ABSTRACT

Title of Dissertation:	QUANTITATIVE ANALYSIS OF INTACT PROTEINS AND RNAS CARRIED BY IMMUNOSUPPRESSIVE EXOSOMES
	Lucía Giorgina Geis Asteggiante, Doctor of Philosophy, 2016
Dissertation directed by:	Professor Catherine Fenselau, Department of Chemistry and Biochemistry

Myeloid-derived suppressor cells (MDSC) are immature myeloid cells that accumulate in the tumor microenvironment of most cancer patients. They are a major obstacle to immunotherapy because they suppress both adaptive and innate immune responses. MDSCs collected from tumor-bearing mice release nano-sized vesicles, called exosomes, which carry biologically active molecules and participate in intercellular communication. Exosomes released by MDSC stimulate migration of other MDSC towards the tumor microenvironment and convert macrophages to a tumor-promoting phenotype. Among the proteins identified in MDSC-released exosomes, S100A8 and S100A9 are low-mass, highly abundant, pro-inflammatory mediators already known to contribute directly to the immune suppressive functions of MDSC.

The aim of this work was to successfully interrogate the exosomal intact protein cargo using top-down proteomics, a strategy for protein analysis that has not previously been applied to exosomes of any kind. Several protein forms (proteoforms) were fully

characterized, which is critical as post-translational modifications regulate protein functions, cellular location and protein interactions. Additionally, since the tumor promoting activity of MDSC is enhanced by inflammation, we focused on evaluating the effect of increased inflammation on the proteoforms relative abundance using current top-down label-free quantitation techniques (peak intensities and peak areas), and comparing them to our recently validated spectral counting approach. Using spectral counting we were able to estimate differences in abundances of both S100A8 and S100A9 proteoforms.

Furthermore, it has been previously reported that exosomes can carry micro RNAs and messenger RNAs. In order to investigate if MDSC-derived exosomes also contain RNAs, a collaborative study was carried out entailing the qualitative and quantitative analysis of miRNAs, mRNA and proteins present in MDSC and their exosomes, and evaluate their changes due to heightened inflammation. The MDSC and exosome protein cargo was analysed by bottom-up proteomics in this case, and the RNA cargo by next generation sequencing. A large number of mRNA and miRNA species were found to be carried by MDSC-derived exosomes and, strikingly, their putative functions were associated to MDSC expansion and suppressive function, and cancer development.

QUANTITATIVE ANALYSIS OF INTACT PROTEINS AND RNAS CARRIED
BY IMMUNOSUPPRESSIVE EXOSOMES

by

Lucía Giorgina Geis Asteggiante

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Catherine Fenselau, Chair
Professor Najib El-Sayed, Dean's Representative
Professor Nathan Edwards
Professor Nicole LaRonde
Professor Neil Blough

Dedication

To my husband, my family and my most dear friends.

You have loved and nurtured me. I will always be grateful.

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3'-UTR	3'-untranslated region
AGO-2	Protein argonaute-2
Alix	Apoptosis-linked gene 2-interacting protein X
ARGI	Arginase I
C/EBPβ	CCAAT/enhancer binding protein-β
CAH	Carbonic anhydrase
CD100	Semaphorin-4D
CD101	Immunoglobulin superfamily member 2
CD107b	Lysosome-associated membrane glycoprotein 2
CD114	Granulocyte colony-stimulating factor receptor
CD115	Macrophage colony-stimulating factor 1 receptor
CD11a	Integrin alpha-L
CD11b	Integrin alpha-M
CD121b	Interleukin-1 receptor type 2
CD133	Prominin-1
CD147	Isoform 2 of Basigin
CD148	Receptor-type tyrosine-protein phosphatase eta
CD156c	Disintegrin and metalloproteinase domain-containing protein 10
CD157	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2
CD163	Scavenger receptor cysteine-rich type 1 protein M130
CD171	Neural cell adhesion molecule L1
CD172a/Sirp-α	Tyrosine-protein phosphatase non-receptor type substrate 1
CD18	Integrin beta-2
CD182	C-X-C chemokine receptor type 2
CD205	Lymphocyte antigen 75
CD29	Integrin beta-1
CD298	Sodium/potassium-transporting ATPase subunit beta-3
CD300f	CMRF35-like molecule 1
CD39	Ectonucleoside triphosphate diphosphohydrolase 1
CD41	Integrin alpha-IIb
CD42b	Platelet glycoprotein Ib alpha chain
CD42d	Platelet glycoprotein V
CD45	Receptor-type tyrosine-protein phosphatase C
CD49b	Integrin alpha-2
CD49f	Integrin alpha-6
CD56	Neural cell adhesion molecule 1
CD5L	CD5 antigen-like
CD61	Integrin beta-3
CD62L	L-selectin

CD62P	P-selectin
CD66a	Carcinoembryonic antigen-related cell adhesion molecule 1
CD98	4F2 cell-surface antigen heavy chain
CID	Collision induced dissociation
cpm	Counts per million
DAVID	Database for Annotation, Visualization and Integrated Discovery
DE	Dynamic exclusion
DTT	Dithiothreitol
EIC	Extracted ion chromatograms
ESCRT	Endosomal sorting complex required for transport
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FADD	Fas-Associated protein with Death Domain
FDR	False discovery rate
FN	False negative
FP	False positive
G-CSF	Granulocyte colony-stimulating factor
GELFrEE	Gel-eluted liquid fraction entrapment electrophoresis
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GO	Gene Ontology
HCD	Higher-energy collisional dissociation
HF	High field
HMGB-1	High-mobility group box 1
hnRNP	Heterogeneous ribonucleoproteins
IKKβ	Inhibitor of nuclear factor kappa-B kinase subunit beta
IL-1β	Interleukin-1β
IL-6	Interleukin-6
ILV	Intra-luminal vesicles
IRAK	IL-1R-associated kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS/MS	Liquid chromatography tandem mass spectrometry
limma	Linear model for microarray
Ly6C	Lymphocyte antigen 6C1
Ly6C	Lymphocyte antigen 6C1
Ly6G	Lymphocyte antigen 6G
Ly6G	Lymphocyte antigen 6G
MAF	Macrophage-activating factor
MDSCs	Myeloid-derived suppressor cells
MEF2C	Myeloid ELF1-like factor 2C
MHC	Major histocompatibility complex
MIF	Macrophage migration inhibitor factor

MMP9	Metalloproteinase-9
MS	Mass spectrometry
MVB	Multi-vesicular bodies
MW	Molecular weight
NFIA	Nuclear factor I/A
NF-κB	Nuclear factor-κB subunit 1
NK	Natural killer
ORF	Open reading frame
PA28α	Proteasome activator complex subunit 1
PA28β	Proteasome activator complex subunit 2
PDCD4	Programmed cell death 4
PGE$_2$	Prostaglandin E2
PepArML	Peptide identification arbiter by machine learning
PF4	Platelet factor 4
PIR	Protein Information Resource
PIR-B	Leukocyte immunoglobulin-like receptor subfamily B member 3
PrSM	Protein spectrum match
PSM	Peptide spectrum match
PTEN	Phosphatase and tensin homologue
PTMs	Post-translational modifications
RNase	Ribonuclease A
rRNA	Ribosomal RNA
Rsc	Ratios from spectral counts
RUNX1	Runt-related transcription factor 1
S/N	Signal-to-noise
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SHIP1	SH2-domain-containing inositol-5-phosphatase 1
SILAC	Stable Isotope Labeling by Amino acids in Cell culture
SPRED-1	Sprouty Related EVH1 Domain Containing 1
STAT	Signal transducer and activator of transcription
STAT	Signal transducer and transcription activator
SVA	Surrogate variable analysis
TIC	Total ion chromatogram
TLR	Toll-like receptors
TMT	Tandem mass tag
TNF-α	Tumor necrosis factor alpha
TRAF6	TNFR-associated factor 6
T$_{Reg}$	Regulator T cells
TSP-1	Thrombospondin-1
VEGF	Vascular endothelial growth factor
Vps	Vacuolar protein sorting-associated proteins

Chapter 1: Introduction

1.1. Research significance and objectives

Myeloid-derived suppressor cells (MDSCs) are immature myeloid cells produced during aberrant myelopoiesis that accumulate in the tumor microenvironment of most cancer patients.[1] The tumor microenvironment is a complex environment in which tumor and host cells interact.[2] There MDSCs suppress both adaptive and innate immune responses.[1] MDSCs collected from tumor-bearing mice profusely release nano-scale membrane-bound extracellular vesicles, called exosomes, which carry biologically active proteins between cells.[3] These exosomes stimulate migration of MDSC, due to the presence of pro-inflammatory mediators (S100A8 and S100A9), and mediate MDSC conversion of macrophages to a tumor-promoting phenotype.[3] MDSC immune suppressive activity is enhanced under heightened inflammation conditions, promoting tumor progression and metastasis.[4–7] Therefore, these cells hinder current immunotherapies. The broad objectives of our research group is to study the cargo carried by MDSC-derived exosomes, in order to help understand MDSC immune suppression mechanisms, exosomal communication within the tumor microenvironment and more broadly, the biological characteristics of exosomes. We also expect our study to help define new therapeutic targets.

My Ph.D. research work has focused on the interrogation and relative quantitation of the protein, mRNA and miRNA cargoes of exosomes shed by MDSC using mass spectrometry and next generation sequencing. In this work, MDSCs and

MDSC-derived exosomes from tumor-bearing mice were obtained by injecting mice with 4T1 and 4T1/IL-1β tumor mammary carcinoma cells. The 4T1/IL-1β tumor cells are 4T1 tumor cells that were transduced to express the cytokine interleukin-1β, in order to increase inflammation in the tumor microenvironment. This disturbance allows us to study the MDSC and exosomal cargo formed under "conventional" and "inflammatory" tumor microenvironment conditions. Both MDSC and exosomes were provided by our collaborator Dr. Ostrand-Rosenberg at the University of Maryland Baltimore County.

The first objective of my research was the successful identification of the exosome protein cargo using top-down mass spectrometry, a proteomics approach that has not been previously applied to any kind of exosomes. Top-down proteomics allows the characterization of intact proteins and provides information on the many possible protein forms (proteoforms) that can be present in a biological sample due to post-translational modifications (PTMs) and sequence variants. The full characterization of proteoforms, such as those of S100A8 and S100A9, is critical because PTMs may regulate protein functions, cellular location and protein interactions.[8–10]

Considering that inflammation-induced MDSCs suppress immune response and promote tumor progression more aggressively than conventional,[4–7] the qualitative and quantitative effects of increased inflammation on exosomal proteoforms could provide new insight on exosome and MDSC activity. However, top-down proteomics is still a relatively new field of study and approaches for protein quantitation are not yet

established. The second objective aimed to evaluate the applicability of spectral counting, a well-established quantitation approach in bottom-up proteomics (peptide analysis), to top-down proteomics workflows. This work included the assessment of sensitivity, precision and accuracy of spectral counting for proteoform relative quantitation using a mixture of protein standards.

Apart from carrying proteins, exosomes can also carry RNAs.[11] Several miRNAs have been reported to regulate MDSC activity.[12] The transfer of miRNA and mRNA via exosomes into surrounding cells has been reported previously.[13–15] The third objective was to identify and quantify the mRNA and miRNA species present in MDSC and MDSC-derived exosomes collected under conventional and inflammatory conditions. This analysis provided predictive information about signaling pathways in receiver cells that may be affected by the exosome cargo.

1.2.1. Proteomics approaches

Proteomics comprises the study of the proteome in which mass spectrometry (MS) plays a key role. Two main proteomics approaches are bottom-up and top-down. Bottom-up proteomics entails chemical or enzymatic digestion of the proteins into peptides prior to liquid chromatography tandem mass spectrometry (LC-MS/MS). In a typical MS/MS analysis, peptides are ionized by electrospray ionization (ESI), precursor ions isolated and collisionally excited to form fragments. The mass spectra of fragment ions are then compared to MS/MS spectra predicted from amino acid sequences to identify the peptides and infer proteins.[16] The major advantages of bottom-up proteomics are that it is straightforward, simple to carry out in solution, can be performed at lower mass resolution, and the bioinformatics tools used for data analysis are well established.[17] However, it also has limitations such as low sequence coverage, limited ability to identify protein variants, and lack of reliable information on post-translational modifications.[18] In the case of top-down proteomics, no digestion is performed prior to LC-MS/MS; hence intact proteins must be ionized, precursor ions isolated and fragmented. This approach offers information complementary to bottom-up, providing molecular weight determination, identification of protein variants, identification of one or multiple PTMs and determination of PTMs sites. However, top-down also has its own limitations. High mass resolution is needed for determining precursor and product ions masses. Also in order to obtain good fragmentation from collision induced dissociation (CID) or electron transfer dissociation (ETD), more ions

need to be accumulated, activated and detected, requiring longer instrumental duty cycles and limiting the ability to analyze complex mixtures on a chromatographic time scale.[17,19,20] For these reasons higher protein concentration and fractionation are generally required to achieve sufficient data quality for identification and quantitation. Additionally, only a few bioinformatics tools for top-down are freely or commercially available, and many of them are still under development.

1.2.2. Electrospray Ionization (ESI)

Several ionization techniques are used for protein analysis. Most commonly, the transition from analytes in solution-phase to ionized molecules in the gas-phase is achieved using a nanoESI source, which is set to deliver a solution at nanoscale flow rates (nL/min) into the MS. This source ionizes the analytes under atmospheric pressure by applying a potential difference (kV) between the emitter tip (used to deliver the solution into the MS) and the counter electrode, producing an intense electric field.[21] When the voltage is applied the charged components in the solution start separating and cations migrate towards the end of the tip. As the solution leaves the capillary, a Taylor cone is formed and charged droplets are emitted.[22] As the droplets travel towards the counter electrode they evaporate and shrink, until the surface tension of the shrunk droplets equals the repulsion produced by the charged species present. At this point, the Rayleigh instability limit is reached and smaller droplets are formed by fission.[22] The droplet radius is just a few nm when the charged species are released into the gas-phase.

There are three proposed mechanisms that can explain the release of molecular ions depending on their size and shape: (1) the ion evaporation model (IEM), based on the observation that high intensity electric fields applied to nanodroplets, containing low molecular weight species, cause the direct expulsion of small ions from the surface;[21,22] (2) the charge residue model (CRM), based on the transfer of charge from the solvent of a droplet to a large globular specie contained in it that is released into the gas phase when dryness is reached;[21,22] and (3) the chain ejection model (CEM) in which non-polar polymer chains, such as an unfolded protein, locate their non-polar regions toward the surface of the droplet and are sequentially extruded until completely ejected from the droplet.[22] Unfolded proteins are ionized with multiple charges, allowing proteins with large masses to fall inside the instrument *m/z* range (200 – 2000) and be detected.

1.2.3. Orbitrap instruments

Top-down mass spectrometry requires the use of mass analyzers that provide high mass resolution and accuracy. The orbitrap was invented in the 1990s by Alexander Makarov.[23,24] It contains a spindle-like axial central electrode and a pair of outer electrodes positioned symmetrically to each other (see Figure 1.1).[25] In this instrument ion trapping is achieved using only electrostatic fields, and hence, is not limited by the need for an intense magnetic field. The orbitrap electrostatic potential distribution is shown in Eq. (1); it is composed of an axial quadrupole field (produced by the outer electrodes) and a radial logarithmic field (from the central electrode).[25] The orbitrap design was based on previous studies on a Kingdon Trap by Knight[26] and

Gillig et al.,[27] and uses the spindle-like geometry for the central electrode in order to provide a purely harmonic potential in the axial direction with a frequency that is independent of the ion's initial conditions.[24,25] A detailed description of ion trajectories and frequencies can be found elsewhere.[24] The radial, axial and rotational frequencies are shown here in Eq. (2-4). Both radial and axial frequencies can be used for *m/z* determination, however the axial frequency is preferred because it is independent of the ion's initial velocity and radius. Ions are injected into the orbitrap radially through a slot offset from the center of the trap, in order to start coherent axial oscillations.[24] A deflector is used to compensate for electric field perturbations caused by the slot present in the outer electrode. As the ion packets move coherently an image current is detected, the axial harmonic frequency obtained by Fourier transform and *m/z* calculated.[25]

$$U(r,z) = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}(R_m)^2 \ln\left[\frac{r}{R_m}\right] + C$$

Eq. (1)

Axial oscillation frequency: $\quad \omega = \sqrt{(q/m)k}$

Eq. (2)

Radial oscillation frequency: $\quad \omega_r = \omega\sqrt{\left(\frac{R_m}{R}\right)^2 - 2}$

Eq. (3)

Rotation frequency:

Eq. (4)

$$\omega_\varphi = \omega\sqrt{\frac{\left(\frac{R_m}{R}\right)^2 - 1}{2}}$$

where k is the axial restoring force, r and z are coordinates, R_m is the characteristic radius, C is a constant, q is the ion charge, R is the radius of the ion trajectory.[24]

There are two orbitrap designs commercially available, the standard orbitrap and the high field (HF) orbitrap, which mainly differ in their geometry.[28] The standard

orbitrap design is available in several instruments (LTQ orbitrap XL, Velos and Q Exactive), however newer instruments use the HF orbitrap (Elite, Q Exactive HF, Fusion and Fusion Lumos). The major advantage of the HF orbitrap design is the increase in resolving power by approximately 50% at a central electrode voltage of 3.5 KV and 80% at 5 KV, without compromising sensitivity.[29] Thermo Scientific reports a maximum resolution power (at 200 *m/z*) of 140,000 for the standard orbitrap and 240,000 for the HF design.

In this research work, both the LTQ orbitrap XL and Fusion Lumos were used. Significant improvements were made in the span of 10 years between our acquisitions of the orbitrap XL and Fusion Lumos instruments, as shown in the manufacturer's schematics (Figure 1.1a,b). The HF orbitrap provides higher scan rates and resolving power. Additionally, higher sensitivity is obtained due to improved ion transfer, as lower vacuum pressures are achievable, and effective and rapid precursor isolation by the quadrupole mass filter installed in the front of the system.[30] Another significant improvement in the Fusion Lumos is increased ETD (see below) fragmentation efficiency due to the new design of the ETD reagent (fluoranthene) source, and the enlarged trapping capacity of the high pressure linear ion trap to accumulate reagent and analyte ions for ETD reaction. A gain of 3-fold in fragment ion signal to noise ratio is reported for ETD.[31] Note that the first model of the LTQ orbitrap XL did not have ETD capabilities. In our case, the system was retro-fitted to perform ETD. Initial designs of ETD capable orbitrap instruments located the ETD source at the back of the instrument, which has now been relocated to the front.[30]

(a)

(b)
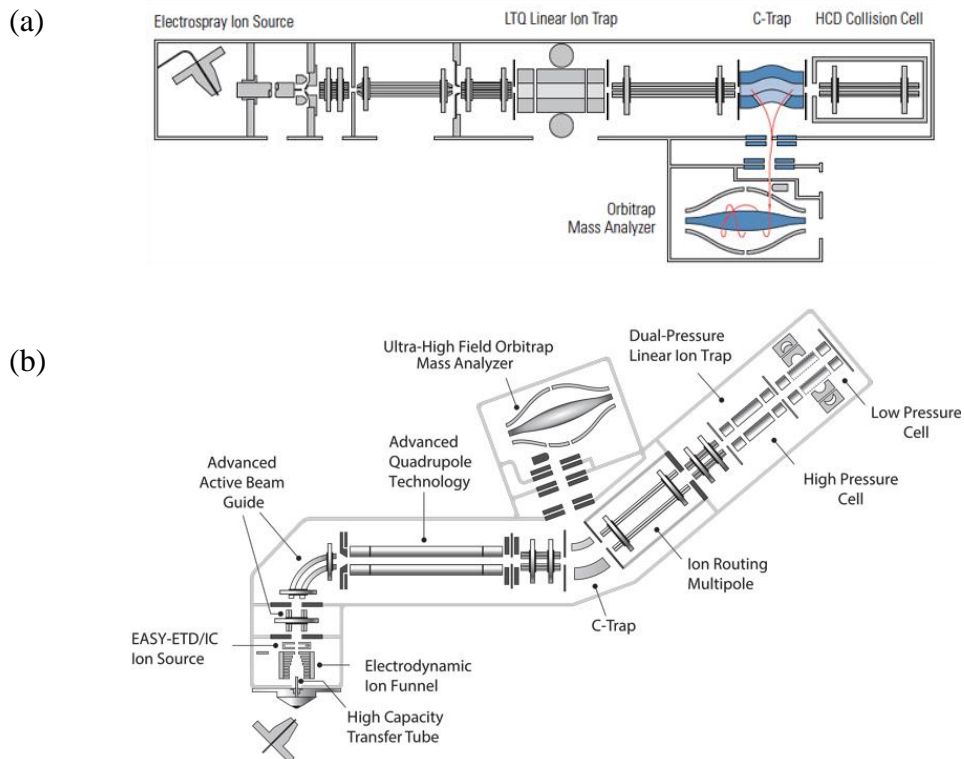
Figure 1.1. Orbitrap mass analyzers used in this study: (a) LTQ orbitrap XL and (b) orbitrap Fusion Lumos. (Schematics available at http://planetorbitrap.com)

1.2.4. Fragmentation techniques

Fragmentation is crucial for protein identification using top-down workflows. The two most commonly used fragmentation techniques are collision induced dissociation and electron transfer dissociation.

Collision induced dissociation (CID) relies on the conversion of kinetic energy of accelerated precursor ions in the gas phase to internal (vibrational) energy by collision with an inert gas (generally helium or argon). The dissociation of a molecule depends on the amount of internal energy acquired from the collision, which depends

on the acceleration voltage applied, the precursor and inert gas masses and the distance difference between the center of mass of the colliding ions.[32] The maximum energy conversion is achieved when the center of mass are aligned and is estimated by Eq. (5), assuming that the inert gas velocity is zero.[33]

$$KE_{lab} = (KE_{com}\, m_T)\,/\,(m_T + m_p) \hspace{3cm} \text{Eq. (5)}$$

where KE_{lab} is the laboratory kinetic energy, m_p and m_T are the precursor ion and inert gas masses and KE_{com} is the gained internal energy.

