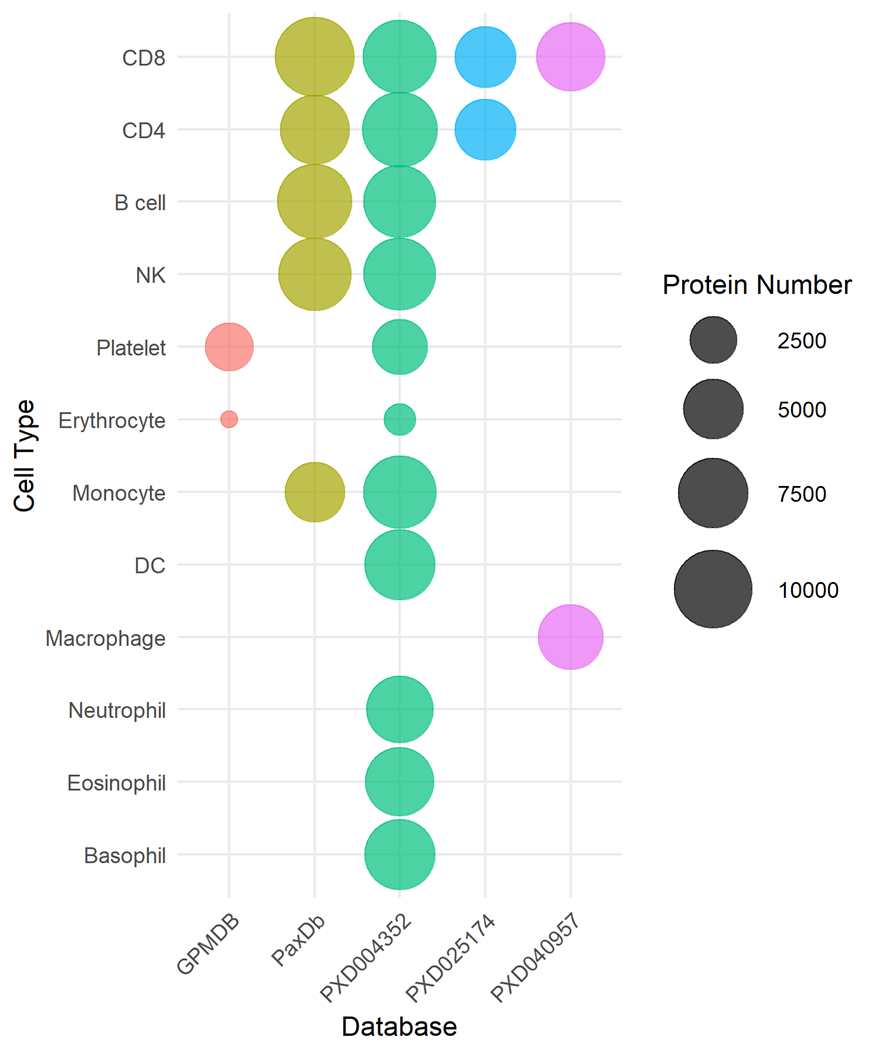
Examining each database individually, PaxDb uniquely identifies 2,304 proteins, GPMDB 315, HPA PEA 256 and PeptideAtlas 59. As a result, the proportion of proteins shared with at least two other databases ranges from 58% in PaxDb to 95% in HPA MS. Given that HPA MS is entirely and PaxDb is partially based on PeptideAtlas, 99% of HPA MS proteins (4,232) are found in PeptideAtlas, and 92% of PeptideAtlas proteins (4,229) are also present in PaxDb. However, since PaxDb integrates data from multiple sources, these 4,229 shared proteins represent only 61% of its total protein content. A specific comparison of the sources used in the PaxDb’s is available in **Supplementary Figure 1**. Regarding the three resources of HPA, only 153 proteins are shared among them. Removing HPA Immuno (it has 45 proteins not shared with the other resources), the number goes up to 948 proteins, 65% of the entire HPA PEA database. 

**Figure 2: Overlaps and differences in the identified plasma proteins in different databases**. Each set in the graph represents a unique combination of identified proteins, while the intersections indicate shared elements among the databases. The smaller bars to the left indicate the number of identified proteins per database. Combinations with less than 100 proteins are not shown.

## **4.5. Cell type database comparison**

Unlike plasma, where proteomic analysis often relies on a combination of techniques, most studies focused on blood cell types primarily use MS to profile their proteomes. The complexity of blood cell types is further compounded by the existence of numerous subtypes within each major cell type, such as dim and bright NK or Th1, Th2 and Th7 helper CD4+ cells. Depending on the study's focus, researchers may either analyse the proteome of a specific subtype or look at the proteome of the entire cell type.

Studies that analyse multiple cell types simultaneously are often considered more reliable, as they ensure that all measurements are performed using the same technique, by the same research team, and under similar conditions. However, such comprehensive studies are relatively rare, and there are very few databases or datasets that include proteomic data for multiple cell types analysed in this way. As a result, comparing datasets or databases across different cell types becomes challenging, as differences in experimental design, methodology, and data processing can introduce significant variability.