However, the energy gained is often lower than that estimated by Eq. (5), and many collisions may be required to accumulate sufficient internal energy to break a bond.[33] CID of peptides and proteins, performed in ion traps and quadrupoles or multipoles, mainly dissociates the polypeptide backbone at the amide bond. This fragmentation produces a sequence of b- ions and y- type ions by breaking the peptide bond (C-N) and retaining the N- or C-terminus of the protein, respectively (see Figure 1.2). Additionally, neutral losses of water or ammonia are usually observed.

The main disadvantages of CID for top-down analysis are the lack of fragmentation on the center sections of the protein sequences limiting residue cleavage coverage,[17,34] and the loss of labile post-translational modifications (PTMs) such as phosphorylation.[18] Note that a variation of CID called higher-energy collisional dissociation (HCD) by Thermo Fisher is also widely used in proteomics. The principle

of this activation method is the same as CID but the energy provided for dissociation is higher.[35]

Electron transfer dissociation (ETD) achieves protein fragmentation through an ion-ion reaction with a polycyclic aromatic radical anion, such as fluoranthene or anthracene.[21,34] This radical anion is produced by negative chemical ionization in a separate area of the instrument and guided to the reaction cell, where it reacts with the precursor ions transferring an electron. The electron transfer causes backbone fragmentation at the amide nitrogen-alpha carbon (N-C_α) bond, forming primarily c- and z-type ions that retain the N- and C-terminus of the protein, respectively (see Figure 1.2).[34,36] ETD is a non-ergodic fragmentation technique that offers the advantage of producing fragmentation at the site where the electron is captured (independent of the bond strength) and preserves side chains and PTMs.[34] ETD fragmentation often generates charge reduced species that do not fragment. It provides more efficient fragmentation of precursor ions that carry higher charge states. Further fragmentation of stable charge reduced species can be obtained with supplemental activation by CID or HCD.[37]
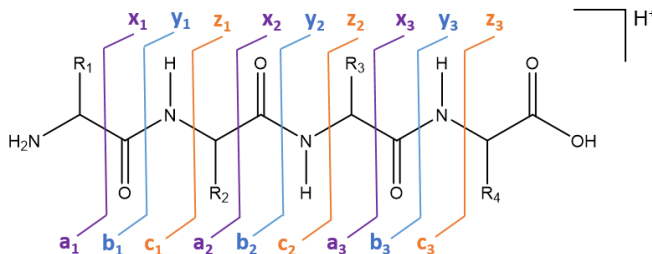
Figure 1.2. Dominant peptide backbone fragmentation by CID is represented in blue and by ETD is represented in orange.

11

1.3.1. Bottom-up proteomics

Automated high-throughput bottom-up protein identification bioinformatics algorithms are well established. In general when performing a sequence database search, the user needs to provide information on the protein sequence database of the organism studied and the protease(s) utilized. An in silico digestion of the protein sequence database is performed in order to obtain all possible peptides that can be produced (including, if desired, flexibility for protease missed cleavages) and each peptide MS/MS fragment ion series, which are generated following current fragmentation theories.[16,38] In order to identify the peptides present in a sample, the experimental peptide and fragment *m/z* values are compared to the predicted MS/MS spectra and a probability score is assigned to each peptide spectrum match (PSM) obtained based on spectrum similarities.[38]

There are many freely available search engines such as MSGF+,[39,40] X!Tandem with native,[41] k-score[42] and s-score[42] scoring, OMSSA,[43] Andromeda,[44] MSAmanda[45] and MyriMatch.[46] Additionally, other popular commercially available search engines are MASCOT[47] and SEQUEST.[48] The search scoring is performed in different ways depending on the search engine used. Commonly used statistical methods for PSM scoring are cross correlation,[48] Bayesian probability,[47] Poisson distribution,[43] dot-product,[40,41] binomial distribution,[44,45] and multivariate hypergeometric distribution.[46] The scores obtained can be transformed into *p*-values and/or expectation values (E-values). However, these individual PSM statistics are not sufficient to assess the protein

identification confidence. Typically, datasets contain millions of spectra to be compared; hence corrections for multiple testing are necessary.

After a list of scored PSMs is obtained, the identification confidence is assessed by estimating the spectral false discovery rate (FDR),[49] which is the proportion of false positive peptide identifications over the total number of peptide identifications. This is generally done by performing a target-decoy search, for which a decoy database of the same size than the targeted database is created by reversing or randomizing the amino acid sequences of the organism protein sequence database, which is then concatenated with the target database. The search is then performed against the concatenated database. As both databases have the same size, it can be assumed that their false positives distributions are similar and that the number of decoy identifications obtained estimates the number of false positives observed in the target search, allowing the estimation of the spectral FDR.[50] PSMs are then filtered using a cutoff spectral FDR chosen by the user and grouped into their corresponding proteins, following the parsimony principle, for protein inference.

Although proteins are inferred from their peptides, the confidence of these peptide identifications does not translate directly into protein identification confidence. The main reason being that the identified PSMs (peptides) can be shared between more than one putative protein, grouping into a smaller number of inferred proteins, even though a large number of PSMs were identified. [51,52] However, each PSM that erroneously matches any of the proteins in the sequence database will directly increase

the number false positive identifications compared to the smaller number of inferred proteins, causing the FDR at the protein level to be larger than those observed for its PSMs.[51,52] Post-search algorithms able to estimate protein level FDR are ProteinProphet[53] and MAYU.[54] ProteinProphet's cumulative score takes into account the combination of the PSMs probabilities and is adjusted for the effect of non-random peptide grouping.[53] MAYU estimates protein FDR using the target-decoy results, and it assumes that the number of false positive protein identifications follow a hypergeometric distribution.[54]

Additionally, there is software available to combine results from several search engines. Peptide identification arbiter by machine learning (PepArML)[55] is a freely available platform capable of combining the results of up to 7 different search engines: MSGF+, X!Tandem with native, k-score, and s-score scoring, OMSSA, MASCOT, and MyriMatch. This platform provides a higher number of peptide identifications than the individual searches.[55] In a similar fashion, Proteome Discoverer v2.1 is a commercially available software capable of combining results from various search engines (e.g. MSAmanda, MASCOT and SEQUEST).

There are many approaches available to perform bottom-up quantitation. This research centers on the use of label-free quantitation approaches, such as spectral counting, chromatographic peak intensities or peak areas, which allow retrospective analysis of samples with no additional sample preparation steps, reducing costs and time. Spectral counting for relative quantitation of proteins is based on the principle that the likelihood of identifying a peptide is higher if the abundance of the protein of

origin is high. Hence, it is expected that the counts of peptide identifications or protein spectrum matches (PSMs) of a protein correlate with its abundance.[56–59] Zybailov *et al.* has demonstrated that spectral counting provides reproducible results for peptides and larger dynamic ranges than approaches based on peak intensity or area.[60] However, peak intensities or areas provides more accurate results.[61,62]

1.3.2. Top-down proteomics

Top-down bioinformatics tools are less mature than their bottom-up counterparts. ProSightPC was the first search engine to match high-resolution, accurate MS and MS/MS data of intact proteins with protein sequence database, such as Uniprot Knowledge base.[63,64] The user provides the protein sequence database of interest, which could include information on previously reported PTMs or PTMs that are inferred by sequence similarity in Uniprot. In the absolute mass search, the software considers each MS/MS spectrum acquired inside the precursor ion mass window selected. The observed precursor ion mass and fragmentation is compared with the theoretical, matching peaks counted and an expected value (E-value) computed to measure the statistical significance of the protein spectrum match (PrSM) or protein identification.[63,64] A small E-value means that the expected number of proteins with this many matching peaks that are identified by chance is small. Inside ProSightPC, the SequenceGazer tool is available for manual inspection of the assigned fragments and propose putative PTM assignments or residue deletion, with recalculated E-values. In late 2015, a plug-in of ProSightPC became available for Proteome Discoverer 2.1,

called ProSightPD 1.0. This new format of ProSightPC includes all the above-explained capabilities, though those tasks previously performed in SequenceGazer are now substituted by ProsightLite.[65] Additionally, similarly to bottom-up bioinformatics, after target-decoy searches are performed ProSightPD is able to provide a list of protein identifications that are under a selected FDR threshold (typically 1% FDR).

TopPIC is a freely-available software that similarly to ProSightPC is capable of searching against a protein sequence database, identifying proteins and providing an E-value and/or false discovery rate (FDR) value (if searched against a decoy database).[66–68] The most significant advantage over ProSightPC is that TopPIC searches are significantly faster. A recent software update allows the user to add a list of putative PTMs to be considered as variable modifications during the search. Up to 2 "unknown" PTMs may be assigned using the MIScore method.[69] Unfortunately, no interactive PTM manipulation tool is provided.

MASH Suite is another freely-available software capable of performing top-down protein identifications.[70] The first version of this software was only capable of targeted analysis, which required *a priori* knowledge of the protein(s) under study. In 2016, MASH Suite Pro[71] was released and this version of the software, similarly to TopPIC, includes search engine capabilities and the discovery of unknown PTMs using the MSAlign+ algorithm.[68]

Independently of the software used, prior to start a search, the raw data obtained from the instrument needs to be deconvoluted using a deconvolution algorithm (THRASH,[72] MSDeconv[73] or Xtract[74]), which provides the decharged monoisotopic and average masses of the precursor and fragment ions. It is important to notice that TopPIC and ProSightPC software do not offer quantitation tools, as quantitation approaches in top-down are not very advanced. The significance associated to top-down proteomics quantitation is high because it opens the door to the study of proteoform relative abundances, which could provide useful biological information.[75] Label-free approaches using extracted ion chromatogram intensities or areas have been successfully used[76,77] and will be further discussed in Chapter 3.

1.4. Myeloid derived suppressor cells

Myeloid-derived suppressor cells (MDSCs) are a heterogeneous population of immature myeloid cells, which are present in all cancer patients. MDSCs are identified by the plasma membrane markers CD11b and Gr1 in mice, and by the markers CD11b, CD14, CD15, CD33 and absence of HLA-DR in humans.[1] MDSC are produced by abnormal differentiation of the myeloid linage and are known to accumulate in the tumor microenvironment, where they suppress both adaptive and innate immunity.[78] Under normal conditions, haematopoietic stem cells differentiate into immature myeloid progenitor cells, which further differentiate into three types of mature myeloid cells: dendritic cells, macrophages and granulocytes. However, in the presence of soluble factors released by tumor and host cells in the tumor microenvironment, the production of MDSC is promoted.[79,80] Additionally, it has been demonstrated that

inflammation increases MDSC abundance and suppression activity, facilitating tumor progression.[4–7] Parker *et al.* recently described some of the pro-inflammatory tumor-released factors that can drive MDSC activation, expansion, accumulation and immune suppressor activity, which include: vascular endothelial growth factor (VEGF), granulocyte-macrophage colony-stimulating factor (GM-CSF), granulocyte colony-stimulating factor (G-CSF), prostaglandin E2 (PGE_2), interleukin-6 (IL-6), interleukin-1β (IL-1β), S100A8/A9 and high-mobility group box 1 (HMGB-1), among others.[1] Many of these tumor-released factors disturb normal cell functions by regulating transcription factors, such as the signal transducer and activator of transcription (STAT) 1, 3 and 6, and the nuclear factor–κB.[1,79] Moreover, miRNAs can also regulate MDSC production, accumulation and function, which will be discussed in more detail in Chapter 4.

As stated previously, MDSCs are able to suppress the adaptive and innate immune response by suppressing T cells and natural killer (NK) cells.[1] There are several mechanisms known by which MDSCs can suppress T cells (reviewed in detail [1,80,81]), namely: (1) amino acid starvation by local depletion of L-arginine, L-tryptophan and L-cysteine; (2) production of nitric oxide; (3) production of reactive oxygen species; (4) inhibition of T cell migration; (5) induction of apoptosis; (6) production of regulatory T cells and Th17 cells. Contact with MDSCs or close proximity appears necessary as suppression mechanisms are mediated by cell-surface receptors and labile soluble factors.[80,81] Additionally, exosomes released from MDSC have been found to mediate MDSC's ability to suppress the tumor immune response.[3]

Proteomic studies by Sinha *et al.* demonstrated that MDSCs from tumor-free and 4T1 mammary carcinoma tumor-bearing mice have N-glycan receptors that bind S100A8/A9 heterodimer.[82] The authors also found that S100A8 and S100A9 are pro-inflammatory mediators with chemotactic activity for MDSC, and are secreted by both 4T1 tumor cells and MDSCs, providing autocrine stimulation that helps maintain the pro-inflammatory microenvironment.[82] There are few reports that analyze the MDSC protein cargo by mass spectrometry. Boutte *et al.* found that pathways, such as platelet aggregation and angiogenesis, were enriched when comparing the protein content of MDSC from metastatic (4T1) and non-metastatic (67NR) mammary carcinoma tumor-bearing mice by shotgun proteomics and spectral counting.[83] Chornoguz *et al.* compared the protein cargo of MDSCs from 4T1 and 4T1/IL-1β tumor-bearing mice. Note that 4T1/IL-1β cells are transduced 4T1 cells that express IL-1β, a cytokine that intensifies inflammation. One biological replicate was analyzed and 789 proteins were identified and roughly quantified using peptide peak areas. Several pathways were found to be up-regulated under heightened inflammation, including the caspase network, Fas, TGF-β and IL-1 pathway.[84] Moreover, a semi-quantitative analysis of surface proteins on MDSC from 4T1 and 4T1/IL-1β tumor bearing mice by Choksawangkarn *et al.*, reported that S100A8 and S100A9 can be localized in the plasma membrane and are found in greater abundance under heightened inflammation.[85]

1.5. Exosomes shed by myeloid derived suppressor cells

Exosomes are membrane-bound extracellular vesicles of 30 to 100 nm in diameter that are secreted by all eukaryotic cells and many prokaryotic cells.[11] These extracellular vesicles are formed following the endocytic pathway, where intra-luminal vesicles (ILVs), which are formed by invagination of the membrane of early endosomes, accumulate in multi-vesicular bodies (MVBs). These MVB can have two fates: (1) degradation by fusing the lysosome, or (2) exocytosis by releasing the ILVs into the extracellular space, where they are called exosomes.[11,86] Extracellular vesicles are classically isolated from cell culture and biological fluids using centrifugation speeds of $\approx 100,000 \times g$, after removal of larger vesicles and cell debris at lower centrifugation speeds.[87] Additionally, materials pelleted by ultra-centrifugation may contain protein aggregates that co-isolate with exosomes and need to be removed using a sucrose density gradient.[88] After isolation, the exosome size and shape needs to be confirmed by transmission electron microscopy (TEM) or cryo-electron microscopy (cryo-EM). Compared to TEM, where exosomes often present an artifactual cup-shape morphology, cryo-EM maintains the exosome intact, and a round shape is clearly observed.[89] Exosomes are also often identified by proteins markers that are found to be enriched in exosomes, such as the tetraspanins[90] CD63, CD9 and CD81, heat shock proteins Hsc70 and Hsp90, major histocompatibility complex (MHC) class II, and proteins from the endosomal sorting complex required for transport (ESCRT), such as apoptosis-linked gene 2-interacting protein X (Alix) and tumor-suppressing gene 101 (Tsg101).[89,91,92] Recently, Bobrie *et al.* reported that many of these proposed protein

markers may actually be shared with other extracellular vesicle subpopulations that are inadvertently co-isolated by ultra-centrifugation.[93]

Interest on exosomes has increased during the past 10 years.[94] Even though exosomes were initially thought simply to carry unwanted molecules, extensive research has demonstrated that they actually carry bioactive proteins, lipids, mRNA and miRNA.[11,89,95] A broad number of biological roles have been suggested for these vesicles depending on the cell type from which they originate. Exosomes have been shown to participate in intracellular communication between B lymphocytes and CD4$^+$ T cells.[88] Human exosomes from dendritic cells showed *in vivo* tumor growth suppression.[96] Additionally, pathological functions such as immunity suppression, tumor progression, and metastasis have also been reported.[14,89,97] Therapeutic uses of exosomes including their use for delivery of drugs or other molecules to specific tissues have been discussed by others.[14,98]

MDSC from 4T1 and 4T1/IL-1β tumor-bearing mice release exosomes of 25 - 30 nm in diameter. These exosomes have been reported to play a role on MDSC's immune suppressive response, in part, through the pro-inflammatory mediators S100A8 and S100A9 that provide chemotactic activity for MDSC, and also by their participation on MDSC-macrophage cross-talk, polarizing macrophages to a tumor-promoting phenotype.[3] The MDSC-derived exosome protein cargo has been analyzed previously by shotgun proteomics and spectral counting. A total of 412 proteins were identified, and an increase in abundance for proteins related to innate immunity and

exosome migration was observed under heightened inflammation.[3] Further work by Burke *et al.*, demonstrated that exosomes released by MDSC from 4T1/IL-1β tumor-bearing mice carry ubiquitinated proteins, using a double immune-precipitation approach prior to LC-MS/MS analysis.[99]

The knowledge acquired on MDSCs and MDSC-derived exosomes by our research group and others during the past several years demonstrates that both MDSC and their exosomes play a key role in the tumor microenvironment, obstructing current immunotherapies and possibly providing a link between inflammation and cancer progression. However, there is still much of these mechanisms that remains unknown. Our research group in collaboration with Dr. Ostrand-Rosenberg at UMBC and Dr. Edwards at Georgetown University, have focused on increasing the understanding of MDSC immune suppression mechanisms by interrogating the protein contents and functions of MDSC and their exosomes. In this dissertation, efforts were made to study their protein content using both bottom-up and top-down mass spectrometry approaches. Additionally, working together with Dr. El-Sayed at UMD, the mRNA and miRNA contents were also studied using next generation sequencing. In all cases, relative quantitation was performed in order to provide a deeper understanding of quantitative differences between exosomes and their parental cells and also between conventional and heightened inflammation conditions. We anticipate that this information may help elucidate the role played by MDSC and MDSC-derived exosomes as immune suppressors in the tumor microenvironment and their relationship with increased inflammation, tumor progression and metastasis.

Chapter 2: Identification of intact proteins carried by immunosuppressive exosomes *(adapted from reference [100])*.

This work was jointly authored with Dr. Avantika Dhabaria, who performed the sample preparation, fractionation and LC-MS/MS analysis, and Lucia Geis Asteggiante, who performed the data processing and analysis.

2.1. Introduction

Exosomes are extracellular vesicles of 30 – 100 nm in diameter that are shed by many different types of cells and carry proteins, mRNA and miRNA.[11] Myeloid derived suppressor cells (MDSC) are immature myeloid cells that play an important role on immunity suppression in cancer patients.[1] These cells are known to shed exosomes that participate on MDSCs' suppressive activity.[3] Up to this time, proteomic analyses of the exosome protein cargo has been carried out by digesting proteins into peptides prior to LC-MS/MS. This approach, called bottom-up proteomics, has many advantages, as it provides a larger proteome coverage and high throughput.[17] However, when proteins are digested several pieces of information are lost. The study of intact proteins by top-down mass spectrometry offers relevant complementary information, such as identification of sequence variants and PTMs, including the determination of modification sites and their spatial relationship.[18] Full characterization of these different protein forms, called proteoforms, is crucial as little is known about their possible impact on MDSCs' activity.

Top-down proteomics of complex mixtures generally require extensive separation in order to enhance protein identification performance. Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) is a denaturing electrophoretic

approach where proteins are separated based on their molecular weight (3.5-150 KDa). The separation is carried out similarly to a sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE), but in this case proteins are actually eluted out of the gel into a chamber containing running buffer and equipped with a molecular weight cut-off filter that separates the collection and anode chamber.[101,102] The system is stopped at set time intervals and liquid fractions are manually collected from the chamber. As the gel columns used are short, fractionation is quick and diffusion and dilution of proteins is not significant.[101] Protein recoveries are reported to be >60%.[103] Several top-down analyses have successfully used GELFrEE.[100,104–107]

The protein cargo of exosomes shed by MDSC has been previously interrogated by bottom-up proteomics, where 412 proteins were identified.[3] The aim of this study is to identify and, when possible, fully characterize the low mass (< 25KDa) protein cargo by top-down mass spectrometry. MDSC-derived exosomes from 4T1/IL-1β mammary carcinoma tumor-bearing mice are lysed and their protein cargo fractionated using GELFrEE and reversed-phase liquid chromatography prior to high resolution, high accuracy MS/MS analysis.

2.2. Materials and methods

2.2.1. Exosomes shed from myeloid derived suppressor cells

BALB/c mice were injected with 7000 wild-type syngeneic 4T1 mammary carcinoma cells transduced to constitutively express the cytokine IL-1β, as previously

reported.[3,84] Populations of MDSCs collected from blood of 1 to 3 mice (5×10^6-10^7 cells per mouse) that were found to be > 90% Gr1$^+$CD11b$^+$ by flow cytometry, were incubated for 16 h in serum-free media at 37°C and 5% CO_2. Exosome isolation was performed following Burke *et al.*[3] Briefly, cultures were centrifuged first at 805×g for 5 min, supernatant collected and then centrifuged again at 2,090×g for 30 min in order to remove the cells. Exosomes in the supernatants were pelleted by ultracentrifugation at 100,000×g for 20h, purified by sucrose density gradient (0.25 to 2 M) and characterized by transmission electron microscopy. All animal experiments were approved by UMBC and UMCP Institutional Animal Care and Use Committees.

2.2.2. Exosome lysis and protein fractionation

Exosomes were lysed using an 8M urea in 50 mM ammonium bicarbonate buffer with protease inhibitor cocktail (Sigma-Aldrich. St. Louis, MO). A lysate aliquot was buffer exchanged using 3 KDa molecular weight cut-off filters, to reach a final concentration of 0.8 mM urea, prior to estimating the lysate protein content using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Rockford, IL). Note that exosomes shed from 1×10^8 cells contain approximately 100 μg of proteins. The exosome lysate was visually inspected by SDS-PAGE using a 15 μg lysate aliquot and stained with Commassie blue and silver stain (Thermo Fisher Scientific, Rockford, IL). Additionally, histone content was determined using the EpiQuik Total Histone Extraction Kit (Epigentek, Farmingdale, NY). All steps of the extraction kit protocol

were followed with the exception of sample reduction, as dithiothreitol (DTT) interferes with the BCA protein assay kit.

For GELFrEE separation, 300 µg of protein were precipitated using a mixture of $CHCl_3$-MeOH-H_2O (4:1:3, v/v)[108] and re-suspended in GELFrEE sample buffer (an SDS containing proprietary buffer). Prior to loading the sample into the 12% tris-acetate polyacrylamide cartridge (Expedon, San Diego, CA), proteins were reduced with 53 mM DTT at 50°C for 10 min. A total of 12 fractions were collected during the 2 h electrophoretic separation using the range of voltages (50 to 85 V) specified in the manufacturer's instructions. In order to remove the SDS from the sample, fractions were re-precipitated and re-suspended in solvent A (97.5:2.5 H_2O-acetonitrile in 0.1% formic acid) and then further separated using reversed-phase (C3 chemistry) liquid chromatography. GELFrEE separation efficiency was visualized by SDS-PAGE using 10 µL aliquots for each fraction and silver staining (Thermo Scientific, Rockford, IL).

2.2.3. Liquid chromatography mass spectrometry

Samples were analyzed using a Prominence LC system (Shimadzu, Columbia, MD) coupled to an LTQ-orbitrap XL (Thermo Fisher Scientific, Waltham, MA). Chromatographic separation was obtained using a C3 reversed-phase trap (0.3×5 mm, 5 µm particle size) and column (0.1×150 mm, 5 µm particle size), both obtained from Agilent Technologies (Wilmington, DE). The injected sample was loaded into the trap, desalted and concentrated using Solvent A at a flow rate of 10 µl/min for 15 min. The sample was subsequently eluted from the trap and further separated in the analytical

column by linearly increasing Solvent B (acetonitrile-H_2O, (97.5:2.5) in 0.1% formic acid) from 0 to 80%, during the span of 200 min under a flow rate of 300 nL/min.

The analysis of intact proteins by mass spectrometry was achieved setting a mass resolution of 60,000 (at m/z 400) for both precursor and product ions. Ions were collected based on a target automatic gain control of 10^6 and 10^5 for precursor and product ions, respectively. Additionally, in order to improve MS1 and MS2 signal to noise ratio, 5 microscans were averaged. Data dependent acquisition was carried out so that the 3 most abundant precursor ions that carried charges higher than +4 (including precursor ions with an undetermined number of charges) are isolated inside a 10m/z window. The isolated precursor ions were fragmented using CID set to 25% normalized collision energy. Dynamic exclusion was used in order to increase the variety of precursor ion selected by excluding precursor ions for future selection for 240s after being selected twice.

2.2.4. Bioinformatics

Spectra were deconvoluted using the THRASH algorithm[72] and searched against the Uniprot Knowledge Base *Mus musculus* protein sequence database (January and May 2014) using ProSight PC 3.0. In order to account for putative PTMs, a 2,500 Da precursor mass tolerance window and 15 ppm product ion mass delta were set. A protein was considered identified when an E-value $<10^{-5}$ was obtained. Additionally, the identity of putative PTMs and site locations were assessed manually using the SequenceGazer tool, requiring a mass difference < 15 ppm between the theoretical and

observed mass and an E-value $<10^{-5}$ in order to consider a proteoform as fully characterized.

2.3. Results and discussion

The exosomal protein cargo has been successfully interrogated previously by bottom-up analysis.[3] Based on protein identifications reported by Burke *et al.*, a histogram was created showing the exosome protein cargo molecular weight (MW) distribution (Figure 2.1). Considering that on the chromatographic time scale the LTQ orbitrap XL capability for intact protein analysis is restricted to proteins of MW $\leq$ 25 KDa, we could expect to find as many as 76 distinct proteins (18% of the identified proteins).
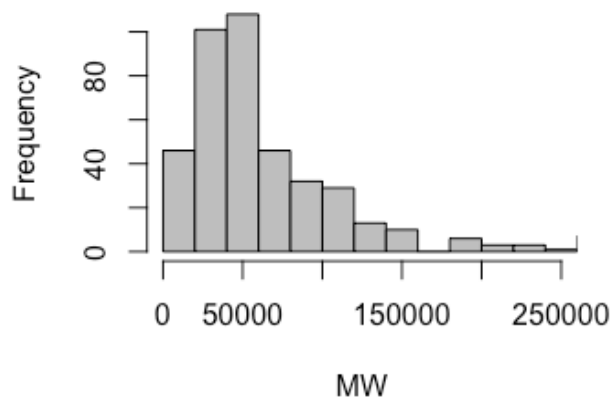
Figure 2.1. Molecular weight (MW) distribution of identified exosomal proteins based on findings reported by Burke *et al.*[3]

However, it is relevant to notice that top-down analysis is less sensitive than bottom-up analysis, due to the fact that larger ions are ionized and transferred less efficiently than peptides and that, as intact proteins carry multiple charges, the

28

precursor ion signal observed is diluted. Since GELFrEE fractionation follow the same principles as SDS-PAGE, an exosome lysate was initially visualized by SDS-PAGE and stained with both Commassie blue and silver stains, in order to obtain a rough estimate of protein abundance by MW range. Figure 2.2 shows that there is a higher abundance of proteins with MW ≤15-16KDa. Considering the previously reported bottom-up analysis, at least 36 distinct proteins could be present in this MW range.[3]

From the 12 fractions obtained by GELFrEE and further separated by reversed-phase (C3) fractionation, a total of 209 proteoforms from 21 distinct proteins were identified in the mass range of 5 to 16 KDa. Separation obtained by GELFrEE was helpful for identification purposes as many proteoforms were identified, but it was not very effective since in many cases the same proteoforms were found to be spread throughout many fractions (see Figure 2.3).
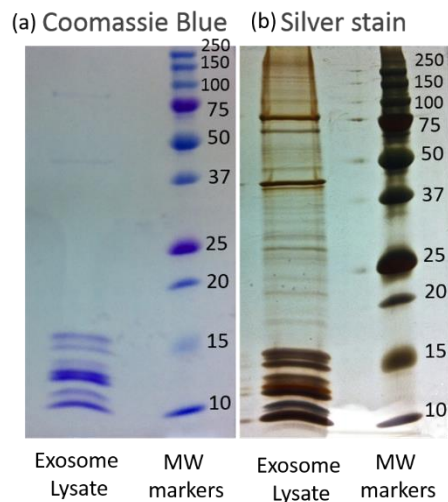
Figure 2.2. SDS-PAGE gel of a 15 μg aliquot of exosome lysate stained with (a) commassie blue and (b) silver stain. MW markers are shown for reference.
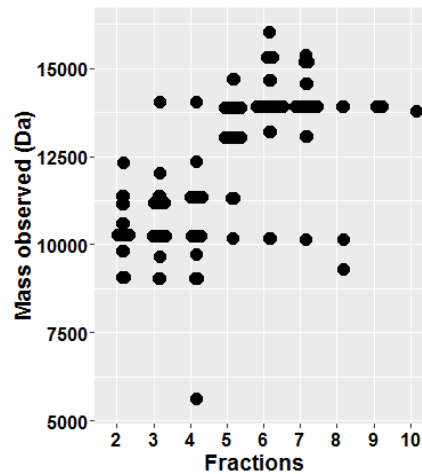
Figure 2.3. Masses observed (Da) for identified proteoforms by top-down mass spectrometry in each GELFrEE fraction collected.

29

Table 2.1 summarizes the identified (low mass) proteins, number of putative proteoforms and fraction number in which their proteoforms were found. A detailed list of proteins identified and proteoforms characterized is provided in Appendix 2.1, which includes theoretical and observed masses, E-values, number of ions in which the identification and characterization is based and, proposed putative PTMs.

Table 2.1. Distinct proteins identified by top-down mass spectrometry and number of putative proteoforms found.

Accession number	Protein description	Number of putative proteoforms identified	Fraction number observed
P09602	Non-histone chromosomal protein HMG-17	1	8
P47945	Metallothionein-4	1	4
P14069	S100 A6	1	2
P27005	S100 A8	42	2, 3, 4, 5, 6, 7, 8
P31725	S100 A9	1	3, 4, 5, 6, 7
P22752	Histone H2A.1	5	5, 6, 9
P27661	Histone H2A.X	1	7
Q6GSS7	Histone H2A.2A	2	6, 7, 9
Q8BFU2	Histone H2A.3	8	4, 5, 6, 7, 8, 9
P70696	Histone H2B.1A	1	3
Q64475	Histone H2B.1B	13	4, 5, 6, 7, 8, 9
Q6ZWY9	Histone H2B.1C	11	5, 6, 7, 8, 9
P10853	Histone H2B.1F	14	5, 6, 7, 8, 9, 10
Q64478	Histone H2B.1H	7	5, 6, 7, 8, 9
P10854	Histone H2B.1M	10	5, 6, 7, 8
Q64525	Histone H2B.2B	7	5, 6, 7, 8, 9
Q6LBF0	Histone H3.1	4	5, 6, 7
P84228	Histone H3.2	14	5, 6, 7
P84244	Histone H3.3	21	5, 6, 7
P02301	Histone H3.C	24	5, 6, 7
P62806	Histone H4	21	2, 3, 4, 5

2.3.1. Exosomes carry histone proteoforms

Several nucleic acid binding proteins were found to be carried by exosomes. Table 2.1 shows that 16 of the 21 proteins identified were sequence variants of the core histones H2A, H2B, H3 and H4. A total of 163 histone proteoforms were identified and only H3.1 (Q6LBF0), H3.3 (P84244), H3.C (P02301) and H4 (P62806) were also observed unmodified. Some of the proteoforms found were fully or partially characterized, considering the mass difference between the theoretical and observed masses and the fragmentation obtained. The loss of N-terminus Met, characterized by a mass difference of -131.04 Da, was found for all H2B variants except for H2B.1A (P70696). In the case of H2B.1A, only one proteoform was found, comprising the loss of -2203.20 Da. The loss of N-terminus Met with acetylation of the new N-terminus corresponding to a total mass difference of -89.02 Da (Figure 2.4a), was observed for H2A.1 (P22752), H2A.X (P27661) and H2A.2A (Q6GSS7). In the case of H2B.1C (Q6ZWY9), loss of initial Met and methylation of the new N-terminus Pro was observed (Figure 2.4b). H2B methylation of the N-terminus Pro has been observed in *Drosophila melanogaster* Kc cells under heat shock conditions.[109]

Interestingly, histones with proteolytically cleaved N-tails were also found. Using the SequenceGazer tool, the mass difference observed for histone H3.2 (P84228) and H3.3 (P84244; -2367.33 and -2294.28) could be assigned to the loss of 22 and 21 residues from the N-terminus, respectively. Their sequence and fragments observed are shown in Figure 2.4c,d. This process called "histone clipping" has been previously

reported for histone H3 in *Tetrahymena thermophila*,[110] mouse[111] and human[112]embryonic stem cells, human fibroblasts and melanocytes,[113] *S. cerevisiae*,[114] and human primary hepatocytes studied *in vivo* and hepatocarcinoma cells *in vitro*.[115] Several cellular processes seem to be regulated by histone proteolytic processing, including cell differentiation,[111,112] cell senescence,[113] gene transcription,[113,116] and yeast sporulation.[114] To the best of our knowledge, the proteolytic cleavage of histone H4 (P62806) N-tail has not been previously reported. Our work provided the full characterization of a H4 cleaved proteoform, with a mass difference of -2300.43 Da, that can be explained by the loss of the first 23 N-terminal amino acid residues. As shown in Figure 2.5, the use of top-down proteomics provided significant fragmentation (22 b-ions and 27 y-ions) that strongly supports the observation that exosomes shed by MDSC carry several histone proteoforms, including forms that had gone through a regioselective proteolytic process.

In the interest of estimating the contribution of histones to the lysate total protein content, a commercially available histone extraction kit indicated for isolation of histones from cell lysates and tissues was used. Based on this kit, it was estimated that ≈ 56% of the protein content corresponded to histones. However, later top-down analysis of the supernatants collected from this extraction procedure showed the additional presence of some S100A6 (P14069), S100A8 (P27005) and S100A9 (P31725). Hence, the previously reported percentage may overestimate the total amount of histones present in the exosome lysate.
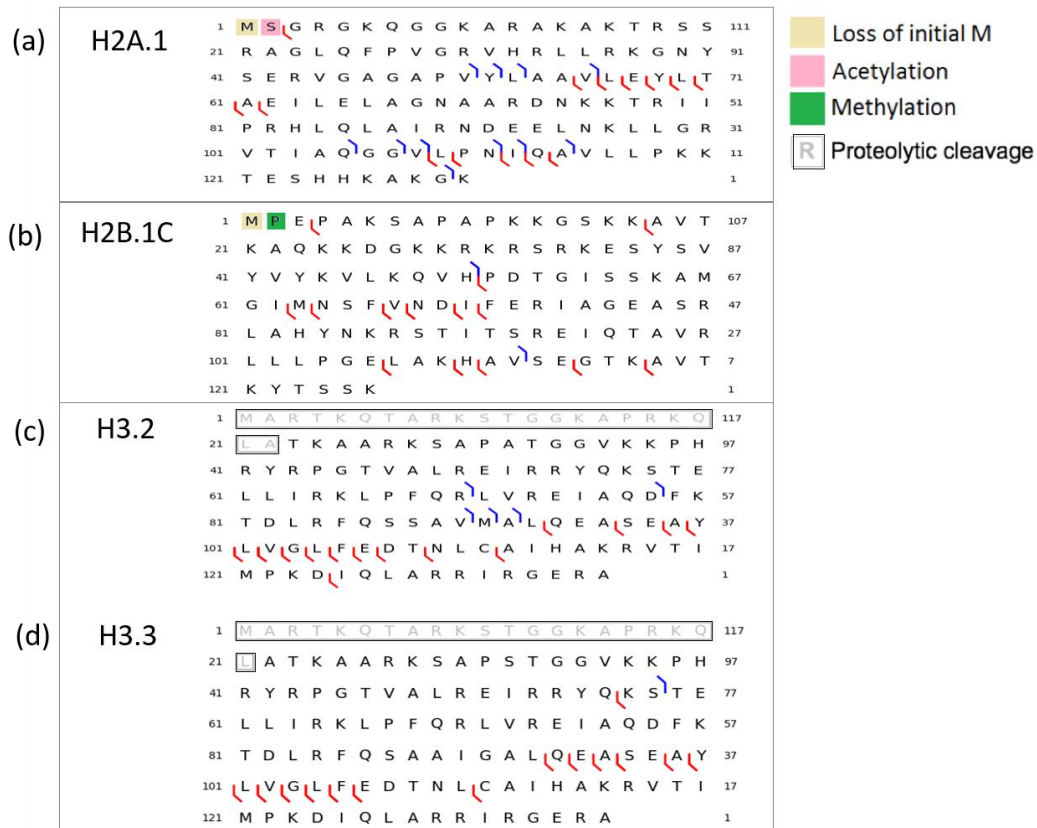
Figure 2.4. Sequence of characterized histone proteoforms (a) H2A.1 (P22752; 14,038 Da), (b) H2B.1C (Q6ZWY9; 13,780 Da) and the proteolytically cleaved histones (c) H3.2 (P84228; 13,011 Da) and (d) H3.3 (P84244; 13,024 Da). Fragment ions obtained by CID are shown in color, where b-ions are represented in blue and y-ions in red. (Reprinted with permission from Ref 100, Copyright 2014, *Int. J. Mass Spectrometry*. Elsevier B.V. DOI: 10.1016/j.ijms.2014.08.035).
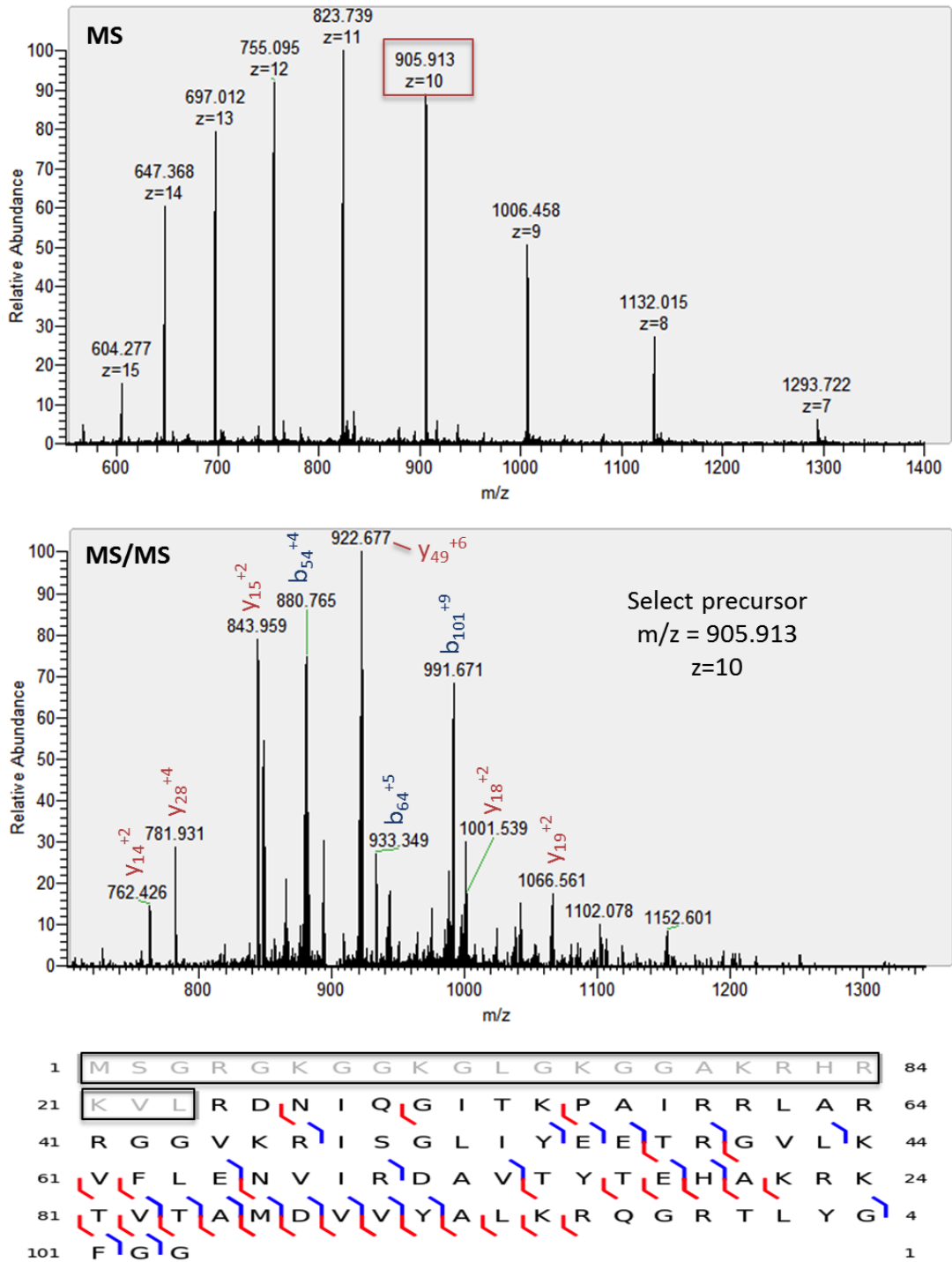
Figure 2.5. MS spectra corresponding to precursor ions of the proteolytically cleaved H4 proteoform (P62806; 9,044 Da), annotated MS/MS spectra and sequence. Fragment ions obtained by CID are shown in color, where b-ions are represented in blue and y-ions in red. (Reprinted with permission from Ref 100, Copyright 2014, *Int. J. Mass Spectrometry*. Elsevier B.V. DOI: 10.1016/j.ijms.2014.08.035).