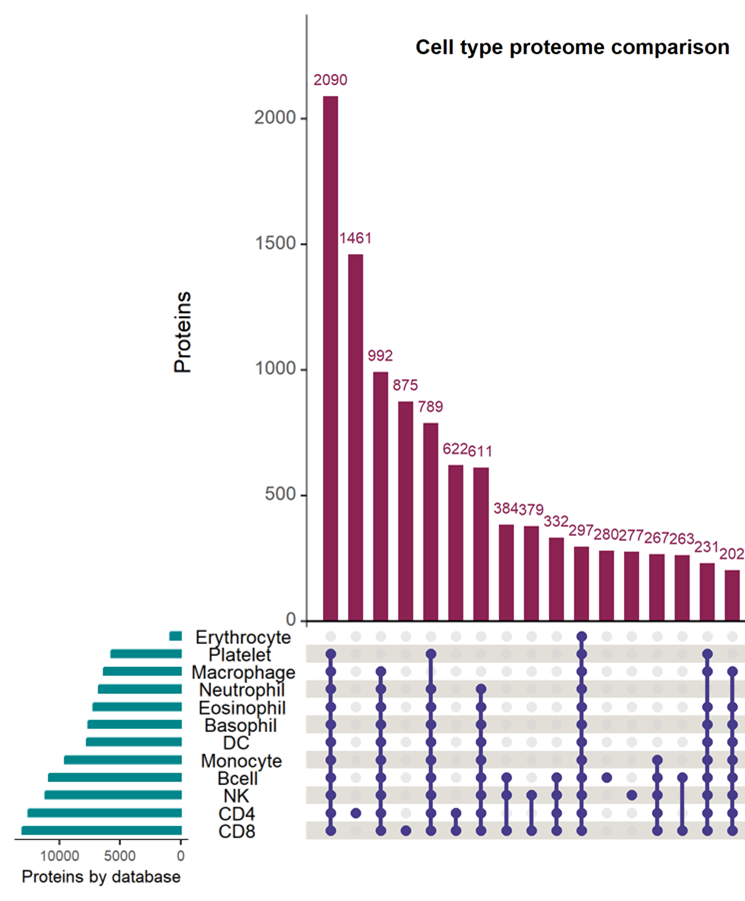
**Figure 3** presents the proteins reported across various cell-type-specific resources, including databases such as GPMDB and PaxDb, as well as relevant original datasets deposited in ProteomeXchange. Notably, resources like PaxDb (based on PXD00056) and the dataset PXD004352 analyse multiple cell types simultaneously, in contrast to the majority of other studies that focus on just one or two cell types. Additionally, PXD004352 investigates several subtypes for each of the cell types depicted in the figure (see **Supplementary Data**). The proteomes of all subtypes have been combined and averaged, allowing for a comparative analysis across the datasets. In terms of the proteomes for each cell type, most studies reported between 5,000 and 10,000 proteins per cell type, except erythrocytes, which yielded only 400-800 proteins. The dataset PXD004352 stands out by reporting a higher number of proteins overall, largely due to its analysis of multiple subtypes. Generally, CD8+, CD4+, B cells, and NK cells exhibit the highest protein counts, while erythrocytes and platelets have the fewest.

### Figure 3: Quantification of immune cell proteomes across resources. Bubble plot illustrating the number of proteins quantified in various immune cell types across multiple proteomic resources.

Beyond the number of proteins reported by each resource, the proteomes were also compared to evaluate overlap in protein reports (**Table 3**; detailed breakdown by resource and cell type in **Supplementary Figure 2**). While individual studies reported between 5,000 and 10,000 proteins per cell type, combining multiple resources raised the total to approximately 6,000–13,000 proteins—though this may inflate the false discovery rate (FDR) if not properly controlled [106](https://www.zotero.org/google-docs/?w7BIXu). Notably, 40–50% of each cell type’s combined proteome was reported in only a single dataset. Strikingly, for platelets and erythrocytes, only 17% and 1% of their combined proteomes, respectively, were shared between GPMDB and PXD004352. In fact, GPMDB reported only 120 proteins in erythrocytes, of which just 4 overlapped with those in PXD004352. Overall, the lack of overlap highlights the variability in results across different studies, reflecting the impact of the techniques and methodologies used.