2.3.2. Proteoforms of S100A8/A9 pro-inflammatory mediators in exosomes

The exosome lysate also contained 3 distinct proteins belonging to the S100 protein family. A total of 42 proteoforms of S100A8, and one proteoform each of S100A6 and S100A9 were observed (see details in Appendix 2.1). In the case of S100A6 (P14069; 9,955 Da), the proteoform observed presented the loss of N-terminus Met1 and acetylation of the new N-terminus Ala2 (Figure 2.6a). The characterized S100A9 (P31725; 12,965 Da) proteoform comprised the loss of N-terminus Met1, acetylation of Ala2 and methylation of His107 (Figure 2.6b). This finding supports previously reported S100A9 PTMs characterized by bottom-up analysis of GluC and LysC digests and the incorporation of tritium labeled methyl groups.[117] Raftery *et al.* also reported the presence of a disulfide-bond between Cys79 and Cys90, however, we do not expect to conserve disulfide bonds as samples are reduced prior to GELFrEE fractionation.
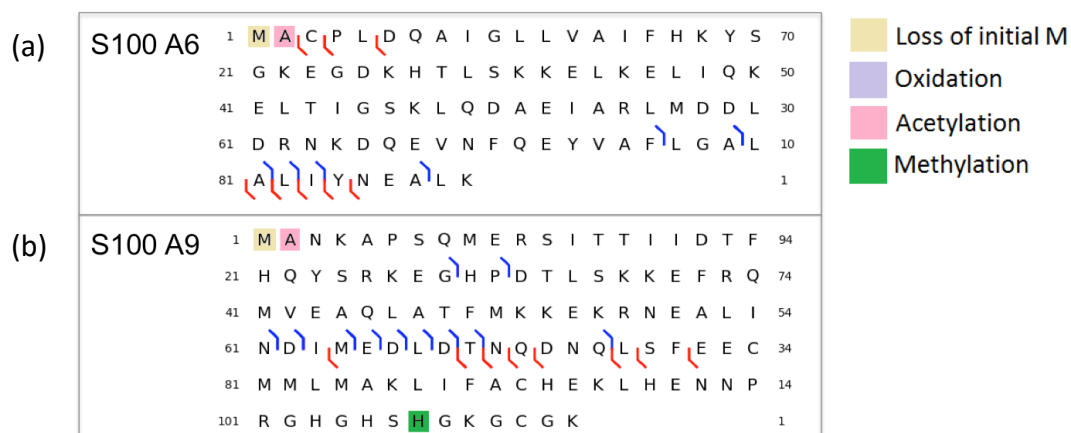
Figure 2.6. Sequence of characterized S100 family proteoforms of (a) S100A6 (P14069; 9,955 Da) and (b) S100A9 (P31725; 12,965 Da). Fragment ions obtained by CID are shown in color, where b-ions are represented in blue and y-ions in red. (Modified from Ref 100. Reprinted with permission, Copyright 2014, *Int. J. Mass Spectrometry*. Elsevier B.V. DOI: 10.1016/j.ijms.2014.08.035)

Several S100A8 proteoforms were well characterized and examples are shown in Figure 2.7a-e. The S100A8 (P27005; 10,157 Da) proteoform formed by the loss of N-terminus Met1 corresponded to the highest peak intensity signal of the non-fractionated lysate, and was estimated to represent 70% of the total ion chromatogram (TIC) intensity compiled by extracted ion chromatograms. The identification of Met1 acetylated S100A8 form (P27005; 10,330 Da) was supported by the fragmentation of the peptide bond between Met1-Pro2 (y_{88} ion) and Thr39-Thr40 (b_{39} ion) (see Figure 2.7b). An oxidized form of S100A8 (P27005; 10,173 Da) was identified. Based on the fragment ions available, there is enough evidence to clearly propose the loss of N-terminus Met1, as supported by the series of b ions observed, but oxidation could be located on any methionine or cysteine residue from position 15-44 (see Figure 2.7d). The S100A8 proteoform (P27005; 10,200 Da) comprises the loss of N-terminus Met1 and the putative acetylation of Pro2, though acetylation could be alternatively placed in any serine or lysine residue from position 2 to 39 (see Figure 2.7e). Moreover, unmodified S100A8 (P27005; 10,288 Da) was also observed. Modifications corresponding to additions of larger masses, ranging from 117.08 to 2,043.89 Da, were observed but the identity of those mass additions was not determined.
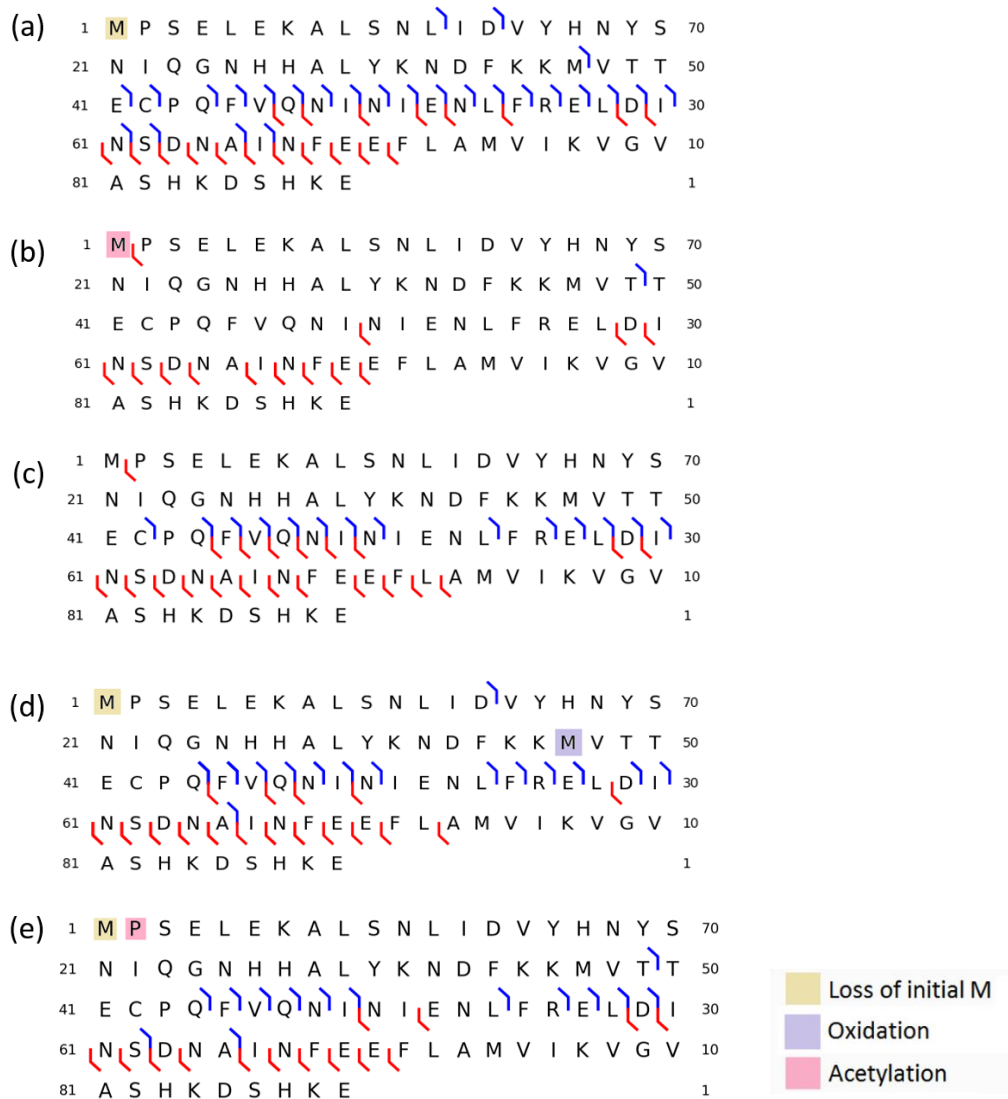
Figure 2.7. Sequence of characterized S100A8 (P27005) proteoforms (a) 10,157 Da, (b) 10,330 Da, (c) 10,288 Da, (d) 10,173 Da, (e) 10,200 Da. Fragment ions obtained by CID are shown in color, where b-ions are represented in blue and y-ions in red. (Modified from Ref 100. Reprinted with permission, Copyright 2014, *Int. J. Mass Spectrometry*. Elsevier B.V. DOI: 10.1016/j.ijms.2014.08.035)

S100A8 and S100A9 are pro-inflammatory mediators known to have chemotactic activity for MDSC and carried by the exosomes[3] and to be present in the cytosol and surface of MDSC.[84,85] Functional studies of S100A8/A9 are generally performed using antibodies that interact with different regions of these proteins. As this interaction could be affected by the presence of post-translational modifications, the characterization of S100A8/A9 proteoforms may add relevant information for future functional bioassays.

Even though a large number of proteoforms were identified in this study, it is clear that achieving full characterization is challenging. Fragmentation of intact proteins by CID in the LTQ orbitrap XL, under a chromatographic time scale, provides limited fragmentation. In this work, fragment ions were observed mostly in towards the C-terminus sections of the protein sequence, with only a few fragments observed on the first 30 N-terminal residues. The lack of fragment ions in the N-terminus region is especially problematic for histone proteoforms, as many of their PTMs are located there. Additional fragmentation using complementary approaches, such as ETD or EThcD, should help overcome this limitation and will be considered for upcoming studies.

2.4. Summary

   The low mass protein cargo of MDSC-derived exosomes was successfully interrogated using extensive protein fractionation and top-down mass spectrometry. GELFrEE coupled to reversed-phase fractionation provided a large number of protein identifications, though significant overlap of proteins was observed between adjacent fractions. More than 200 proteoforms from 21 distinct proteins were identified and several of them characterized.

   Forty-three proteoforms corresponded to S100A8 and S100A9, which are chemoattractants relevant to MDSC's activity. Additionally, more than 160 histone proteoforms corresponding to 16 histone variants were also identified. Even though histones represent a large part of the exosome cargo, there is no known function associated to their presence in exosomes. Perhaps they play a role in supporting the RNA cargo carried in exosomes, and/or upon delivery of that cargo to other cells present in the tumor microenvironment. This increased knowledge about the diversity of proteoforms carried by MDSC-derived exosomes should help better understand how MDSC produces immune suppression and promotes tumor progression.

Chapter 3: Evaluation of spectral counting for proteoform quantitation in top-down proteomics *(adapted from reference [118])*.

3.1. Introduction

The study of intact proteins by top-down mass spectrometry provides the identification of proteoforms, which are different forms of a protein due to single nucleotide polymorphisms, alternative mRNA splicing, PTMs, and truncations.[119,120] These proteoforms may have varied regulatory functions, controlling processes such as gene expression, cell cycle, proliferation, differentiation and cell-cell communication.[8] Hence, the ability to fully characterize and quantify proteoforms is crucial to better understand biological processes.

Quantitation strategies for intact proteins are still under development. The applicability of stable-isotope labelling for proteoform quantitation was initially demonstrated using a $^{14}N/^{15}N$ metabolic label in yeast,[121] and Stable Isotope Labeling by Amino acids in Cell culture (SILAC) for the *in vitro* expressed growth factor receptor-bound protein 2 (28KDa).[122] Collier *et al.*, later quantified the proteome of human embryonic stem cells using SILAC.[123] Even though these metabolic labeling strategies have provided fruitful results, their elevated costs and restriction to *in vitro* studies and animal models limit their use.[124] Moreover, Hung and Tholey evaluated the use of tandem mass tag (TMT), where the proteins of interest are labeled before LC-MS/MS analysis and reported that the lack of complete labeling limits this approach.[125]

To this date, the only label-free quantitation approach broadly applied to top-down proteomics uses precursor ions (MS1) integrated peak areas or peak intensities from extracted ion chromatograms (EIC). Even here, there is a lack of consensus on which signal should be used for quantitation. Mazur *et al.* summed the peak areas obtained from the 5 most abundant protein charge states EIC to show differences in abundance in human apolipoproteins.[126] Peak intensities of the 5 most intense isotopomers normalized to the overall peak intensity were used to estimate the percentage of phosphorylated cardiac troponin I proteoforms.[77] Similarly, Chen *et al.* estimated the percentage of phosphorylated H1 proteoforms, though in this case only the peak intensity of the most abundant isotopomer was used.[127] The use of EIC peak areas, considering all the observable protein charge states, was observed in a study on human saliva where the abundances of 83 proteoforms produced by two different glands (parotid and submandibular/sublingual gland) were compared.[128] In this case, normalization was performed against the total ion chromatogram (TIC) maximum intensity.[128] The Kelleher group created a pipeline termed "Quantitative Mass Targets" for top-down comparative analysis, which uses deisotoped precursor intensities normalized to the TIC maximum intensity.[76,129,130] Currently, automatic tools for peak area/intensity estimation are not available, limiting the capability for high-troughput analysis.

Discovery-based proteomics is generally performed in a data-dependent manner, where the *n* most abundant precursor ions are selected for fragmentation. Spectral counting is widely used for relative peptide quantitation and is based on the

premise that abundant precursor ions have a higher probability and frequency of being selected for fragmentation, and are more probable to provide a positive identification based on the produced fragment ions.[56–59] Hence, a correlation between the number of counts of peptide identifications or peptide spectrum matches (PSM) and the protein abundance can be suggested. This approach is quite easy to implement as database search engines provide the number of identified peptides during data processing. For various reasons it has not yet been applied to top-down proteomics.

This study aims to evaluate the applicability of spectral counting to top-down proteomics workflows. We provide proof of concept that spectral counts can be used to determine proteoform differential abundances by performing spiking experiments and comparing the performance of spectral counting, normalized peak areas and normalized peak intensities. Furthermore, all three approaches were applied to the protein cargo from MDSC-derived exosomes produced under conventional and heightened inflammation conditions.

3.2. Materials and methods

3.2.1. Exosomes shed from myeloid derived suppressor cells

BALB/c mice were injected with 7000 wild-type syngeneic 4T1 mammary carcinoma cells or 4T1 cells transduced to constitutively express the cytokine IL-1β (4T1/IL−1β), as previously reported.[3,84] Populations of MDSCs collected from blood of 1 to 3 mice (5×10^6-10^7 cells per mouse) were confirmed to be > 90% Gr1$^+$

CD11b$^+$ by flow cytometry, and were incubated 16 h in serum-free media at 37°C and 5% CO_2. Exosomes were isolated by ultracentrifugation and characterized by sucrose density gradient and transmission electron microscopy as previously reported by Burke *et al.*[3] In this study, exosomes shed from MDSCs from 4T1 and 4T1/IL-1β tumor-bearing mice are termed "conventional" and "inflammatory", respectively. All animal experiments were approved by UMBC and UMCP Institutional Animal Care and Use Committees.

Exosomes samples were prepared following Burke *et al.* with modifications.[3] Lysis was performed using 8M urea in 50 mM ammonium bicarbonate buffer with protease inhibitor and phosphatase inhibitor number 2 and 3 cocktails from Sigma-Aldrich (St. Louis, MO) and deacetylase inhibitor cocktail from Santa Cruz Biotechnologies (Dallas, TX). Lysates were buffer exchanged to reach a final concentration of 8 mM urea, using 3 KDa molecular weight cut-off filters from Millipore (Cork, Ireland). Protein content was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Rockford, IL). All lysates were prepared in solvent A (97.5:2.5 H_2O-acetonitrile in 0.1% formic acid) and analyzed by LC-MS/MS. Four biological replicates for each condition, were each analyzed in triplicates.

3.2.2. Preparation of spiked exosome lysates

Exosome lysates containing 500 ng of protein were spiked with 5 protein standards as specified in Table 3.1, in order to reach a total protein amount of 680 ng for subsequent LC-MS/MS analysis. The protein standards used were: ribonuclease A

from bovine pancreas (RNase, 13.7KDa), equine apomyoglobin (16.9KDa), and carbonic anhydrase (CAH, 29KDa), all obtained from Sigma Aldrich and, recombinant human tumor necrosis factor alpha (TNF-α, 17.3KDa) and apolipoprotein D (20.3KDa) from Novoprotein (Summit, NJ).

Table 3.1. Description of spiked standard proteins into eight aliquots of an exosome lysate.[118]

Lysate (ng)	TNF-α (ng)	CAH (ng)	RNase (ng)	Apomyoglobin (ng)	Apolipoprotein D (ng)
500	40	10	50	60	20
500	20	50	10	40	60
500	60	-	60	20	40
500	50	20	40	60	10
500	50	40	20	10	60
500	10	40	20	50	60
500	5	100	5	70	-
500	1	100	1	78	-

3.2.3. Liquid chromatography mass spectrometry

Samples were analyzed using an Ultimate 3000 RSLCnano system (Dionex, Sunnyvale, CA) coupled to an orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific, Waltham, MA). Chromatographic separation was obtained using a PepSwift RP-4H monolith trap (0.1 × 5 mm) and ProSwift RP-4H column (0.2 × 250 mm), both obtained from Thermo Fisher Scientific. Spiked samples were loaded into the trap, desalted and concentrated using Solvent A at a flow rate of 5 μl/min for 5 min, and then eluted from the trap and further separated in the analytical column by linearly increasing Solvent B (acetonitrile-H_2O, (75:25) in 0.1% formic acid) from 5 to 55%, during the span of 70 min under a flow rate of 1.5 μL/min. In the case of exosome lysates, samples were desalted and concentrated for 10 min and then separated using a linear gradient from 1 to 55% B for 145 min.

The analysis of intact proteins by mass spectrometry was achieved setting a mass resolution of 120,000 (at m/z 200) for both precursor and product ions. Ion routing multipole pressure was set to 3mTorr. Ions were collected based on a target automatic gain control of 10^6 and $5x10^5$ for precursor and product ions, respectively. The maximum injection time for both precursors and product ions was set to 200ms. Additionally, in order to improve MS1 and MS2 signal to noise ratios, 5 microscans were averaged for spiked samples and 2 microscans for exosome samples. Data dependent acquisition was carried out in a fixed 10s duty cycle, in which the top n most abundant precursor ions that carry charges higher than +4 (including precursor ions with undetermined charge-states) are isolated by the quadrupole within a 1 m/z isolation window. The isolated precursor ions were fragmented using CID set to 25% normalized collision energy, or EThcD with 6 ms reaction time for ETD and 10% supplemental activation with HCD. Dynamic exclusion (DE) was used in order to increase the variety of precursor ions selected. The effect of DE was assessed using spiked samples. In the case of exosome samples, precursor ions selected were excluded for 60s after selection.

3.2.4. Bioinformatics and relative quantitation

Spectra were searched against the *Mus musculus* database (Uniprot KnowledgeBase, Oct. 2015) and a decoy database (created by reversing the mouse database protein sequences) using the ProSightPD node 1.0 in Proteome Discoverer 2.1. In the case of spiked samples, the protein sequences of the standard proteins used were manually added into the mouse database. The search parameters included

precursor and product ion mass tolerance windows of 2.5 KDa and 15 ppm, respectively. A protein/proteoform was considered identified if a protein spectrum match (PrSM) had an estimated E-value $\leq 10^{-5}$ and false discovery rate $\leq 1\%$.

The total number of spectral counts were estimated based on the number of times a proteoform was identified by counting the protein spectrum matches (PrSMs) provided in ProSightPD search output both manually and using an in-house script. Proteoform peak intensities and areas were manually estimated in Xcalibur from deconvoluted extracted ion chromatograms. A minimum signal-to-noise (S/N) ratio of 3 was required for spectra deconvolution. Additionally, in order to improve quantitation precision and accuracy,[76] the obtained peak intensities and areas were normalized by the total ion chromatogram maximum intensity and total area, respectively.

The evaluation of intact protein relative quantitation using spectral counts, normalized peak areas and normalized peak intensities in top-down proteomics entailed: (1) the initial confirmation that those metrics present a linear relationship with the amount of standard protein spiked; (2) the comparison between the observed protein ratios and expected ratios calculated using the known amounts of protein standard spiked; (3) the estimation of statistical significance of differential abundance between two samples, using Fisher's Exact test for differential spectral counts and Student's *t*-test for comparison of means from normalized peak areas and intensities. In all cases, the obtained p-values were corrected for multiple testing by estimating the Benjamini-Hochberg FDR.[49]

The total number of spectral counts obtained from MDSC-derived exosome lysates were 3-fold lower for exosomes shed under conventional conditions compared to those under heightened inflammation. Hence, a global adjustment was performed to force centering the median protein ratio to 1. Median correction was also performed for peak areas and intensities. When comparing the proteoform corrected spectral counts observed between inflammatory conditions, statistical significance was evaluated using Fisher's Exact and χ^2 tests,[131] and a sample-label permutation test with a maximum of 100,000 trials per proteoform. Note that using an early stopping rule the number of trials were reduced.[132] In the case of normalized peak areas and intensities, statistical significance was assessed using the Mann-Whitney test, which is less sensitive to the presence of outliers. As previously stated, Benjamini-Hochberg FDR values were also estimated. Excel and/or R programs were used for performing calculations, statistical analysis and creation of graphs.

3.3. Results and discussion

3.3.1. Effect of dynamic exclusion and spectra averaging on spectral counting

Data-dependent acquisition is the preferred mode for proteomics discovery studies. This acquisition mode provides higher protein/proteoform coverage as it allows a more varied selection of precursor ions using the mass spectrometer dynamic exclusion (DE) feature. Dynamic exclusion aims to avoid the repeated selection of highly abundant precursor ions by excluding previously selected *m/z* for a user-defined

period of time. In top-down proteomic workflows, precursor ions are generally excluded for long periods of time, in order to increase the number of protein/proteoform identifications. However, the increase in precursor ion variety is achieved at the cost of reducing the likelihood of obtaining repeated measurements,[133] which reduces the statistical power of spectral counting. Hence, DE is an important factor to consider and needs to be evaluated.

A set of 5 standard proteins were used for DE optimization and spectral counting assessment. The identity and quality of the standard proteins was evaluated by LC-MS/MS, and 4 out of the 5 proteins were successfully characterized as individual proteoforms. Apolipoprotein D was not identified in any of the mixtures. Additionally, the identification of apomyoglobin in spiked samples was severely hindered by co-eluting high intensity S100A8 proteoforms. Hence, three standard proteins, CAH, TNF-α and RNase, were found suitable for further analysis.

The effect of DE on spectral counts was evaluated by spiking an exosome lysate (500ng) with 50ng of three protein standards and analyzing the sample in triplicates using the DE settings shown in Table 3.2. Note that the largest exclusion period selected (200s) represents typical DE parameters used in discovery top-down analyses and is longer than the observed LC peak widths of up to 180s.

Table 3.2. Dynamic exclusion settings evaluated.