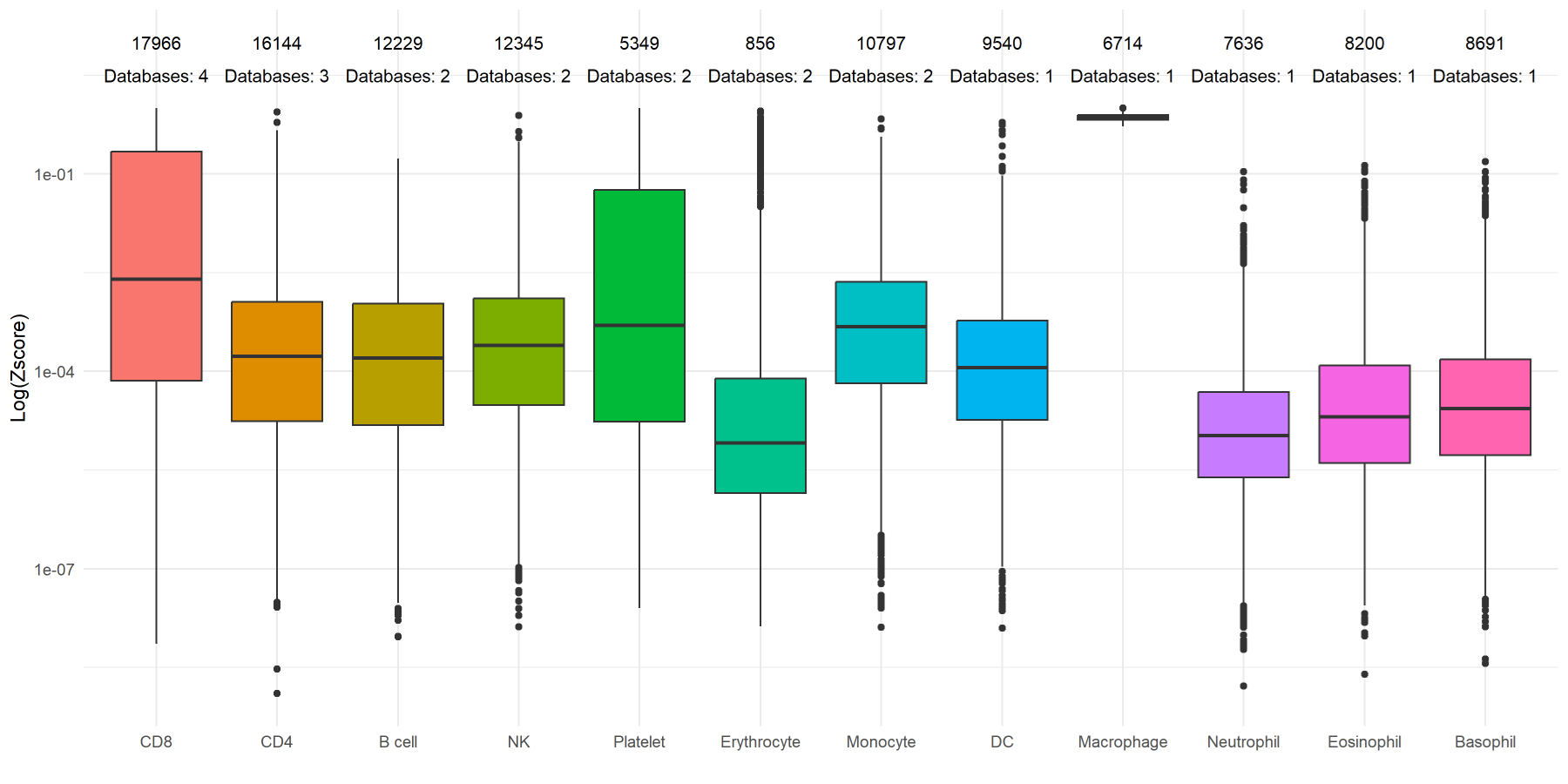
**Table 3: Blood cell types proteome by combining resources**. Combined reported proteins, proteins reported in one and all resources and unique proteins for each cell type are detailed. Reported proteins’ refer to those listed across sources but do not imply confirmed identification; they may include unverified or context-dependent entries.

| **Cell type** | **Combined reported proteins** | **Proteins reported in only one resource** | **Overlapped protein among all resources** |
| --- | --- | --- | --- |
| **CD8+** | 13053 | 4755 (36%) | (28%) |
| **CD4+** | 12573 | 6064 (48%) | 1278 (31%) |
| **B cell** | 10864 | 4587 (42%) | 6277 (58%) |
| **NK** | 11187 | 5729 (52%) | 5458 (48%) |
| **Platelet** | 5713 | 4743 (83%) | 970 (17%) |
| **Erythrocyte** | 846 | 842 (99%) | 4 (1%) |
| **Monocyte** | 9549 | 5731 (60%) | 3818 (40%) |
| **DC** | 7735 | - | - |
| **Macrophage** | 6328 | - | - |
| **Neutrophil** | 6742 | - | - |
| **Eosinophil** | 7222 | - | - |
| **Basophil** | 7631 | - | - |

Comparing the proteomes of different blood cell types offers valuable insights into the common and unique proteins that define each type, shedding light on cellular function and differentiation. **Figure 4** illustrates the overlap and distinctions between blood cell type proteomes. Collectively, the combined proteome of all blood cell types analysed in this study comprises 15717 proteins. Among these proteins, only 297 are shared across all cell types. However, when erythrocytes are excluded—given that their entire proteome is limited to approximately 850 proteins—the number of shared proteins increases significantly to 2,387. Across the 12 studied cell types, approximately 4,500 proteins are each present in at least 10 different cell types. These broadly distributed proteins represent 60–80% of the proteomes of platelets, macrophages, neutrophils, eosinophils, and basophils. In contrast, CD4 and CD8 T cells each have a substantial portion of their proteomes—10% and 7%, respectively—composed of proteins not detected in any other cell type. Notably, these two cell types share 10,874 proteins, representing approximately 85% of their proteomes — a figure that is comparable to, though slightly lower than, those reported in other studies [107](https://www.zotero.org/google-docs/?iYG0So). Besides CD4 and CD8 cells, only NK cells and B cells exhibit more than 100 unique proteins, underscoring the overall similarity among the proteomes of blood cell types.

**Figure 4: Overlaps and differences among blood cell types proteome**. Each set in the graph represents a unique combination of identified proteins, while the intersections indicate shared elements among the databases. The smaller bars to the left indicate the number of identified proteins per database. Combinations with less than 200 proteins are not shown.

In addition to the number of identified proteins and their overlap between cell types, the expression levels of these proteins are also significant (**Figure 5**). While most cell types exhibit similar average protein expression distribution, there are notable exceptions: CD8 cells, platelets, and macrophages. CD8 cells have the largest reported proteome among the analysed cell types, and four resources contribute to their study. This extensive coverage may account for their broad expression distribution. Conversely, platelets possess the second-smallest proteome and are only covered by two datasets, implying that platelet proteome could be composed of proteins spanning several orders of magnitude.

Macrophages stand out as the only cell type not included in the PXD004352 dataset—the largest resource analysed in this study—and are examined by just one resource. This limited representation may explain their distinct average expression level and narrow distribution.

**Figure 5: Boxplot of blood cell type proteome expression levels.** The intensities of each database were normalized and combined for each cell type, and the number of reported proteins and databases for each cell type are also shown.

# **5. The importance of individual public proteomics datasets**

While expression databases like PeptideAtlas, PaxDB or HPA are invaluable, they are built upon individual datasets submitted to ProteomeXchange, each reflecting specific experimental conditions and biological insights. Although the cumulative number of datasets in a repository or the total number of proteins cataloged in a database is often used to assess the breadth of proteomics knowledge, the most reliable insights come from analysing individual datasets in isolation. Variability in experimental conditions—including sample preparation, instrument sensitivity, data processing pipelines, and biological variables—means that direct comparisons between datasets can be challenging. Differences in methodologies can lead to inconsistencies in protein identification and quantification, making it difficult to extract universal conclusions from aggregated data alone. As a result, the highest-confidence proteomics data is often obtained by focusing on well-designed, high-quality individual datasets rather than by averaging results across multiple studies.

Moreover, the number of proteins reported within a single dataset serves as an indicator of the current state of the art in proteomics technologies. Advances in MS sensitivity, antibody specificity, and data analysis pipelines directly impact how many proteins can be detected in a given experiment. By analysing the protein yield of recent high-quality datasets, researchers can estimate the current limits of detection for specific techniques and gain insights into how proteomics methodologies are evolving. **Table 4** highlights the most significant datasets based on the number of reported proteins, with a focus on plasma and specific blood cell types. It provides details on the databases where these datasets are available, if applicable, along with key technical specifications.

**Table 4: Relevant blood plasma and cell type datasets.** Datasets are sorted by technique, tissue or cell type and sample size, in that order.

| **Technique** | **Tissue or**  **cell type** | **Sample size** | **Dataset ID** | **Protein number** | **Instrument** | **Depletion,**  **Fractionation** | **Acquisition method and**  **Labelling** | **Conditions** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MS** | Plasma | 141 | PXD017052 | +2000 | Orbitrap Fusion Lumos Tribrid  Triple TOF 6600+ | MARS-14 column  XBridge  BEH C18 column  Nanoparticles | DDA  DIA | Normal and early non-small-cell lung cancer |
| **MS** | Plasma | 42 | PXD010899 | 2308 | Q-Exactive | Agilent Plasma 14 Multiple Removal System  HiRIEF | DDA  TMT-10 plex | Normal |
| **MS** | Plasma | 16 | - | +4500 | Q Exactive Plus | Seppro IgY14 LC20 column  bRP | DDA  iTRAQ, TMT6-10 | Hypertrophic obstructive cardiomyopathy |
| **MS** | Plasma | 16 | MSV000079033 | +2500 | Q-Exactive | IgY14, Supermix LC10  bRP | DDA  TMT6, TMT10 | Hypertrophic obstructive cardiomyopathy |
| **MS** | Cell type | 175 | PXD004352 | +9500 | Q Exactive HF | - | DDA | Normal |
| **MS** | Cell type | 156 | PXD047742 | 10000 | Q-Exactive Plus  Orbitrap Exploris 240 | StageTip-based high-pH | DDA  DIA | Normal |
| **MS** | Cell type | 48 | PXD040957 | +5000 | Q-Exactive | - | DDA | Hepatocellular carcinoma |
| **MS** | Cell type | 6 | PXD000561 | 5000-10000 | Orbitrap Elite Orbitrap Velos | In-gel digestion bRPLC | DDA | Normal |
| **ARBP** | Plasma | 38380 | - | 2919 | Olink Explore 3072 | - | - | Biobank |
| **ARBP** | Plasma | 35559 | - | +4600 | SomaScan v/4 | - | - | Biobank |
| **ARBP** | Plasma | 3301 | - | 3,622 | SomaScan assay, Olink | - | - | Normal |
| **ARBP** | Plasma | 372 | - | 1463 | Olink Explorer | - | - | Normal and Type 2 Diabetes |

# **6. Conclusions and Challenges in Blood Proteomics Research**

Blood proteomics research presents distinct challenges due to the inherent complexity and dynamic nature of blood components. Plasma, in particular, poses significant difficulties as it contains proteins originating from various tissues, often spanning a vast concentration range. This complexity requires tailored approaches to maximize protein detection, quantify expression levels accurately, and ensure reproducibility across studies. Addressing these challenges is crucial to improving our understanding of blood proteomics and its potential applications in both health and disease.

Plasma proteomics remains a technically challenging field due to the extreme dynamic range and complexity of the plasma proteome, which spans over 10 orders of magnitude in protein abundance. Despite technological advances, many low-abundance proteins remain undetected, and current datasets report fewer than 8,300 unique proteins, well below the estimated total of 10,000, highlighting substantial gaps in proteome coverage. MS remains the primary method for protein discovery, offering unbiased and comprehensive profiling. However, its sensitivity limits the detection of very low-abundance proteins. In contrast ARBT such as PEA and SOMAmer assays, offer higher sensitivity and scalability, identifying a greater number of plasma proteins in high-throughput settings. Nonetheless, ARBT approaches face challenges with off-target effects and limited specificity, necessitating rigorous reagent validation.

A major issue across both platforms is the lack of consistency and reproducibility among datasets. Few proteins are consistently reported across different plasma proteomics resources. The diversity of MS platforms, depletion and fractionation strategies, and labeling techniques further complicates data integration and cross-study comparisons. Notably, the databases analysed in this study—PeptideAtlas, HPA, PaxDb, and GPMDB—yielded divergent results in both protein counts and identities. For ARBT data, the absence of large, centralized repositories hampers efforts to consolidate and compare findings across studies. To overcome these limitations, future research should emphasize integrative strategies that leverage the strengths of both MS and ARBT platforms. MS is well-suited for initial discovery, while ARBT excels in validation and large-scale analysis.