Dynamic exclusion setting #	Number of times selected in a 30s period	Exclusion window (s)
1	1	200
2	1	60
3	2	60

The results observed for the DE settings evaluated are shown in Figure 3.1. Compared to the typical DE setting used for top-down proteomics, highlighted in grey in Figure 3.1, the reduction of exclusion time to 60s provided a higher number of spectral counts per proteoform, as repeated selection of precursor ions (m/z) is allowed. However, requiring the exclusion of a precursor ion only after the ion was selected twice in a 30s period did not provided a higher number of counts. At least 10 counts were observed for RNase and TNF-α. The relative standard deviation of counts ranged between 2 - 8% and 7 - 76% for RNase and TNF-α, respectively. The variability observed for TNF-α can be explained by the presence of co-eluting S100A9 proteoforms from the exosome lysate. Significantly fewer counts were observed for CAH, which is expected since the identification and quantitation of larger molecules on the chromatographic time scale is limited by current instrumental ionization and detection capabilities. Lower signal to noise ratios (S/N) are generally observed due to ESI production of multiply charged species, and the reduced ion transfer and fragmentation efficiency.[134] Additionally, larger proteins compete for ionization against smaller easier to ionize proteins present in the complex sample. The effect of DE was also evaluated for a set of 5 proteoforms present in the exosome lysate, which were selected in order to represent a broad range of normalized intensities (0.2 – 37% of the maximum TIC intensity). Independently of their signal intensities, DE showed a similar effect on spectral counts to that observed for the standard proteins.
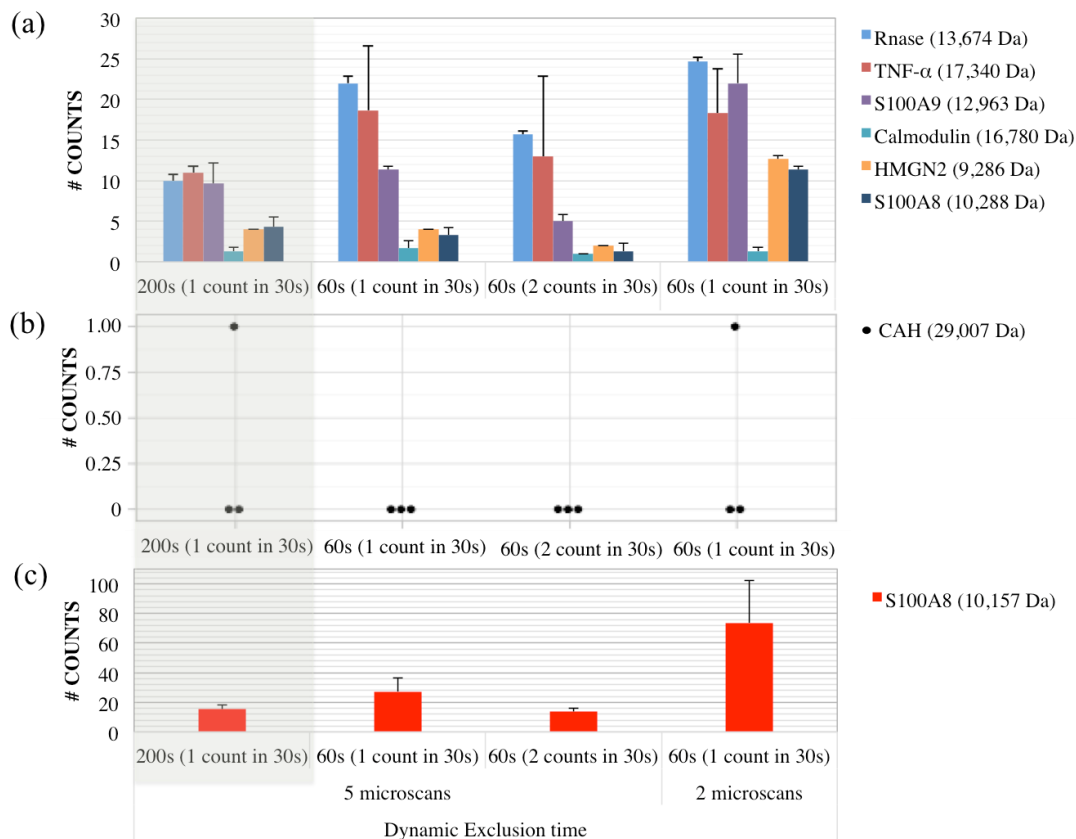
Figure 3.1. Effect of dynamic exclusion and microscans averaging on PrSMs counts of (a) 2 protein standards spiked (50 ng) and 4 proteoforms found in the exosome lysate (representing in average 0.2 – 4.2% of the TIC intensity), (b) CAH standard spiked, and (c) S100A8 (10,157 Da) proteoform, which represents in average 37% of the TIC intensity. Error bars represent the standard deviation of the mean. Typically used dynamic exclusion settings for top-down proteomics are highlighted in light gray.[118]

An additional parameter that significantly affects the application of spectral counting in top-down proteomics is the number of averaged microscans used. The process of injecting, storing and detecting an ion in the orbitrap is called a microscan. For intact protein analysis, 5 or more microscans are generally averaged in order to increase the S/N of precursor and product ions, as high-resolution and good quality spectra are needed to determine their monoisotopic masses. This results in long duty cycles, reduced ion sampling rates and reduced number of proteoforms identified.

50

Nevertheless, considering the higher speed rates and fragmentation efficiencies provided by the state-of-the-art orbitrap Fusion Lumos,[30] the use of fewer microscans was evaluated. A fixed duty cycle of 10s and DE of 60s for precursor ions that were selected once in a 30s window were set, and precursor and product ions acquired averaging 2 and 5 microscans. A 2.5-fold increase in the number of MS/MS acquired was observed when 2 microscans were averaged (7-8 MS/MS) compared to 5 microscans (2-3 MS/MS). Additionally, an increase in the number of spectral counts was also observed for most proteoforms (see Figure 3.1). In this study, spectral counting will be ultimately applied to perform relative quantitation of the MDSC-derived exosome protein cargo under two inflammation conditions. Hence, using DE set to exclude precursor ions for 60s after been selected once and the averaging of 2 microscans will provide the highest number of proteoform identifications.

3.3.2. Exploring spectral counting applicability to top-down proteomics

The evaluation of spectral counting for relative quantitation of proteoforms was performed using standard proteins spiked into a complex sample at various amounts and, comparing its performance with currently used label-free quantitation approaches: normalized peak intensities and normalized peak areas. This process entailed: (1) determining the relationship between the spiked amount and signal measured; (2) exploring the capabilities of each approach for detecting differential abundances; and (3) determining the accuracy of each approach to estimate protein ratios. For this purpose, 8 aliquots of an exosome lysate were spiked with varied amounts of standard proteins (see Table 3.1), and analyzed in triplicates by LC-MS/MS. Proteoforms were

identified in ProSightPD by searching against a modified Uniprot mouse protein sequence database which included the protein sequences of the three standard proteins spiked. Spectral counts were obtained by manually counting those PrSM that passed the identification criteria stated in the "Materials and Methods" section. All acquired spectra were deconvoluted. EIC peak areas and intensities were manually estimated and normalized by TIC maximum area/intensity, usually showing relative standard deviations $\leq 25\%$.

Depending on the approach under evaluation, the signal measured (spectral counts, normalized peak areas or normalized peak intensities) was plotted against the standard protein amounts added to the sample. Figure 3.2a-c shows that irrespective of the signal measured, a positive linear relationship was observed for the three standard proteins spiked, with increasing measured signal as the amount spiked increases. Additionally, even though all three approaches fit a linear model (R^2 ranging from 0.66 to 0.90), normalized peak intensities presented the highest R^2 for all 3 standard proteins (Figure 3.2b).

The ability of each quantitation approach to detect statistically significant differences in abundance was explored using spiked samples analyzed in triplicates. Expected protein ratios were calculated for each standard protein as the ratio of pairs of known spiked amounts and the statistical significance of the differential abundances determined (see Figure 3.3a-c). The number of false positives and false negatives were also estimated. In this study, a false negative (FN) is defined as the case where a

difference in abundance goes undetected (adjusted p-value > 0.05), even though the known protein ratio is ≥ 2 (highlighted in light red in Figure 3.3). Conversely, a false positive (FP) is defined as the case where difference in abundance is detected (adjusted p-value ≤ 0.05), but the known protein ratio is < 2 (highlighted in light blue in Figure 3.3).
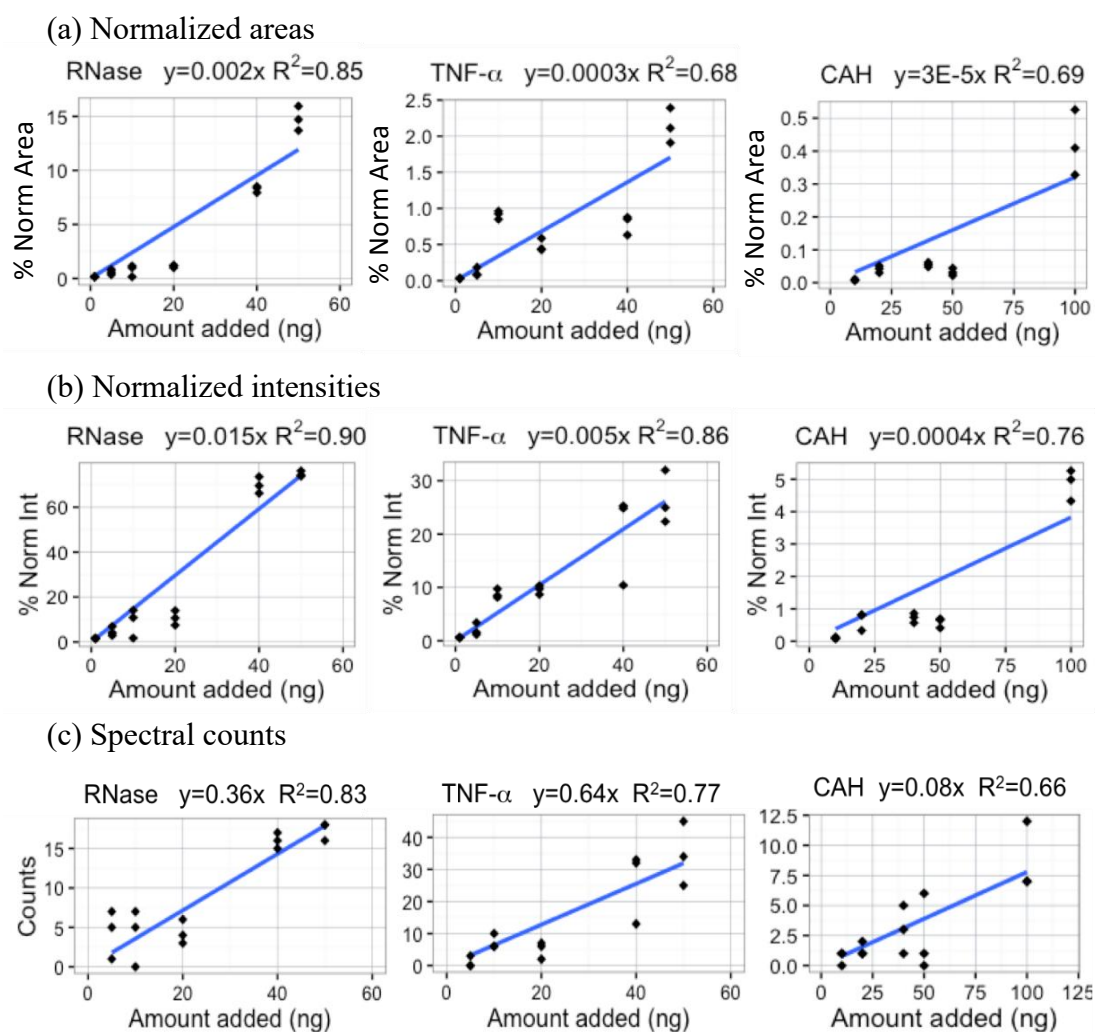
Figure 3.2. Relationship observed between (a) normalized area, (b) normalized intensity, (c) spectral counts, and the amount of standard protein spiked into the exosome lysate.[118]

In the case of spectral counts, differences in abundance were deemed significant if the Fisher's Exact test p-value, adjusted for multiple testing by Benjamini-Hochberg, was ≤0.05. As shown in Figure 3.3a, differences in abundance were often detected for RNase and TNF-α for known protein ratios ≥ 2 and, CAH for protein ratios ≥ 3. The percentage of FN observed for TNF- α, RNase and CAH were 18%, 27% and 63%, respectively. Additionally, FP were only observed for TNF- α for known protein ratios < 2. As stated above, it is expected that larger proteins, such as CAH, will produce only a moderate number of counts. When determining the significance of differences in abundance, statistical tests such as Fisher's Exact and χ^2 tests, which consider the number of counts observed as independent observations, tend to overestimate significance. Nevertheless, the overestimation did not seem to greatly affect marginal p-values, as supported by the few FP observed.

In the case of normalized peak intensities and areas, the statistical significance of differences in abundance was estimated using a Student's t-test with Benjamini-Hochberg correction. Figure 3.3b shows a large proportion of FN observed when using normalized peak intensities, with 50%, 33% and 36% FNs for TNF- α, RNase and CAH, respectively. Additionally, no FP were observed for this approach. Normalized peak areas performed similarly to spectral counting, as the percentage of FNs observed were 13%, 31% and 72% for TNF- α, RNase and CAH, respectively (see Figure 3.3c). However, a FP was found for both TNF-α and RNase. Hence, in this study spectral counting showed comparable or higher sensitivity than normalized peak areas and intensities. This observation was unexpected as quantitation using peak areas or

intensities is generally consider to be robust. Increasing the number of replicates could

help reduce the variability observed and provide better sensitivity.
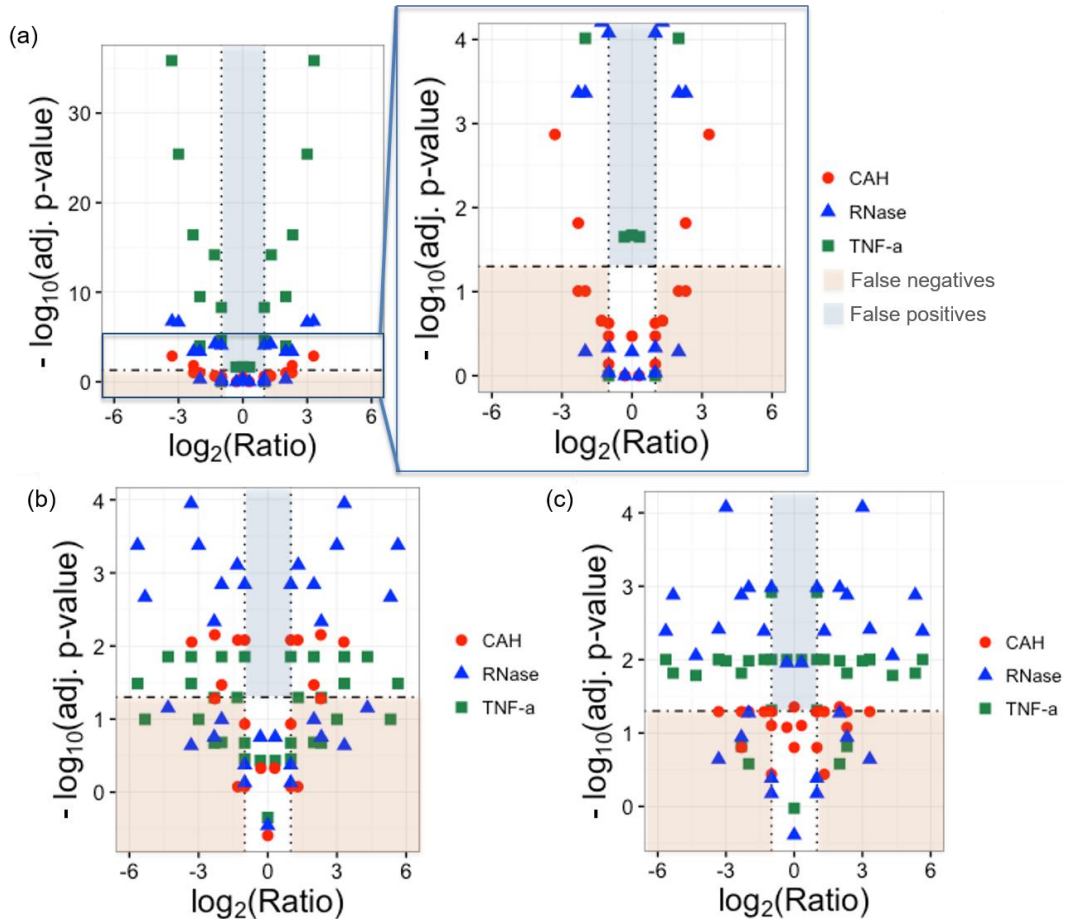
Figure 3.3. Volcano plots (-$\log_{10}$(p-value) vs. expected protein ratio) showing the ability of (a) spectral counting, (b) normalized peak intensities, and (c) normalized peak areas to detect differences in abundance. The horizontal dashed line represents the adjusted p-value threshold of 0.05 (-$\log_{10}$(p-value) ≥ 1.3). Differences in abundance of two-fold are marked by dotted vertical lines. An re-scaled plot was added for spectral counting in order to be easily compared to the other techniques. In this study, false positives (highlighted in light blue) are defined as cases with known spiked protein ratios < 2 ($\log_2$(Ratio) between -1 and 1), with adjusted p-values below 0.05 (-$\log_{10}$(p-value) > 1.3). False negatives (highlighted in light red) are cases with known spiked protein ratios > 2 ($\log_2$(Ratio) < -1 and $\log_2$(Ratio) > 1) and adjusted p-values above 0.05 (-$\log_{10}$(p-value) < 1.3).[118]

Protein ratios were estimated for the three approaches evaluated and plotted against the expected known protein ratios after performing a $\log_2$ transformation (Figure 3.4a-c; additional plots without $\log_2$ transformation are provided in Appendix 3.1). A linear relationship was observed ($R^2 > 0.5$) in all cases, except for protein ratios estimated for RNase by spectral counting, which presented a weak correlation ($R^2 = 0.34$, Figure 3.4a). The moderate R^2 values observed are expected due to the variability observed at low spiking levels. This observation is even more noteworthy for spectral counting, because as the amount of standard protein spiked into the complex sample decreases, co-eluting proteins become more prominent reducing the likelihood of identifying the spiked protein. The ability of each approach to estimate protein ratios was evaluated based on the linear regression slopes observed, where a slope equal to 1 represents an accurate estimate. Normalized peak areas and peak intensities, provided more accurate estimates, as supported by the observed correlation slopes ranging from 1.00 to 1.43 (Figure 3.4b-c). In the case of spectral counting, protein ratio estimates obtained were less accurate, ranging from 0.68 – 1.40, but still followed the expected trend. Spectral counting can therefore be used to estimate protein ratios.
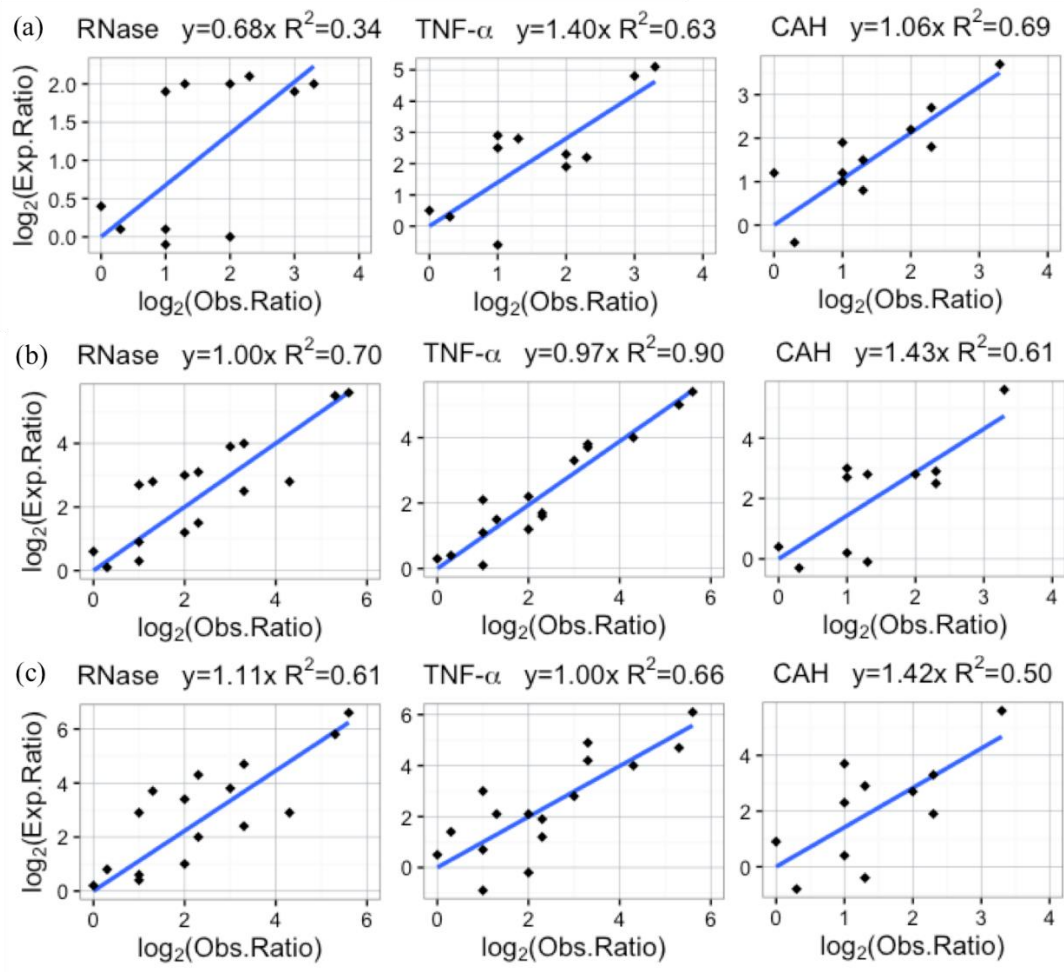
Figure 3.4. Relationship between expected and observed protein ratios estimated by (a) spectral counting, (b) normalized intensities and (c) normalized areas.[118]

3.3.3. Spectral counting applied to the immunosuppressive exosome protein cargo

The next step in this study was to apply all three previously evaluated approaches to quantitatively study the proteoform cargo of exosomes shed by myeloid-derived suppressor cells (MDSCs) collected from tumor-bearing mice with heightened and conventional inflammation. For each condition, 4 biological replicates were analyzed in triplicate as stated in the "Materials and Methods" section. Protein

identifications and spectral counts were based on the combined search results for each inflammation condition. A total of 20 distinct proteins were identified. Fourteen distinct proteins with a total of 52 proteoforms were found in conventional exosomes, and 20 distinct proteins with a total of 70 proteoforms were found in inflammatory exosomes. A total ion chromatogram showing protein fractionation can be found in Figure 3.5. The present study identified fewer proteoforms when compared to our previous interrogation of the exosomal protein cargo under heightened inflammation conditions, where more than 200 proteoforms of the core histones and S100 family proteins were identified.[100] This observation is not surprising as the earlier study had included fractionation on a GELFrEE apparatus. Nevertheless, it is relevant to note that in the present work, an additional 12 distinct proteins corresponding to 29 proteoforms were identified (e.g. Figure 3.6a-d,h). Additionally, extensive fragmentation within the N-tail of histone H4 (P62906) allowed confirmation of proteoforms previously reported[100] (e.g. Figure 3.6e,g) and characterization of additional proteoforms (e.g. Figure 3.6f). Using the orbitrap Fusion Lumos, proteoform characterization was often straightforward due to faster acquisition speeds and availability of both CID and EThcD fragmentation. Depending on the proteoform, residue cleavage ranged from 13 to 97% based on the combined spectra of one precursor ion fragmented by both EThcD and CID (see Appendix 3.2). The complete list of proteoform identifications including putative PTMs is shown in Appendix 3.2, and a selection of sixteen fully characterized proteoforms are shown in Figure 3.6. Several proteoforms of the biologically active pro-inflammatory mediators S100A8 and S100A9 were also identified. Characterizing their proteoforms is relevant as these proteins are chemotactic for MDSC and are

present in high abundance in MDSC and their exosomes.[3,82,100] However, to the best of our knowledge, there is currently no information linking one or more distinct proteoforms to their chemotactic activity or linking increased inflammation to changes in the relative abundances of individual proteoforms.
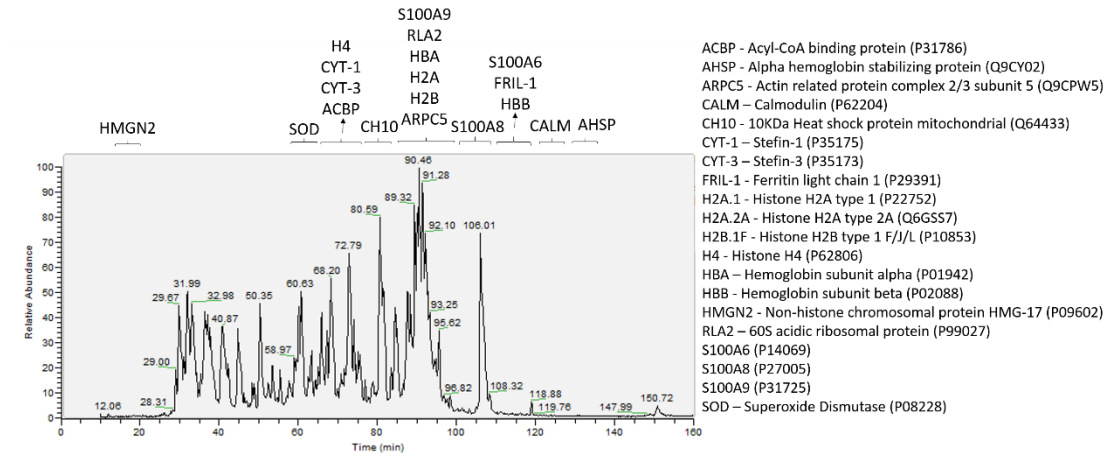
Figure 3.5. Total ion chromatogram of an exosome lysate showing the separation of distinct proteins in a monolith RP-4H column.