Regarding the blood cell proteome, firstly, there is a notable lack of studies investigating blood cell types, particularly using ARBT techniques, in comparison to other tissues. Compounding this issue is the challenge of data availability and integration. While ProteomeXchange hosts over 800 plasma proteomics datasets, the numbers are significantly lower for specific blood cell types—only 125 for platelets and as few as 19 for NK cells. Secondly, while individual blood cell types can be studied in isolation, research that simultaneously analyses multiple cell types is far more informative. Such comparative studies are especially valuable as they ensure consistency in techniques, conditions, and patient cohorts, allowing for more accurate cross-cell type comparisons.

Another significant concern is the inconsistency observed across different datasets. Our analysis revealed that the overlap between studies examining the same cell type was below 50%, with some datasets sharing as little as 1%–7% of reported proteins. This lack of reproducibility underscores the need for standardized methodologies and improved data integration strategies. Lastly, the inherent complexity of blood cell populations presents further challenges. Each major cell type is composed of several subtypes with distinct functions and proteomic profiles. However, most existing studies fail to analyse these subtypes independently, instead grouping them together and potentially overlooking critical subtype-specific insights. Taken together, these challenges—ranging from limited comparative studies and inconsistent data to insufficient resolution at the subtype level—make it difficult to integrate existing datasets into a cohesive and reliable blood cell proteomic reference.

Considering these limitations, we propose that future studies adopt broader approaches that simultaneously investigate multiple blood cell types and their respective subtypes. Such comprehensive analyses could greatly enhance our understanding of blood cell proteomics and provide new insights into their roles in immunity, inflammation, and disease progression.

**7. Acknowledgements**

We would like to thank Alejandro Brenes, Edinburgh University, for helpful discussions. We want to Y.P.-R. is funded by Wellcome grants (numbers 208391/Z/17/Z, 223745/Z/21/Z) and EMBL core funding. A.L-S was supported by MCIN (Recovery, Transformation and Resilience Plan) and the Basque Government “Biotechnology Complementary Plan Applied to Health” with funding from European Union NextGenerationEU (PRTR-C17.I1; PRTR-C17.I01.P01.S13) (AAAA\_ACG\_AY\_2539/22\_05), EMBO Scientific Exchange Grant (number 10015/2023). C.M is funded by the Grupos Consolidados Gobierno Vasco 2021, grant number 449IT1720‐22 and Proyectos de Generación de Conocimiento from the Ministerio de Ciencia, Innovación y Universidades, under the grant PID2022‐136788OB‐I00

# Bibliography

[1. Benjamin, R. J. & McLaughlin, L. S. Plasma components: properties, differences, and uses. *Transfusion* **52**, (2012).](https://www.zotero.org/google-docs/?7nHXa2)

[2. Palstrøm, N. B., Matthiesen, R., Rasmussen, L. M. & Beck, H. C. Recent Developments in Clinical Plasma Proteomics—Applied to Cardiovascular Research. *Biomedicines* **10**, 162 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[3. Heil, L. R. *et al.* Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J. Proteome Res.* **22**, 3290–3300 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[4. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* **22**, 19–37 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[5. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome. *Molecular & Cellular Proteomics* **1**, 845–867 (2002).](https://www.zotero.org/google-docs/?7nHXa2)

[6. Deutsch, E. W. *et al.* Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **20**, 5241–5263 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[7. Deutsch, E. W. *et al.* The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Research* **51**, D1539–D1548 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[8. Dai, C. *et al.* quantms: a cloud-based pipeline for quantitative proteomics enables the reanalysis of public proteomics data. *Nat Methods* **21**, 1603–1607 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[9. Uhlén, M. *et al.* A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Molecular & Cellular Proteomics* **4**, 1920–1932 (2005).](https://www.zotero.org/google-docs/?7nHXa2)

[10. Wang, M. *et al.* PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life. *Molecular & Cellular Proteomics* **11**, 492–500 (2012).](https://www.zotero.org/google-docs/?7nHXa2)

[11. Craig, R., Cortens, J. P. & Beavis, R. C. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *J. Proteome Res.* **3**, 1234–1242 (2004).](https://www.zotero.org/google-docs/?7nHXa2)

[12. Issaq, H. J., Xiao, Z. & Veenstra, T. D. Serum and Plasma Proteomics. *Chem. Rev.* **107**, 3601–3620 (2007).](https://www.zotero.org/google-docs/?7nHXa2)

[13. Plebani, M. *et al.* Serum or plasma? An old question looking for new answers. *Clinical Chemistry and Laboratory Medicine (CCLM)* **58**, 178–187 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[14. Hortin, G. L., Jortani, S. A., Ritchie, J. C., Valdes, R. & Chan, D. W. Proteomics: A New Diagnostic Frontier. *Clinical Chemistry* **52**, 1218–1222 (2006).](https://www.zotero.org/google-docs/?7nHXa2)

[15. Chen, G. *et al.* Comprehensive Identification and Characterization of Human Secretome Based on Integrative Proteomic and Transcriptomic Data. *Front. Cell Dev. Biol.* **7**, 299 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[16. Beimers, W. F. *et al.* Technical Evaluation of Plasma Proteomics Technologies. *J. Proteome Res.* acs.jproteome.5c00221 (2025) doi:10.1021/acs.jproteome.5c00221.](https://www.zotero.org/google-docs/?7nHXa2)

[17. Dodig-Crnković, T. *et al.* Facets of individual-specific health signatures determined from longitudinal plasma proteome profiling. *EBioMedicine* **57**, 102854 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[18. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Molecular & Cellular Proteomics* **11**, M111.014050 (2012).](https://www.zotero.org/google-docs/?7nHXa2)