The total number of spectral counts was estimated for each inflammation condition and compared to evaluate possible differences in sampling depth. Inflammatory exosomes presented a higher total number of counts (6434 PrSMs) when compared to conventional exosomes (2131 PrSMs), even though the same amount of lysate was injected. Normalization was therefore needed to correct for the bias observed. Assuming that systematic bias is not the major reason for the differences observed and that most of the proteoforms identified show no differences in abundance, a global median normalization can be performed. This normalization approach re-

59

scales the number of spectral counts measured for the inflammatory samples in order to obtain an overall median protein ratio of one ($\log_2$(Ratio)=0), expecting the protein ratios measured to be symmetrically distributed around one.[135–138] Nevertheless, there are inevitable systematic biases intrinsic to LC-MS/MS analysis. As discussed previously, discovery proteomic analysis uses data-dependent spectra acquisition, which by definition gives preference to abundant precursor ions. Additionally, lack of reproducibility in the chromatography can severely affect quantitation. The bias observed in sampling depth was also observed for normalized peak areas and peak intensities, thus the same correction was performed separately for each approach (Figure 3.7).

Relative quantitation was performed on the 22 proteoforms that were present in at least two biological replicates for each condition. In the case of spectral counting, differences in abundance were evaluated using the count-based Fisher's Exact test, χ^2 test and sample-label permutation test. In each case, p-values were adjusted for multiple test correction by Benjamini-Hochberg. As stated in the previous section, the count-based Fisher's Exact test and χ^2 test tend to overestimate significance. Since we analyzed a sufficiently large number of data files (12 data files per condition), a sample-label permutation test was applied in order to obtain empirical p-values and confirm the results provided by the Fisher's Exact test. When comparing the p-values of the proteoforms with differences in abundance obtained for the Fisher's Exact test and the sample-label permutation test (see Table 3.3) the latter proved more conservative.

(a) Acyl-CoA binding protein (9,905 Da)

(b) Actin related protein complex 2/3 subunit 5 (16,308 Da)

(c) Calmodulin (16,780 Da)

(d) 10KDa Heat shock protein mitochondrial (10,867 Da)

(e) Histone H4   (9,044 Da)

(f) Histone H4 (11,343 Da)

(g) H4 (11,360 Da)

(h) Hemoglobin subunit alpha (15,078 Da)

(i) Non-histone chromosomal protein HMG-17 (9,286 Da)

(j) S100A6 (10,074 Da)

(k) S100A8 (10,157 Da)

(l) S100A8 (10,173 Da)

(m) S100A8 (10,189 Da)

(n) S100A8 (10,205 Da)

(o) S100A8 (10,221 Da)

(p) S100A9 (12,963 Da)

Legend:
- Acetylation
- Cysteinylation
- R Citrullination
- Dimethylation
- trimethylation
- hydrogen loss
- methylation
- oxidation
- AAS - Ser to Thr
- sulfinic acid
- N-term acetylation
- sulfonic acid

Figure 3.6. Annotated sequences of a selection of 16 proteoforms found in the exosome lysates. Proteoform name, accession number and observed mass are shown in each case. Putative PTM assigned are color-coded. Full characterization was possible due to the elevated fragmentation density observed. Fragment ions obtained by CID (b- and y- ions) are shown in blue and ETD (c- and z- ions) are represented in red.
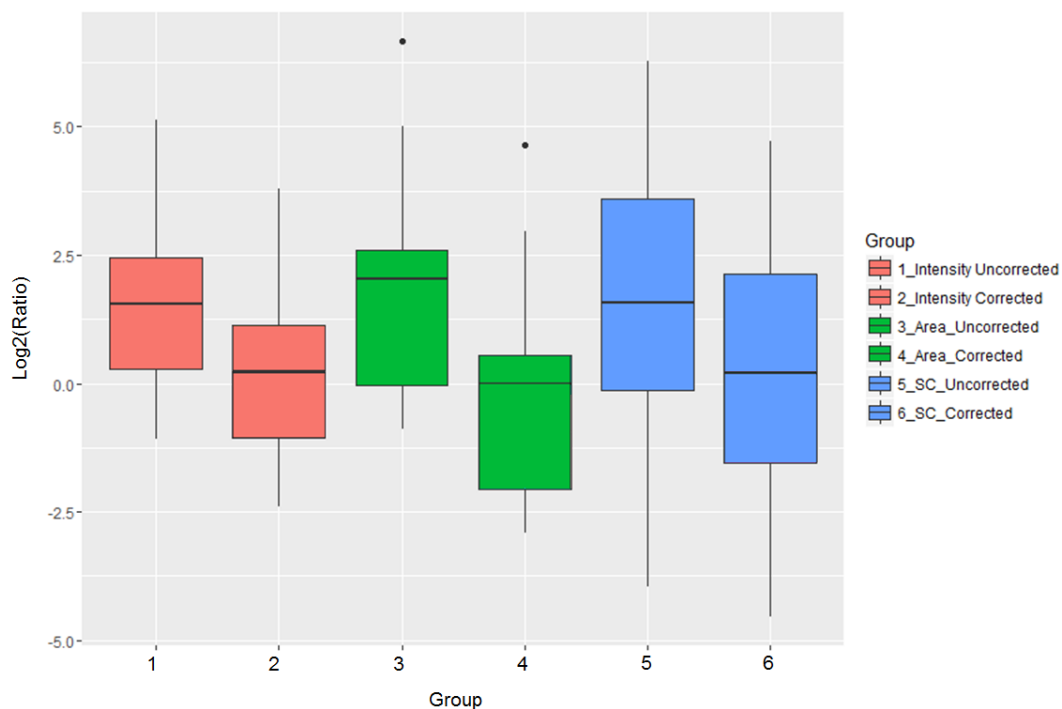
Figure 3.7. Effect of normalization in the distribution of $\log_2$(protein ratios). Uncorrected and corrected $\log_2$(protein ratios) are shown as estimated by normalized peak intensities (red), normalized peak areas (green) and spectral counts (blue).[118]

Based on the sample-label permutation test, changes in abundance of only six proteoforms were significant, and these results often agreed with normalized peak areas and peak intensities that were found significant by Mann-Whitney tests after Benjamini-Hochberg correction. Hence, the results obtained with the sample-label permutation test were used for further discussions. Additionally, as shown in Figure 3.8a,b there is good agreement between protein ratios estimated by normalized peak intensities or normalized peak areas and spectral counting. However, spectral counting provides a slightly higher number of proteoforms with significant differences in abundance (Figure 3.8c-e).

62
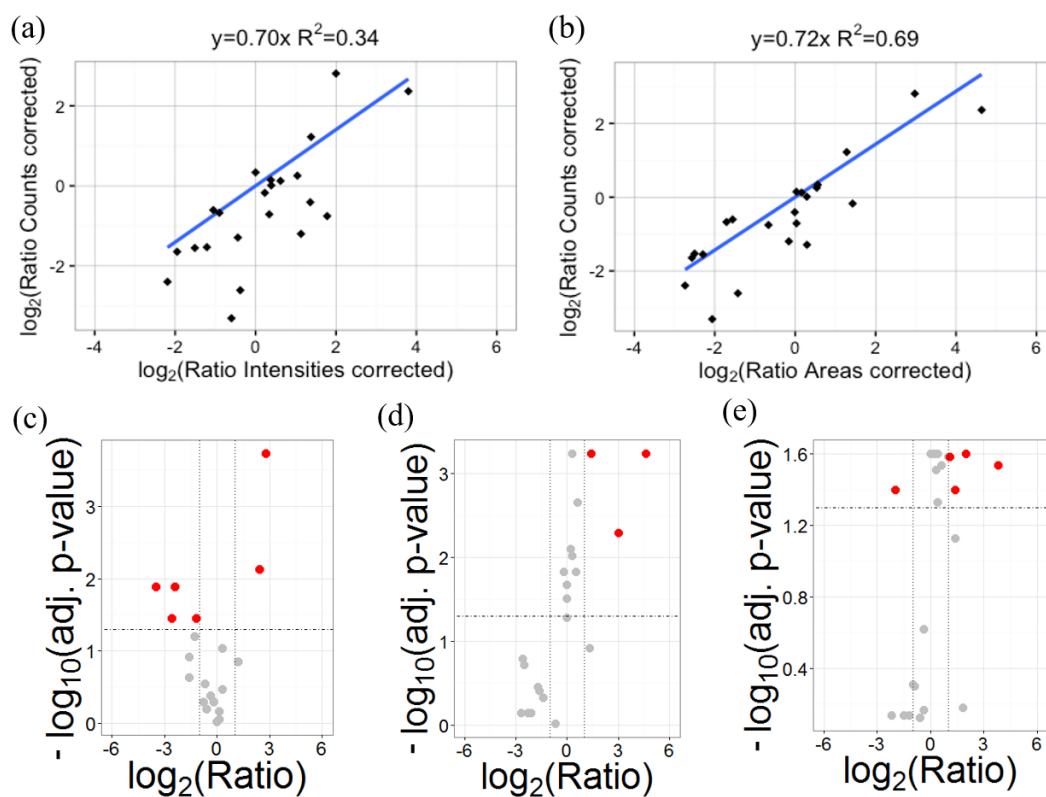
Figure 3.8. Comparison of $\log_2$(ratios) found for the 22 quantified proteoforms in the exosome samples estimated using (a) peak intensities and (b) peak areas vs. those obtained by spectral counting. Proteoforms that showed statistically significant differences in abundance by (c) spectral counting, (d) peak areas and (e) peak intensities are marked in red. The horizontal dashed line in plots (c-e) represents the adjusted *p*-value threshold of 0.05 (-$\log_{10}$(*p*-value) $\geq$ 1.3). Differences in abundance of two-fold are marked in plot (c-e) by dotted vertical lines.[118]

Table 3.3. Summary of protein ratios and statistical significance for the 22 proteoforms present in at least two biological replicates estimated by spectral counting, normalized peak areas and normalized peak intensities. The six proteoforms with statistically significant differences in abundance based on spectral counting and sample-label permutation test FDR are shown in ***bold italics***.[118]

Observed mass (Da)	Spectral counting				Peak Areas		Peak Intensities	
	log$_2$ Ratio (Counts)	Fisher's Exact FDR	Perm. Sample FDR	Chi-square FDR	log$_2$ Ratio (Area)	Mann-Whitney FDR	log$_2$ Ratio (Intensity)	Mann-Whitney FDR
Q9CPW5: Actin related protein complex 2/3 subunit 5								
16308.28	-1.6	1.4×10^{-8}	1.2×10^{-1}	6.2×10^{-8}	-2.6	1.6×10^{-1}	-2.0	4.0×10^{-2}
P62204: Calmodulin								
16779.85	-1.6	1.1×10^{-4}	2.3×10^{-1}	1.2×10^{-4}	-2.5	1.9×10^{-1}	-1.2	7.3×10^{-1}
16795.82	-1.6	5.4×10^{-4}	1.2×10^{-1}	5.6×10^{-4}	-2.3	7.1×10^{-1}	-1.5	7.3×10^{-1}
P22752: Histone H2A type 1								
12354.95	1.2	7.1×10^{-3}	1.4×10^{-1}	5.6×10^{-3}	1.3	1.2×10^{-1}	1.4	7.5×10^{-2}
P01942: Hemoglobin subunit alpha								
15063.71	-0.6	1.9×10^{-1}	6.3×10^{-1}	1.3×10^{-1}	-1.6	3.9×10^{-1}	-1.0	4.9×10^{-1}
15077.81	-0.6	2.7×10^{-1}	6.3×10^{-1}	2.0×10^{-1}	-1.7	3.5×10^{-1}	-0.9	5.0×10^{-1}
P09602: Non-histone chromosomal protein HMG-17								
9286.02	0.1	9.2×10^{-1}	6.8×10^{-1}	6.8×10^{-1}	0.0	2.1×10^{-2}	0.4	4.7×10^{-2}
P14069: Protein S100-A6								
10074.27	-0.7	1.3×10^{-1}	2.8×10^{-1}	9.5×10^{-2}	0.0	3.1×10^{-2}	0.3	3.1×10^{-2}
P31725: Protein S100-A9								
12963.26	***2.4***	***3.0×10^{-12}***	***7.5×10^{-3}***	***6.5×10^{-11}***	***4.6***	***5.8×10^{-4}***	***3.8***	***2.9×10^{-2}***
P27005: Protein S100-A8								
10101.98	0.3	8.7×10^{-1}	3.4×10^{-1}	5.9×10^{-1}	0.5	1.5×10^{-2}	1.0	2.6×10^{-2}
10140.03	***2.8***	***4.8×10^{-5}***	***1.9×10^{-4}***	***7.4×10^{-5}***	***3.0***	***5.1×10^{-3}***	***2.0***	***2.5×10^{-2}***
10157.05	0.3	1.3×10^{-10}	9.2×10^{-2}	4.4×10^{-6}	0.6	2.2×10^{-3}	0.0	2.5×10^{-2}
10172.99	***-1.2***	***2.8×10^{-5}***	***3.5×10^{-2}***	***4.1×10^{-5}***	***-0.2***	***1.5×10^{-2}***	***1.1***	***2.6×10^{-2}***
10179.05	0.1	9.7×10^{-1}	8.8×10^{-1}	7.7×10^{-1}	0.2	8.0×10^{-3}	0.6	2.9×10^{-2}
10189.07	***-2.6***	***9.9×10^{-10}***	***3.5×10^{-2}***	***7.2×10^{-9}***	***-1.4***	***4.7×10^{-1}***	***-0.4***	***6.8×10^{-1}***
10205.02	***-3.5***	***3.5×10^{-7}***	***1.3×10^{-2}***	***2.6×10^{-6}***	***-2.1***	***7.2×10^{-1}***	***-0.6***	***7.5×10^{-1}***
10212.54	-0.8	4.9×10^{-1}	5.1×10^{-1}	3.1×10^{-1}	-0.7	9.6×10^{-1}	1.8	6.6×10^{-1}
10221.02	***-2.4***	***8.0×10^{-12}***	***1.3×10^{-2}***	***7.6×10^{-11}***	***-2.7***	***7.1×10^{-1}***	***-2.2***	***7.3×10^{-1}***
10276.01	-0.4	5.1×10^{-1}	4.1×10^{-1}	3.6×10^{-1}	0.0	5.2×10^{-2}	1.4	4.0×10^{-2}
10288.07	-0.2	4.1×10^{-1}	5.1×10^{-1}	3.6×10^{-1}	1.4	5.8×10^{-4}	0.2	2.5×10^{-2}
10304.04	-1.3	2.3×10^{-4}	6.3×10^{-2}	2.5×10^{-4}	0.3	9.6×10^{-3}	-0.4	2.4×10^{-1}
10340.04	0.0	9.7×10^{-1}	9.4×10^{-1}	9.8×10^{-1}	0.3	5.8×10^{-4}	0.4	2.5×10^{-2}

The six proteoforms showing statistically significant differences in abundance by spectral counting all belong to the chemotactic S100A8 (five proteoforms) and S100A9 (one proteoform) proteins.

In the case of S100A8, four proteoforms were found in greater abundance (2.6 – 6 fold) in conventional exosomes, and one proteoform in greater abundance (7-fold) in inflammatory exosomes. The S100A8 (P27005) proteoforms found in greater abundance in conventional exosomes were characterized by the loss of Met1 and varied degrees of oxidation. The high fragmentation density (74 – 90 % residue cleavage) obtained allowed the determination of the oxidation site(s). The cysteine in position 42 was oxidized to sulfenic acid (Figure 3.6l). sulfinic acid (Figure 3.6m) and sulfonic acid (Figure 3.6n) for S100A8 proteoforms corresponding to the observed masses 10,173 Da, 10,189 Da and 10,205 Da, respectively. Additionally, the S100A8 proteoform with an observed mass of 10,221 Da, could be explained by the loss of Met1, oxidation of Cys42 to sulfonic acid and oxidation of Met37 (Figure 3.6o). Oxidizable forms of S100A8 have been previously identified *in vivo* in neutrophils,[139] and have been proposed to play a protective role against reactive oxygen species produced during inflammation.[140,141] Unfortunately, we were not able to fully characterize the S100A8 proteoform (P27005; 10,140 Da) that was found in greater abundance in inflammatory conditions, as the product ion spectrum was a mixture of fragments from the co-isolated precursor ions corresponding to the S100A8 proteoforms 10,140 Da and 10,157 Da.

In the case of S100A9, one proteoform (P31725; 12,963 Da) was found in greater abundance (6-fold) in inflammatory exosomes. Proteoform characterization was straightforward as good fragmentation density (60% residue cleavage) was observed. This proteoform was formed by the loss of Met1, acetylation of Ala2, methylation of His107, and the formation of a disulfide bond between Cys91 and Cys111 (Figure 3.6p). A partial characterization of this S100A9 proteoform has been reported previously by Raftery *et al.*[142]

Only one of the S100A8 proteoforms (P27005; 10,173 Da) found to be in greater abundance in conventional exosomes was also found to be significant by normalized peak areas and peak intensities. Conversely, spectral counting, normalized peak areas and peak intensities strongly agreed for those proteoforms found in greater abundance in inflammatory exosomes.

3.4. Summary

A thorough evaluation based on spiking experiments and a robust statistical analysis, allowed the comparison of spectral counting with the most commonly used top-down label-free quantitation approaches: normalized chromatographic peak areas and intensities. This study provided proof of concept that spectral counting is applicable to top-down proteomics workflows, and more importantly, that spectral counting has comparable or better sensitivity than the often used chromatographic approaches. Additionally, we demonstrated that spectral counting provides fair estimates of proteoform ratios. The simplicity associated with performing spectral counting, compared to the tedious integration of peak intensities or peak areas, renders it a helpful screening tool that can be applied to large data sets in order to define putative differentially abundant proteoforms. These proteoforms could be later targeted or labeled in order to obtain a more accurate fold change, if necessary.

This study represents the first relative quantitation analysis of MDSC-derived exosomal proteins at the proteoform level. We observed that a set of potentially active proteoforms of S100A8 and S100A9 showed differences in abundance depending on the tumor microenvironment inflammation condition.

Chapter 4: Expression profiling of the miRNA, mRNA and protein cargo of myeloid derived suppressor cells and their exosomes.

This work was carried out in collaboration with Dr. Ashton Belew (A.B,) who performed the sample preparation for RNA cargo analysis and the next generation sequencing. Data processing and functional annotations of the RNA cargo was performed by A.B. and Lucia Geis Asteggiante (L.G.A). Sample preparation for shotgun proteomics, LC-MS/MS analysis and final integrated data analysis was performed by L.G.A.