[19. Nedelkov, D., Kiernan, U. A., Niederkofler, E. E., Tubbs, K. A. & Nelson, R. W. Investigating diversity in human plasma proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10852–10857 (2005).](https://www.zotero.org/google-docs/?7nHXa2)

[20. Uhlén, M. *et al.* The human secretome. *Sci. Signal.* **12**, eaaz0274 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[21. Zou, E., Xu, X. & Chen, L. Potential of plasma biomarkers for heart failure prediction, management, and prognosis: A multiomics perspective. *Heart Fail Rev* **30**, 55–67 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[22. Colombo, G. *et al.* Plasma Protein Carbonyls as Biomarkers of Oxidative Stress in Chronic Kidney Disease, Dialysis, and Transplantation. *Oxidative Medicine and Cellular Longevity* **2020**, 1–20 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[23. Sarhadi, V. K. & Armengol, G. Molecular Biomarkers in Cancer. *Biomolecules* **12**, 1021 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[24. Eldjarn, G. H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[25. Geyer, P. E. *et al.* The Circulating Proteome─Technological Developments, Current Challenges, and Future Trends. *J. Proteome Res.* **23**, 5279–5295 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[26. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics* **2**, 50 (2003).](https://www.zotero.org/google-docs/?7nHXa2)

[27. Kaier, T. E., Alaour, B. & Marber, M. Cardiac troponin and defining myocardial infarction. *Cardiovascular Research* **117**, 2203–2215 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[28. Tamber, S. S., Bansal, P., Sharma, S., Singh, R. B. & Sharma, R. Biomarkers of liver diseases. *Mol Biol Rep* **50**, 7815–7823 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[29. Kliuchnikova, A. A. *et al.* Blood Plasma Proteome: A Meta-Analysis of the Results of Protein Quantification in Human Blood by Targeted Mass Spectrometry. *IJMS* **24**, 769 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[30. Latypova, L. *et al.* Oxygenation state of hemoglobin defines dynamics of water molecules in its vicinity. *The Journal of Chemical Physics* **153**, 135101 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[31. Chakrabarti, A., Halder, S. & Karmakar, S. Erythrocyte and platelet proteomics in hematological disorders. *Proteomics Clinical Apps* **10**, 403–414 (2016).](https://www.zotero.org/google-docs/?7nHXa2)

[32. Ogawa, C., Tsuchiya, K. & Maeda, K. Reticulocyte hemoglobin content. *Clinica Chimica Acta* **504**, 138–145 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[33. Mohandas, N. & Gallagher, P. G. Red cell membrane: past, present, and future. *Blood* **112**, 3939–3948 (2008).](https://www.zotero.org/google-docs/?7nHXa2)

[34. Carminita, E., C Becker, I. & E Italiano, J. What It Takes To Be a Platelet: Evolving Concepts in Platelet Production. *Circulation Research* **4**, 540–549 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[35. Cooper, N. & Ghanima, W. Immune Thrombocytopenia. *N Engl J Med* **381**, 945–955 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[36. Kremer Hovinga, J. A. & George, J. N. Hereditary Thrombotic Thrombocytopenic Purpura. *N Engl J Med* **381**, 1653–1662 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[37. Laiosa, C. V., Stadtfeld, M. & Graf, T. DETERMINANTS OF LYMPHOID-MYELOID LINEAGE DIVERSIFICATION. *Annu. Rev. Immunol.* **24**, 705–738 (2006).](https://www.zotero.org/google-docs/?7nHXa2)

[38. Scheiermann, C., Frenette, P. S. & Hidalgo, A. Regulation of leucocyte homeostasis in the circulation. *Cardiovasc Res* **107**, 340–351 (2015).](https://www.zotero.org/google-docs/?7nHXa2)

[39. Finlay, D. & Cantrell, D. A. Metabolism, migration and memory in cytotoxic T cells. *Nat Rev Immunol* **11**, 109–117 (2011).](https://www.zotero.org/google-docs/?7nHXa2)

[40. Saravia, J., Chapman, N. M. & Chi, H. Helper T cell differentiation. *Cell Mol Immunol* **16**, 634–643 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[41. Berge, T. *et al.* Quantitative proteomic analyses of CD4+ and CD8+ T cells reveal differentially expressed proteins in multiple sclerosis patients and healthy controls. *Clin Proteom* **16**, 19 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[42. Wang, Q., Qin, Y. & Li, B. CD8+ T cell exhaustion and cancer immunotherapy. *Cancer Letters* **559**, 216043 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[43. *B Cells in Immunity and Tolerance*. vol. 1254 (Springer Singapore, Singapore, 2020).](https://www.zotero.org/google-docs/?7nHXa2)

[44. Kornilov, S. A. *et al.* Multi-Omic characterization of the effects of Ocrelizumab in patients with relapsing-remitting multiple sclerosis. *Journal of the Neurological Sciences* **467**, 123303 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[45. Zhang, E. *et al.* Roles and mechanisms of tumour-infiltrating B cells in human cancer: a new force in immunotherapy. *Biomark Res* **11**, 28 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[46. Cooper, M. A., Fehniger, T. A. & Caligiuri, M. A. The biology of human natural killer-cell subsets. *Trends in Immunology* **22**, 633–640 (2001).](https://www.zotero.org/google-docs/?7nHXa2)

[47. Conigliaro, P., Scrivo, R., Valesini, G. & Perricone, R. Emerging role for NK cells in the pathogenesis of inflammatory arthropathies. *Autoimmunity Reviews* **10**, 577–581 (2011).](https://www.zotero.org/google-docs/?7nHXa2)

[48. Bi, J. & Tian, Z. NK Cell Exhaustion. *Front. Immunol.* **8**, 760 (2017).](https://www.zotero.org/google-docs/?7nHXa2)

[49. Burn, G. L., Foti, A., Marsman, G., Patel, D. F. & Zychlinsky, A. The Neutrophil. *Immunity* **54**, 1377–1391 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[50. Voynow, J. A. & Shinbashi, M. Neutrophil Elastase and Chronic Lung Disease. *Biomolecules* **11**, 1065 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[51. Atiakshin, D. *et al.* Carboxypeptidase A3—A Key Component of the Protease Phenotype of Mast Cells. *Cells* **11**, 570 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[52. Ożańska, A., Szymczak, D. & Justyna, R. Pattern of human monocyte subpopulations in health and disease. *Scandinavian Journal of Immunology* **92**, 1–13 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[53. I. Arleevskaya, M. Mononuclear Phagocytes in Rheumatoid Arthritis Patients and their Relatives - Family Similarity. *TORJ* **5**, 36–44 (2011).](https://www.zotero.org/google-docs/?7nHXa2)

[54. Pan, Y., Yu, Y., Wang, X. & Zhang, T. Tumor-Associated Macrophages in Tumor Immunity. *Front. Immunol.* **11**, 583084 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[55. Moussion, C. & Delamarre, L. Antigen cross-presentation by dendritic cells: A critical axis in cancer immunotherapy. *Seminars in Immunology* **71**, 101848 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[56. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. **422**, (2003).](https://www.zotero.org/google-docs/?7nHXa2)

[57. Altelaar, A. F. M., Munoz, J. & Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14**, 35–48 (2013).](https://www.zotero.org/google-docs/?7nHXa2)

[58. Omenn, G. S. *et al.* The 2024 Report on the Human Proteome from the HUPO Human Proteome Project. *J. Proteome Res.* **23**, 5296–5311 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[59. Aebersold, R. *et al.* How many human proteoforms are there? *Nat Chem Biol* **14**, 206–214 (2018).](https://www.zotero.org/google-docs/?7nHXa2)

[60. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **11**, 49–79 (2009).](https://www.zotero.org/google-docs/?7nHXa2)

[61. Al-Amrani, S., Al-Jabri, Z., Al-Zaabi, A., Alshekaili, J. & Al-Khabori, M. Proteomics: Concepts and applications in human medicine. *WJBC* **12**, 57–69 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[62. Rozanova, S. *et al.* Quantitative Mass Spectrometry-Based Proteomics: An Overview. in *Quantitative Methods in Proteomics* (eds. Marcus, K., Eisenacher, M. & Sitek, B.) vol. 2228 85–116 (Springer US, New York, NY, 2021).](https://www.zotero.org/google-docs/?7nHXa2)

[63. Thomas, S. N., French, D., Jannetto, P. J., Rappold, B. A. & Clarke, W. A. Liquid chromatography–tandem mass spectrometry for clinical diagnostics. *Nat Rev Methods Primers* **2**, 96 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[64. Fernández-Costa, C. *et al.* Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results. *J. Proteome Res.* **19**, 3153–3161 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[65. Silva, A. M. N., Vitorino, R., Domingues, M. R. M., Spickett, C. M. & Domingues, P. Post-translational Modifications and Mass Spectrometry Detection. *Free Radical Biology and Medicine* **65**, 925–941 (2013).](https://www.zotero.org/google-docs/?7nHXa2)

[66. Crutchfield, C. A., Thomas, S. N., Sokoll, L. J. & Chan, D. W. Advances in mass spectrometry-based clinical biomarker discovery. *Clin Proteom* **13**, 1 (2016).](https://www.zotero.org/google-docs/?7nHXa2)

[67. Li, C. *et al.* Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Front. Chem.* **9**, 813359 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[68. Viode, A. *et al.* A simple, time- and cost-effective, high-throughput depletion strategy for deep plasma proteomics. *Sci. Adv.* **9**, eadf9717 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[69. Cao, Z., Tang, H.-Y., Wang, H., Liu, Q. & Speicher, D. W. Systematic Comparison of Fractionation Methods for In-depth Analysis of Plasma Proteomes. *J. Proteome Res.* **11**, 3090–3100 (2012).](https://www.zotero.org/google-docs/?7nHXa2)

[70. Riley, N. M., Bertozzi, C. R. & Pitteri, S. J. A Pragmatic Guide to Enrichment Strategies for Mass Spectrometry–Based Glycoproteomics. *Molecular & Cellular Proteomics* **20**, 100029 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[71. Ferdosi, S. *et al.* Engineered nanoparticles enable deep proteomics studies at scale by leveraging tunable nano–bio interactions. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2106053119 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[72. Blume, J. E. *et al.* Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nat Commun* **11**, 3662 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[73. Maes, E., Cools, N., Willems, H. & Baggerman, G. FACS-Based Proteomics Enables Profiling of Proteins in Rare Cell Populations. *IJMS* **21**, 6557 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[74. Sun, Y. & Sethu, P. Low-stress Microfluidic Density-gradient Centrifugation for Blood Cell Sorting. *Biomed Microdevices* **20**, 77 (2018).](https://www.zotero.org/google-docs/?7nHXa2)

[75. Bennett, H., Stephenson, W., Rose, C. & Darmanis, S. Single-cell proteomics enabled by next-generation sequencing or mass spectrometry. *Nature Methods* **3**, 363–374 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[76. Smith, J. G. & Gerszten, R. E. Emerging Affinity-Based Proteomic Technologies for Large-Scale Plasma Profiling in Cardiovascular Disease. *Circulation* **135**, 1651–1664 (2017).](https://www.zotero.org/google-docs/?7nHXa2)