4.1. Introduction

MicroRNAs are non-coding RNAs composed of 18 to 25 nucleotides that negatively regulate mRNA expression by repressing translation or inducing mRNA degradation.[143,144] These small RNAs exert their repressing function by binding to the 3'-untranslated region (3'-UTR) or open reading frame (ORF) of the target mRNAs.[143,144] Many studies have shown that exosomes carry mRNA and miRNA, which can be transferred to surrounding and distant cells.[145–150] Some of these studies also report that the RNA cargo carried by exosomes is quantitatively different than that of their parental (donor) cells, suggesting a somewhat selective loading. Several sorting pathways have been proposed for loading miRNA into exosomes, though the detailed mechanisms involved have not yet been definded.[151,152] The emerging research on exosomal miRNA profiles comparing healthy and disease donors, combined with the observed stability of miRNAs in bodily fluids, has motivated their potential use as surrogate biomarkers.[13,153] Regarding miRNA functions in exosomes, aside from the traditional role of repressing mRNA expression when delivered into a receiver cell, it has been demonstrated that miR-21 and miR-29a, carried in exosomes shed by HEK-

293 cells, bind and activate the toll-like receptors TLR-8 (human) and TLR-7 (murine) causing a pro-inflammatory response.[154,155]

As stated previously, exosomes are also known to contain protein-coding mRNAs that can also be transferred into receiver cells.[146,147,150,156] The integrity of these mRNAs is relevant as they could be delivered and translated, if the translational machinery is available, modifying the receiver cell. Several studies on varied cellular types have shown that the transferred mRNA is functional and proteins are produced in the receiver cell.[147,156–158] Nevertheless, in complex biological settings such as the tumor microenvironment, it is challenging to determine *in vivo* the identity of the receiver cell(s) and if there is any selectivity in the delivery. Interestingly, Ekström *et al.* used the information provided by pathway analysis of the mRNAs contained in higher abundance in exosomes shed from human mast cells to propose and demonstrate that CD34[+] progenitor cells are putative receiver cells.[147]

Myeloid-derived suppressor cells (MDSCs) are immature myeloid cells known to accumulate in the tumor microenvironment, where they suppress both adaptive and innate immunity.[78] In the presence of inflammation, MDSC abundance in the tumor microenvironment increases and their suppression activity becomes more aggressive, facilitating tumor progression.[4–7] Their ability to suppress the tumor immune response has been shown to be mediated by the release of exosomes (25 to 30 nm in diameter) into the tumor microenvironment.[3] Our recent research has focused on interrogating the protein cargo of MDSC-derived exosomes from mammary carcinoma tumor-

bearing mice shed under different inflammatory conditions, in order to identify and quantify putative biologically active proteins.[3,99,100] However, to the best of our knowledge, the miRNA and mRNA cargo of MDSC-derived exosomes has not yet been interrogated.

The aim of this study was to identify and quantify the protein, mRNA and miRNA contents of MDSCs and MDSC-derived exosomes shed under conventional and heightened inflammation conditions. This work entailed the analysis of several biological replicates of matched parental cells and released exosomes to determine the protein contents by shotgun proteomics and RNA cargo by next generation sequencing (see Figure 4.1). This analysis provided information about signaling pathways in receiver cells that may be affected by the differential exosome cargo profiles.
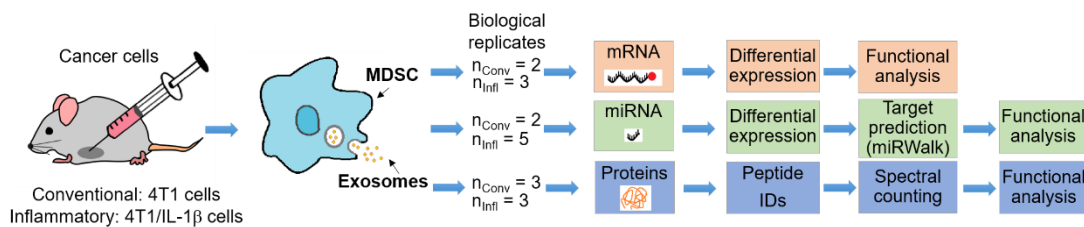
Figure 4.1. Scheme of the experimental design. MDSC and their released exosomes from tumor-bearing mice injected with 4T1 or 4T1/IL-1β mammary carcinoma cells were analyzed for protein, mRNA and miRNA content. The number of biological replicates per condition is shown.

4.2. Materials and methods

4.2.1. MDSC and MDSC-derived exosomes

BALB/c mice were injected with 7000 wild-type syngeneic 4T1 mammary carcinoma cells or 4T1 cells transduced to constitutively express the cytokine IL-

1β (4T1/IL−1β), as previously reported.[3,84] MDSCs were harvested from blood of 1 to 3 mice and MDSC populations confirmed to be > 90% $Gr1^+CD11b^+$ by flow cytometry (1×10^7- 4×10^8 cells) were used for further analyses. An aliquot equivalent to 1×10^6-4×10^7 isolated MDSCs was stored in 1mL of (90:10) fetal calf serum - dimethyl sulfoxide (FCS:DMSO) for both RNA and protein cargo analyses. The reminder of the isolated MDSCs (9×10^6- 3.6×10^8 cells) were incubated for 16 h in serum-free media at 37°C and 5% CO_2. Exosome shed during incubation were isolated by ultracentrifugation and characterized by centrifugation on 0.25M to 2M sucrose gradient and by transmission electron microscopy as previously reported by Burke *et al.*[3] RNAlater from Life Technologies (Carlsbad, CA) was added to all the samples collected for RNA analyses.

In this study, MDSC and MDSC-derived exosomes from 4T1 and 4T1/IL-1β tumor-bearing mice are termed "conventional" and "inflammatory", respectively. All animal experiments were approved by UMBC and UMCP Institutional Animal Care and Use Committees.

4.2.2. RNA Isolation

Total RNA was isolated using the mirVana miRNA Isolation Kit from Life Technologies (Carlsbad, CA) following manufacturer's instructions with modifications. Briefly, MDSC and MDSC-derived exosomes were lysed using the kit lysis buffer containing detergent, to which phenol/chloroform, and guanidinium isothiocyanate were added. One or two additional phenol extraction steps were necessary to remove excess lipids. Lysates were then cleaned up using glass-fiber filter

columns that immobilize RNA. Large RNA species (>100nt) were isolated by precipitation in 33% ethanol, and small RNAs (>10nt) by precipitation in 80% ethanol. The small RNA fraction collected was subsequently visualized using a 10% denaturing polyacrylamide gel and Tris/Borate/EDTA (TBE) running buffer. Two biological replicates of large and small RNAs species were analyzed for conventional MDSCs and MDSC-derived exosomes. In the case of inflammatory MDSCs and MDSC-derived exosomes, three and five biological replicates were analyzed for large and small RNAs species, respectively.

4.2.3. Creation of cDNA libraries and sequencing

MicroRNA cDNA libraries were created using the TruSeq Small RNA Library Preparation Kit from Illumina (San Diego, CA) following manufacturer's instructions. Transcripts (mRNA) cDNA libraries were created using the TruSeq RNA Library Preparation Kit version 2 from Illumina. Manufacturer's instructions were followed for MDSCs and exosomes, with the exception that the initial amounts of exosomal RNA had to be adjusted in order to compensate for the lack of ribosomal RNA and ensure equivalent amounts across samples. The quality of MDSCs' large RNAs species and concentration of all RNA libraries was evaluated using a 2100 Bioanalyzer from Agilent Technologies (Santa Clara, CA). Libraries were sequenced using a HiSeq1500 sequencer from Illumina.

4.2.4. Processing next-generation sequencing data

The quality of the reads obtained was evaluated using FastQC[159] and biopieces (biopieces.org). Large RNA adapter sequences were removed using Trimmomatic[160] and, small RNA adapters using Cutadapt.[161] Libraries were mapped against the Ensembl *Mus musculus* genome with annotations (version GRCm38.75/mm10, Dec 2015) using Bowtie[162] and TopHat;[163] and the *Mus musculus* pre-computed Ensembl transcriptome database (version GRCm38.79/mm10, Dec 2015) using Kallisto.[164] Alignments were performed limiting seed sequences to 10 nt for Bowtie and 20 nt for TopHat, requiring the absence of mismatched seeds and assigning multi-matched reads to one locus. In the case of Kallisto, 19 nt indexes were used for small RNAs and 31 indexes for large RNAs. In all cases, alignments were sorted using SAMtools.[165] For Bowtie and TopHat, aligned reads were counted for each gene using HTSeq.[166] The sorted alignments were mapped against the *Mus musculus* immature miRNA database (http://www.mirbase.org, version 21).[167,168]

4.2.5. Data visualization and clustering

Possible sequencing depth biases due to the variability of biological replicates and sample batch effects were assessed. This process entailed the creation of density plots and boxplots, hierarchical clustering analysis based on Pearson's correlation and Euclidian distance, and principal component analysis, before and after data normalization. Even though miRNA of the MDSC recovered after shedding exosomes was not evaluated in this study, it was still considered for hierarchical clustering as its replicates provided variance for batch effect estimation. Several normalization

approaches were evaluated including quantile,[169] trimmed mean of M-values,[170] relative log expression,[171] upper quartile[172] and variance stabilized data.[173] In the case of miRNA, quantile normalization was selected for data quality assessment. Normalization was not applied to the mRNA reads, because the mRNA profiles between exosomes and MDSC were significantly different. In all cases, a $\log_2$ transformed counts per million (cpm) reads after low read count filtering was performed. A low read count was defined as any feature with counts less than twice the number of samples or cases where any single sample has less than 2 read counts. Moreover, bias due to batch effects was evaluated using several algorithms such as surrogate variable analysis (SVA) and ComBat (Combating Batch Effects When Combining Batches of Gene Expression Microarray Data),[174] remove unwanted variation[175] and batch factor removal via residuals.

4.2.6. Differential expression analysis

MicroRNAs and mRNAs differences in abundance were estimated for all possible pairwise comparisons using the Bioconductor packages linear model for microarray (limma),[176] empirical analysis of digital gene expression data in R (edgeR)[171] and differential gene expression analysis based on the negative binomial distribution (DEseq).[173] The Benjamini-Hochberg method was used to adjust *p*-values for multiple testing. Since the biological replicates were analyzed at two time points with several months in between, data was initially treated separately. As both data sets showed similar trends, the data was combined and the statistical models were adjusted to include batch. Ultimately, limma was selected for further data analysis, as the results

74

using the combined data showed a good agreement between models. Genes with differential abundances were defined as those with a fold-change $\geq |2|$ and adjusted *p*-value ≤ 0.05.

4.2.7. miRNA target gene prediction

The miRWalk2.0[177,178] database web tool was used to determine targets for the top 5 miRNA that presented statistically significant differences in abundance (fold-change $\geq |2|$, adjusted *p*-value ≤ 0.05). miRWalk combines the output of 11 prediction tools: DIANA,[179] PicTar,[180] miRanda,[181] TargetScan,[182] Pita,[183] RNAHybrid,[184] miRDB,[185] mirBridge,[186] miRNAMap,[187] RNA22,[188] doRiNA,[189] and a validated target database miRTarBase.[190] Only those target genes with match seed of 8 nt that were predicted by at least 2 of the tools were considered for Gene Ontology (GO) and/or KEGG pathway analysis.

4.2.8. Gene ontology and pathway annotations

mRNA and miRNA target genes that were found to be significantly enriched (fold-change $\geq |2|$, adjusted *p*-value ≤ 0.05) in our pairwise comparisons were annotated against GO categories using the R package gProfiler.[191] Gene enrichments were considered statistically significant for adjusted *p*-values ≤ 0.05.

4.2.9. Sample preparation for shotgun proteomic analysis

4.2.9.1. Exosome lysis

Exosomes were prepared following Burke *et al.*[3] Briefly, lysis was performed in 8M urea in 50 mM ammonium bicarbonate buffer with protease inhibitor cocktail from Sigma-Aldrich (St. Louis, MO). Lysates were buffer exchanged to reach a final concentration of 8 mM urea, using 3 KDa molecular weight cut-off filters (MWCOF) from Millipore (Cork, Ireland). The protein content was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Rockford, IL). Twenty-five microgram lysate aliquots were: (1) reduced by incubating the sample with 20 mM DTT for 30 min at 56°C; (2) alkylated by incubating the sample with 40 mM iodoacetamide for 30 min in the dark at room temperature; and (3) tryptic digested by incubating the lysates for 16h using Trypsin Gold (Promega Corporation, Madison, WI) in a 1:50 ratio between trypsin and sample protein content. Digestion was stopped by adding formic acid to reach a final concentration of 1%. All digested lysates were lyophilized and reconstituted in solvent A (97.5:2.5 H_2O-acetonitrile in 0.1% formic acid) prior to LC-MS/MS. Three biological replicates for each condition were analyzed, performing five technical replicates per sample.

4.2.9.2. MDSC lysis

MDSCs were thoroughly washed by centrifugation at $900 \times g$ for 10 min at 4°C using 10 mL of cold phosphate buffered saline solution from Sigma-Aldrich. MDSC pellets were lysed by incubation in 8M urea in 50 mM ammonium bicarbonate buffer with protease inhibitor cocktail for 30 min at room temperature. In order to ensure that

cells are completely lysed and the lysate is homogeneous, MDSCs were further lysed mechanically using a set of syringes with needles of sequentially smaller gauge size (18, 20, 21.5) obtained from Becton Dickinson & Co. (Franklin Lakes, NJ) and performing 5-10 strokes each time. Cell debris were removed by centrifugation at 14,000 × g for 10 min, supernatants transferred to 3 KDa MWCOF and buffer exchanged to reach an 8 mM urea concentration. The protein content of MDSC was measured using the Pierce BCA Protein Assay Kit. Twenty-five microgram aliquots were reduced, alkylated and digested as stated previously. Three biological replicates for each condition were analyzed, each with five technical replicates.

4.2.10. Protein analysis by liquid chromatography mass spectrometry

Samples were analyzed using an Ultimate 3000 RSLCnano system (Dionex, Sunnyvale, CA) coupled to an orbitrap Fusion Lumos Tribrid (Thermo Fisher Scientific, Waltham, MA). Chromatographic separation was obtained using a C18 PepMap trap (0.3 × 5 mm, 5μm particle size, 100Å) and C18 Acclaim PepMap RSLC column (0.075 × 250 mm, 2μm particle size, 100Å), both obtained from Thermo Fisher Scientific. Samples were loaded into the trap, desalted and concentrated using Solvent A at a flow rate of 5 μl/min for 10 min, and then eluted from the trap and further separated in the analytical column by linearly increasing Solvent B (acetonitrile-H_2O, (75:25) in 0.1% formic acid) from 5 to 55%, during the span of 150 min under a flow rate of 0.3 μL/min.

Precursor ion mass resolution of 120,000 (at *m/z* 200) and product ion unit mass resolution were used. Ions were collected based on a target automatic gain control of 4×10^5 and 4×10^3 for precursor and product ions, respectively. The maximum injection times for precursor and product ions were set to 50ms and 100ms, respectively. Data dependent acquisition was carried out in a fixed 3s duty cycle, in which the top *n* most abundant precursor ions (intensity $> 1\times10^4$) carrying charges from +2 to +7 were isolated by the quadrupole within a 1.6 *m/z* isolation window. Dynamic exclusion (DE) was set to exclude precursor ions for 60s after being selected once in 30s. The isolated precursor ions were fragmented using CID with helium set to 35% normalized collision energy.

4.2.11. Protein identification and relative quantitation

This study entailed the analysis of four sample types: conventional MDSCs, inflammatory MDSCs, conventional exosomes and inflammatory exosomes. Three biological replicates per sample type were analyzed by LC-MS/MS with five technical replicates per biological replicate. A total of 60 data files were searched against the Uniprot *Mus musculus* reference protein sequence database including isoforms (January 2015) using three search engines (MSGF+, X!Tandem and OMSSA) in PepArML.[55] Search parameters allowed for up to 1 missed cleavage and, a precursor (monoisotopic or first ^{13}C peak) and product ion mass tolerance of 0.05 and 0.5 Da, respectively. Carbamidomethylation was considered as a fixed modification and, oxidation of methionine, deamination of N-terminus glutamine, dehydration of N-terminus glutamic acid and pyro-carbamidomethylation of N-terminus cysteine as

variable modifications. Spectral FDR were estimated in PepArML following Elias and Gygi.[50] Identified peptides were filtered by spectral FDR ≤ 0.31% and a two-unshared peptide global parsimony applied. The resulting protein FDR was of at most 1%, as estimated using MAYU.[54]

Label-free relative quantitation was performed by spectral counting. An in-house software was used to provide the number of peptide spectrum matches (PSM) or spectral counts for each inferred protein based on PepArML search results, after spectral FDR filtering. Differences in abundance between conditions were estimated following Old *et al.* using Eq. (6).[57] Ratios from spectral counts (Rsc) were estimated for each identified protein and statistically significant differences in abundance were determined using the Fisher's Exact test with Benjamini-Hochberg adjustment for multiple testing.[49] Three comparisons were carried out: (1) inflammatory exosomes vs. conventional exosomes; (2) inflammatory MDSC vs. conventional MDSC; and (3) MDSC vs. exosomes irrespective of inflammation condition. Depending on the comparison, Rsc >1 means that a protein is present in greater abundance, by more than 2-fold, in the "inflammatory" or "MDSC" samples. On the contrary, Rsc < -1 refers to an increase in abundance, by more than 2 –fold, in the "conventional" or "exosomes" samples.

$$Rsc = \log_2((n_2+f)/(n_1+f)) + \log_2((t_1-n_1+f)/(t_2-n_2+f)) \qquad \text{Eq. (6)}$$

where n_1 and n_2 refers to the number of spectral counts obtained for a protein in sample 1 and 2, t_1 and t_2 is the total number of spectral counts observed in each sample and f is a correction factor of 0.5 added in order to adjust for cases where a protein is not identified in one of the sample types compared. In this study, the subscript 2 refers to the "inflammatory" condition when comparing between inflammation conditions, or "MDSC" when comparing MDSC vs. exosomes. Inversely, subscript 1 refers to spectral counts of a protein belonging to the "conventional" condition or "exosomes".

4.2.12. Proteomics gene ontology analysis

Identified proteins were annotated by GO categories using the generic and PIR (Protein Information Resource, http://pir.georgetown.edu) GO slims, and the Database for Annotation, Visualization and Integrated Discovery version 6.7 (DAVID).[192,193] Additionally, the Kyoto Encyclopedia of Genes and Genomes (KEGG)[194] and Reactome[195] pathway databases were used to determine putative enriched pathways. Enrichment of GO categories and KEGG pathways were evaluated using all the identified proteins as background and the Fisher's Exact test with Benjamini-Hochberg adjustment for multiple testing.

4.3. Results and discussion

4.3.1. RNA isolation and data quality assessment

The RNA fractions isolated from MDSCs and MDSC-derived exosomes were visualized and RNA concentrations measured by capillary electrophoresis using the Bioanalyzer instrument. The characteristic 18S and 28S ribosomal RNA (rRNA) peaks were observed in the MDSC large RNA fraction as expected (Figure 4.2a), showing no

apparent RNA degradation during sample preparation. Similarly, good isolation was

shown for the MDSC small RNA fractions (Figure 4.2b). Evidence that MDSC-derived

exosomes carry mRNA and miRNA is shown in Figure 4.2c-d. Ribosomal RNA was

not detectable in MDSC-derived exosomes, which is in agreement with previous small

RNA studies on exosomes shed by mast cells,[146,147] melanoma cells,[196] and colon

cancer cells.[197] Crescitelli *et al.* reported that rRNA is generally found in apoptotic

bodies, supporting the conclusion that our exosome samples were not contaminated

with dead cells.[198]
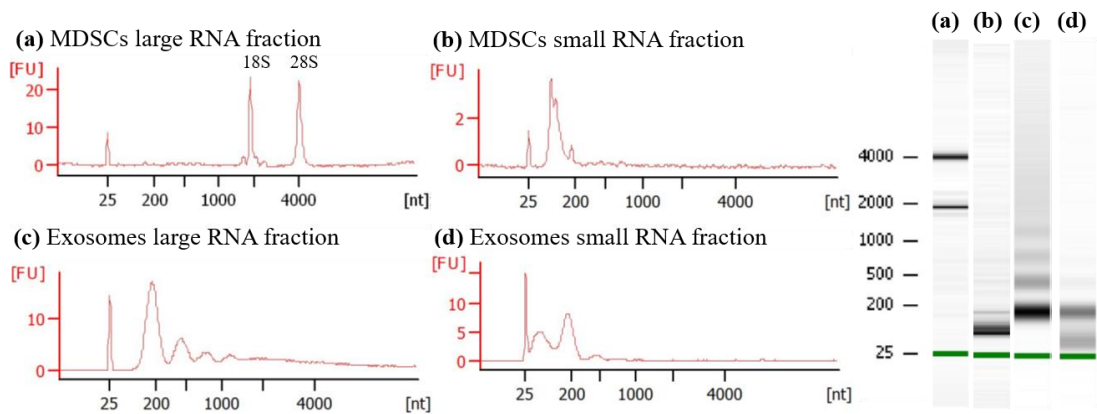
Figure 4.2. RNA size distribution observed by capillary electrophoresis for MDSCs (a) large RNA and (b) small RNA fractions; and exosomes (c) large RNA and (d) small RNA fractions. The y-axis is labeled as [FU] for fluorescence units and x-axis as [nt] for nucleotide length.

The yield of the large and small RNA fractions normalized to the number of

MDSC cells incubated for exosome release was similar for inflammatory and

conventional exosomes. Table 4.1 summarizes the average concentration of small and

large RNA fractions per sample type. Fourteen small RNA libraries and eleven large

RNA libraries were sequenced with good accuracy as reported by FastQC (Appendix

4.2). The reads obtained, which are sequences of nucleotides from the RNA analyzed, were mapped against the Ensembl *Mus musculus* genome. An average of 81,130 and 69,474 reads were mapped to mRNA in exosomes and MDSC, respectively. In the case of miRNA, the number of mapped reads were ~2,500 for both MDSC and their exosomes. The individual reads that could be mapped for miRNA and mRNA samples are summarized in Figure 4.3a-b. The gene expression distribution for each library was evaluated by density plots, showing the number of reads ($\log_2$(filtered counts per million reads)) mapped per gene. A similar density profile was observed for miRNA in both exosomes and MDSC (Figure 4.3c). However, the exosomal mRNA profiles were different than those of the MDSC mRNA libraries (Figure 4.3d). In the case of exosomes, fewer genes were detected but with a larger number of hits per gene, for this reason mRNA libraries were not normalized.