[77. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS ONE* **5**, e15004 (2010).](https://www.zotero.org/google-docs/?7nHXa2)

[78. Feng, W. *et al.* NULISA: a proteomic liquid biopsy platform with attomolar sensitivity and high multiplexing. *Nat Commun* **14**, 7238 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[79. Beutgen, V. M. *et al.* Secretome Analysis Using Affinity Proteomics and Immunoassays: A Focus on Tumor Biology. *Molecular & Cellular Proteomics* **23**, 100830 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[80. Duarte, J. G. & Blackburn, J. M. Advances in the development of human protein microarrays. *Expert Review of Proteomics* **14**, 627–641 (2017).](https://www.zotero.org/google-docs/?7nHXa2)

[81. Katz, D. H. *et al.* Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci. Adv.* **8**, eabm5164 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[82. LaCava, J. *et al.* Affinity Proteomics to Study Endogenous Protein Complexes: Pointers, Pitfalls, Preferences and Perspectives. *BioTechniques* **58**, 103–119 (2015).](https://www.zotero.org/google-docs/?7nHXa2)

[83. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[84. Raffield, L. M. *et al.* Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics* **20**, 1900278 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[85. Haslam, D. E. *et al.* Stability and reproducibility of proteomic profiles in epidemiological studies: comparing the Olink and SOMAscan platforms. *Proteomics* **22**, 2100170 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[86. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Molecular & Cellular Proteomics* **22**, 100577 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[87. Perez-Riverol, Y. Proteomic repository data submission, dissemination, and reuse: key messages. *Expert Review of Proteomics* **19**, 297–310 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[88. Perez-Riverol, Y. *et al.* The PRIDE database at 20 years: 2025 update. *Nucleic Acids Research* **53**, D543–D553 (2025).](https://www.zotero.org/google-docs/?7nHXa2)

[89. Choi, M. *et al.* MassIVE.quant: a community resource of quantitative mass spectrometry–based proteomics datasets. *Nat Methods* **17**, 981–984 (2020).](https://www.zotero.org/google-docs/?7nHXa2)

[90. Deutsch, E. W. The PeptideAtlas Project. in *Proteome Bioinformatics* (eds. Hubbard, S. J. & Jones, A. R.) vol. 604 285–296 (Humana Press, Totowa, NJ, 2010).](https://www.zotero.org/google-docs/?7nHXa2)

[91. Uhlen, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366**, eaax9198 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[92. Dens, C., Adams, C., Laukens, K. & Bittremieux, W. Machine Learning Strategies to Tackle Data Challenges in Mass Spectrometry-Based Proteomics. *J. Am. Soc. Mass Spectrom.* **35**, 2143–2155 (2024).](https://www.zotero.org/google-docs/?7nHXa2)

[93. Moriya, Y. *et al.* The jPOST environment: an integrated proteomics data repository and database. *Nucleic Acids Research* **47**, D1218–D1224 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[94. Ma, J. *et al.* iProX: an integrated proteome resource. *Nucleic Acids Research* **47**, D1211–D1217 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[95. Sharma, V. *et al.* Panorama: A Targeted Proteomics Knowledge Base. *J. Proteome Res.* **13**, 4205–4210 (2014).](https://www.zotero.org/google-docs/?7nHXa2)

[96. Perez‐Riverol, Y., Alpi, E., Wang, R., Hermjakob, H. & Vizcaíno, J. A. Making proteomics data accessible and reusable: Current state of proteomics databases and repositories. *Proteomics* **15**, 930–950 (2015).](https://www.zotero.org/google-docs/?7nHXa2)

[97. Deutsch, E. W. *et al.* Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0. *J. Proteome Res.* **18**, 4108–4116 (2019).](https://www.zotero.org/google-docs/?7nHXa2)

[98. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).](https://www.zotero.org/google-docs/?7nHXa2)

[99. Zhong, W. *et al.* Next generation plasma proteome profiling to monitor health and disease. *Nat Commun* **12**, 2493 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[100. Huang, Q., Szklarczyk, D., Wang, M., Simonovic, M. & Von Mering, C. PaxDb 5.0: Curated Protein Quantification Data Suggests Adaptive Proteome Changes in Yeasts. *Molecular & Cellular Proteomics* **22**, 100640 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[101. Brenes, A. J., Lamond, A. I. & Cantrell, D. A. The Immunological Proteome Resource. *Nat Immunol* **24**, 731–731 (2023).](https://www.zotero.org/google-docs/?7nHXa2)

[102. Nanjappa, V. *et al.* Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucl. Acids Res.* **42**, D959–D965 (2014).](https://www.zotero.org/google-docs/?7nHXa2)

[103. Shao, D. *et al.* HBFP: a new repository for human body fluid proteome. *Database* **2021**, baab065 (2021).](https://www.zotero.org/google-docs/?7nHXa2)

[104. Freeberg, M. A. *et al.* The European Genome-phenome Archive in 2021. *Nucleic Acids Research* **50**, D980–D987 (2022).](https://www.zotero.org/google-docs/?7nHXa2)

[105. Bhattacharya, S. *et al.* ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data* **5**, 180015 (2018).](https://www.zotero.org/google-docs/?7nHXa2)

[106. Reiter, L. *et al.* Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. *Molecular & Cellular Proteomics* **8**, 2405–2417 (2009).](https://www.zotero.org/google-docs/?7nHXa2)

[107. Howden, A. J. M. *et al.* Quantitative analysis of T cell proteomes and environmental sensors during T cell differentiation. *Nat Immunol* **20**, 1542–1554 (2019).](https://www.zotero.org/google-docs/?7nHXa2)