Table 4.1. Average RNA concentrations measured for MDSCs and MDSC-derived exosomes by inflammation condition. Concentrations are normalized by the number of MDSCs or the number of MDSCs that were incubated to release the exosomes are also shown.

Sample type	RNA fraction	Condition	Concentration (ng/uL)	Concentration (ng/uL per 10^6 cells)
Exosomes	Small	Conventional	13	0.14
		Inflammatory	58	0.13
Exosomes	Large	Conventional	29	0.33
		Inflammatory	84	0.28
MDSCs	Small	Conventional	6	0.71
		Inflammatory	43	0.94
MDSCs	Large	Conventional	23	2.65
		Inflammatory	100	3.27

The RNA libraries were classified using PCA, including a surrogate variable (SVA). Clear clusters were observed, shown in Figure 4.4a-b, based on sample type (MDSC and exosome) and to a certain degree also by inflammation condition. The two

first principal components accounted for 67 and 87% of the variance for miRNA and mRNA analyses, respectively. Additionally, the analyzed libraries were inspected by plotting heatmaps using the Pearson's correlation coefficients (Figure 4.4c-d) and Euclidean distance (Appendix 4.3) as similarity measurements in order to perform unsupervised grouping of individual libraries. Figure 4.4c-d also show clear differences in groups based on sample type (MDSC and exosome), and a reasonable segregation by inflammation condition.
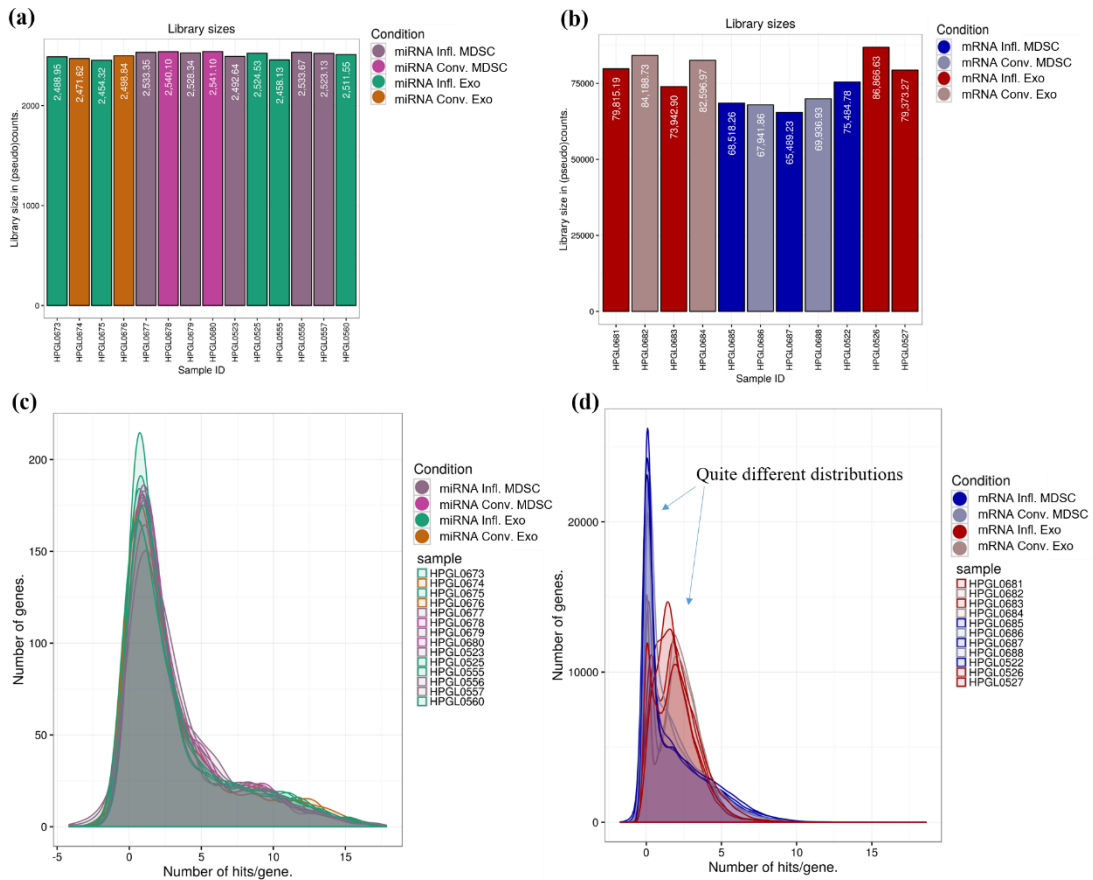
Figure 4.3. Number of mappable sequence reads for each library of (a) miRNA after quantile normalization and (b) mRNA without normalization. Density plot showing the number of genes vs. $\log_2$(count per million reads) after low count filtering per gene for (c) miRNA after quantile normalization and (d) mRNA without normalization.
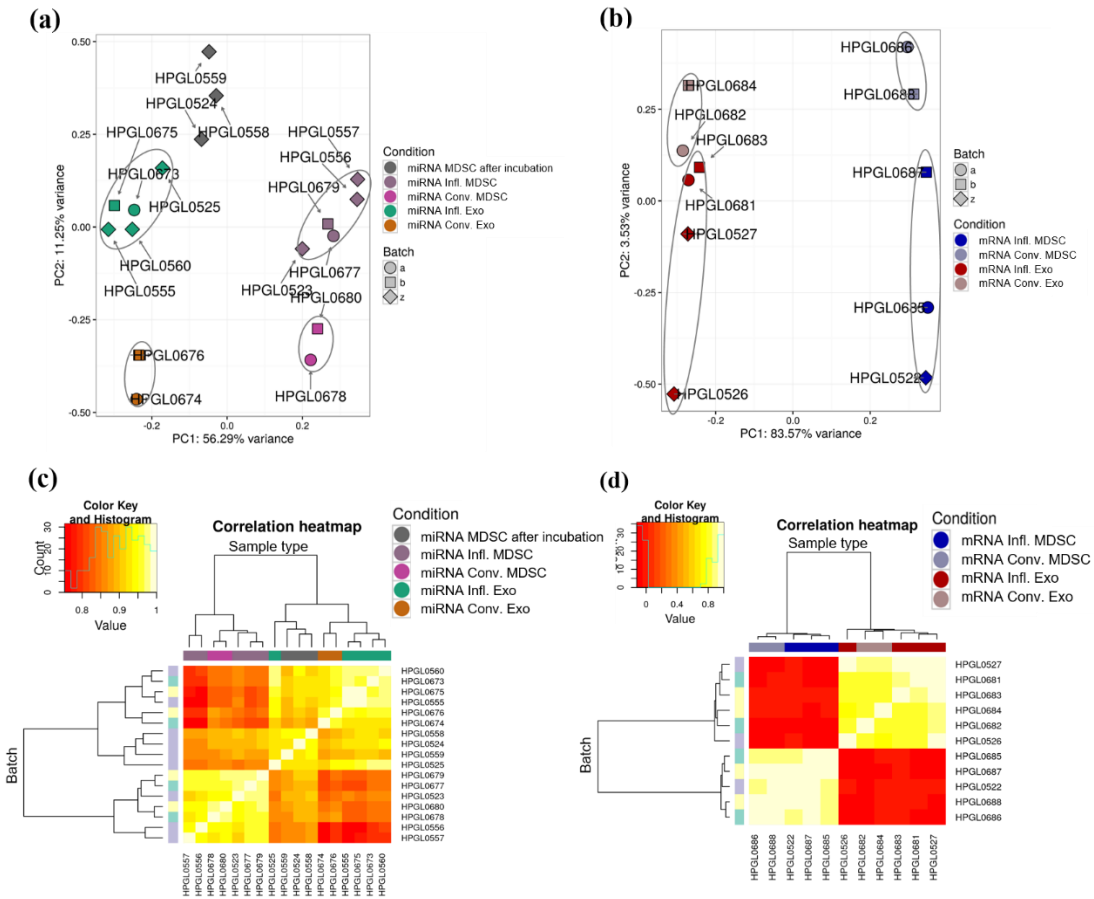
Figure 4.4. Principal component analysis plots for (a) miRNA and (b) mRNA. Heatmap of hierarchical clustering by Pearson's correlation for (c) miRNA and (d) mRNA.

4.3.2. mRNA profiles and putative functions

A total of 53,491 mRNA transcript isoforms were identified with two or more raw read counts in at least one of the samples analyzed. In the case of exosomes, 27,482 and 25,743 mRNA transcripts were identified under inflammatory and conventional conditions, respectively. In the case of MDSC, 23,685 and 22,981 mRNA transcripts were identified under inflammatory and conventional conditions, respectively (see Appendix 4.1).

For relative quantitation purposes mRNA identifications were filtered to remove low read count identifications as stated in the "Methods and Materials" section, reducing the number of total mRNA transcripts to 40,433 confident identifications. Subsequently, four quantitation comparisons were performed: (1) MDSC inflammatory vs. conventional, (2) exosomes inflammatory vs. conventional, (3) exosomes vs. MDSC under conventional inflammation, and (4) exosomes vs. MDSC under heightened inflammation (see Appendix 4.1). The number of mRNA transcripts found to be in greater and lower abundance for each comparison is shown in Table 4.2. A large proportion of the mRNA transcripts found showed statistically significant differences in abundance, when comparing exosomes to their parental cells irrespective of their inflammation conditions. However, differences between inflammation conditions are more subtle, with only 1-6% of the total mRNA transcripts found presenting statistically significant differences in abundance in exosomes and MDSCs. For the purpose of this study, we will focus on comparing the exosome mRNA cargo to that of their parental cells and evaluating the effect of inflammation on the exosome mRNA cargo.

Table 4.2. Number of mRNA transcript isoforms found to have statistically significant differences in abundance.

Comparison	# mRNAs greater in abundance	# mRNAs lower in abundance
Infl. MDSC vs. Conv. MDSC	1,847	2,447
Infl. exosomes vs. Conv. exosomes	339	1,489
Conv. exosomes vs. Conv. MDSC	17,783	7,262
Infl. exosomes vs. Infl. MDSC	18,858	10,457

4.3.2.1. Comparing the mRNA cargo of exosomes and their parental cells

A large number of mRNA transcripts were found to be in greater abundance in exosomes when compared to their parental cells. In order to evaluate the putative functions the enriched cargo can have when delivered and translated into a receiver cell, GO categories and KEGG pathways were annotated. Many GO biological processes and KEGG signaling pathways were found to be enriched in exosomes. Figure 4.5 show the statistically significant (adjusted *p*-value $\leq$ 0.05) enriched categories, including the percentage of functional genes found that belonged to each category. In most cases similar processes and pathways were enriched independently of the inflammation condition.

Several of the enriched GO categories were related to intercellular communication including: "chemotaxis", "cell-cell signaling", "cell surface signaling", "cell aggregation" and "biological adhesion" (see Figure 4.5). Since these categories involve surface proteins, we utilized the enriched GO cellular compartment categories "cell surface" (53% of functional genes, *p*-value = 8.2×10^{-6}), "cell periphery" (48% of functional genes, *p*-value = 3.6×10^{-7}) and "plasma membrane" (47% of functional genes, *p*-value = 1.3×10^{-6}) to determine the mRNA transcripts that encode for surface proteins. A total of 1,804 mRNA transcripts can translate into surface proteins, and 128 out of 1,804 corresponded to integrins and CD proteins, which are interesting proteins as many are membrane receptors or ligands that could modulate various signaling pathways on the receiver cell.[199] A detailed list of this 128 mRNA transcripts and their estimated fold-change is shown in Appendix 4.4. The largest difference in abundance

was observed for *DDR2* mRNA encoding for CD167b (Q62371), which was found to be 1,783-fold more abundant in inflammatory exosomes than their corresponding parental cells, and only 32-fold higher in conventional samples. CD167b is a tyrosine kinase receptor expressed in mesenchymal cells that binds collagen I and III. Elevated levels of *DDR2* mRNA and CD167b have been observed in breast cancer tumors and are associated with poor prognosis and metastasis.[200–202]

Interestingly, KEGG pathways related to cancer, such as "basal cell carcinoma", "proteoglycans in cancer", "choline metabolism in cancer" and "pathway in cancer" were enriched in exosomes. Additionally, KEGG pathways and GO biological processes related to growth, cell proliferation, cell migration and survival[203] were also significantly enriched in exosomes, and VEGF and ErbB signaling pathways related to angiogenesis[204] and chemotaxis[205] were observed to be statistically significant under inflammatory conditions only (see Figure 4.5). Therefore, depending on the receiver cell, these mRNA transcripts could participate in relevant signaling pathways affecting tumor progression, invasion and metastasis.

Most of the mRNA transcripts with the largest differences in abundance (> 2000-fold) in exosomes, irrespective of their inflammation condition, encode for Uniprot TrEMBL predicted proteins. Approximately 10 out of the top 60 mRNA transcripts encoded for Uniprot SwissProt (manually annotated) proteins, including E3 ubiquitin-protein ligase Midline-1 (O70583) and E3 ubiquitin-protein ligase RNF152 (Q8BG47). Normally in the cell, these proteins transfer the ubiquitin carried by the

ubiquitin conjugating enzyme (E2) to a substrate protein and are important as it determines the linkage site.[206] The mRNA transcripts of E2 and E1 ubiquitin activating enzymes were also observed in greater abundance in exosomes.

In the case of the mRNA transcripts encoding for the characteristic pro-inflammatory proteins S100A8 (P27005) and S100A9 (P31725), exosomes contained less of these mRNA transcripts than their parental cells, and no significant difference in abundances were observed between inflammation conditions. However, the mRNA transcripts encoding for HMGB1 (P63158) were found at similar levels in both exosomes and MDSC.

4.3.2.2. Comparing the mRNA cargo of exosomes

The number of mRNA transcript isoforms carried by exosomes that showed differences in abundance when comparing their inflammation condition were markedly lower with only a few mRNA transcripts encoding for Uniprot TrEMBL predicted proteins (Table 4.2). The largest fold-change observed in this comparison was ~170-fold. Even though the number of enriched mRNA transcripts was lower, the putative GO biological processes and KEGG pathways that may be affected if these transcripts are expressed in a receiver cell were very interesting and some differences between inflammation conditions can be observed (see Table 4.3).
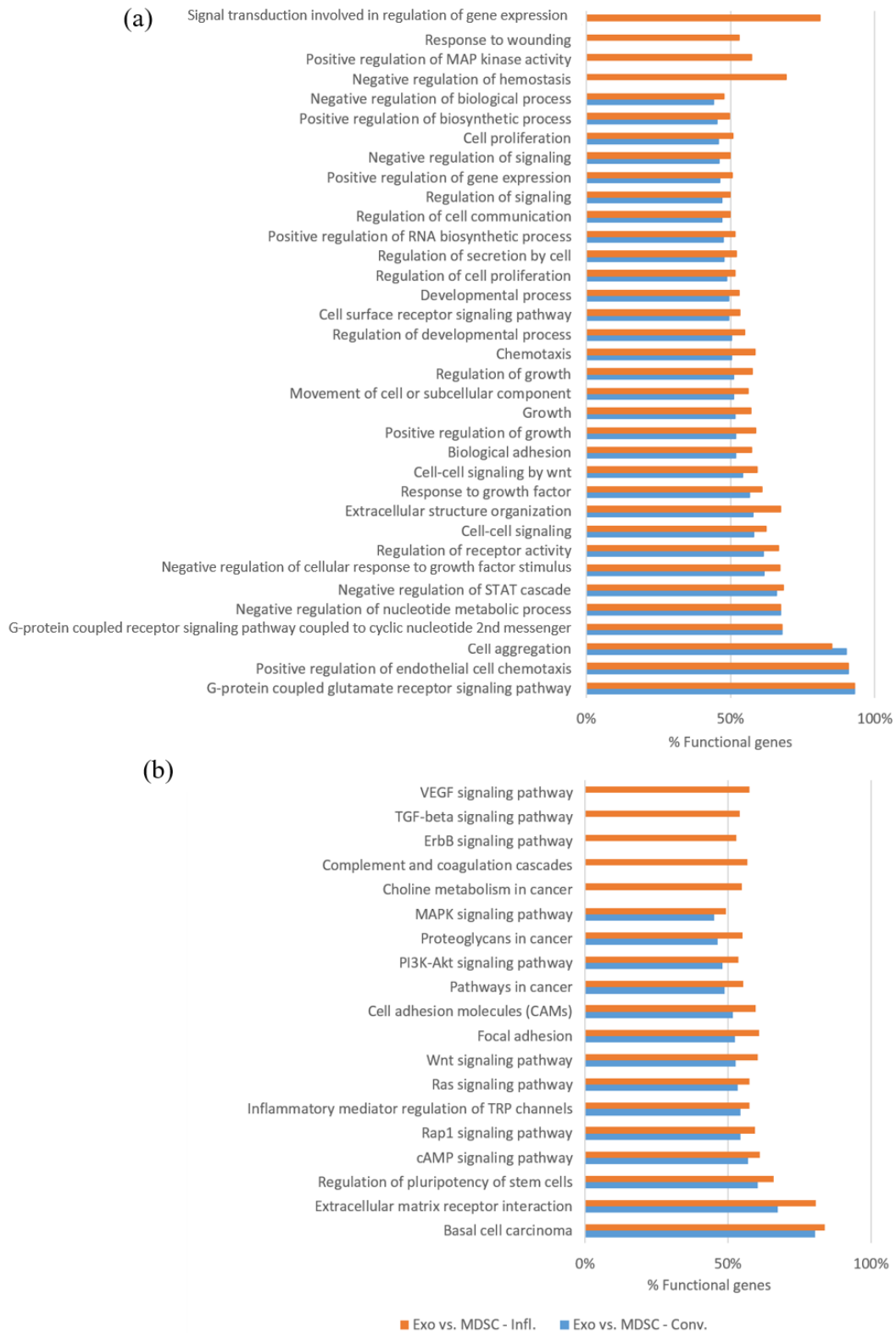
Figure 4.5. Selected enriched (a) GO biological processes and (b) KEGG pathways for transcripts detected in greater abundance in exosomes, when compared against their parental cells, for conventional (blue) and inflammatory (orange) conditions. Categories shown were statistically significant with adjusted p-value ≤ 0.05.

In the case of mRNA transcripts found in greater abundance in conventional exosomes, most of the enriched GO biological processes were related to metabolism, endocytosis and regulation of gene transcription. Several biological process categories composed of a small number of genes, were well represented with 60-75% of the functional genes found in our samples. An example is: "positive regulation of histone deacetylase activity", which is a process relevant for gene transcription and could have varied biological effects.[207] Additionally, no KEGG pathway was found to be significantly enriched.

In the case of mRNA transcript isoforms found in greater abundance in inflammatory exosomes, GO biological processes related to regulation of gene transcription, intercellular communication, and cell differentiation were enriched. In this case the biological processes "regulation of histone H3-K27 acetylation" and "oncogene-induced cell senescence", both related to gene transcription, were found to be highly represented (67%). Additionally, "blood vessel endothelial cells differentiation" a process related to angiogenesis and promoted by VEGF was also found to be enriched under inflammatory conditions.[208] Several KEGG pathways related to cancer were found to be enriched in inflammatory exosomes, including "pathways in cancer", which was previously discussed above to be enriched in exosomes when compared to their parental cells.

90

Table 4.3. Selected enriched GO biological processes and KEGG pathways for mRNA transcripts that were found to be in greater abundance in (a) conventional exosomes and (b) inflammatory exosomes, when comparing inflammation conditions.

(a) Greater abundance in conventional exosomes		
GO Biological Process	*% Functional genes*	Adj. *p*-value
Adherens junction assembly	16%	3.4×10^{-2}
Developmental process	7%	2.1×10^{-5}
Locomotion	8%	1.4×10^{-3}
Metabolic process	6%	1.5×10^{-3}
Negative regulation of gene expression	7%	3.0×10^{-2}
Negative regulation of mesenchymal cell proliferation	50%	1.5×10^{-2}
Negative regulation of pinocytosis	75%	1.9×10^{-2}
Positive regulation of cholesterol homeostasis	60%	4.5×10^{-2}
Positive regulation of histone deacetylase activity	60%	4.5×10^{-2}
Positive regulation of protein ubiquitination	14%	3.5×10^{-2}
Positive regulation of translation	14%	4.0×10^{-2}
Ras protein signal transduction	10%	3.0×10^{-2}
Receptor-mediated endocytosis	11%	4.0×10^{-2}
Regulation of cell communication	7%	1.0×10^{-4}
Regulation of cell differentiation	7%	4.0×10^{-2}
Regulation of receptor internalization	23%	1.4×10^{-2}
Regulation of vascular permeability	26%	3.4×10^{-2}
Response to inorganic substance	9%	3.1×10^{-2}
Ribonucleotide metabolic process	9%	4.8×10^{-2}
(b) Greater abundance in inflammatory exosomes		
GO Biological Process	*% Functional genes*	Adj. *p*-value
Blood vessel endothelial cell differentiation	33%	3.8×10^{-2}
Canonical Wnt signaling pathway	4%	3.4×10^{-2}
Cell surface receptor signaling pathway	2%	7.7×10^{-3}
Cell-cell signaling	3%	3.1×10^{-3}
Chromatin organization	3%	2.8×10^{-2}
Macrophage differentiation	14%	6.8×10^{-3}
Negative regulation of cell differentiation	3%	2.4×10^{-3}
Negative regulation of cell proliferation	3%	3.0×10^{-3}
Negative regulation of growth	4%	2.3×10^{-2}
Negative regulation of transcription from RNA polymerase II promoter	3%	4.9×10^{-2}
Oncogene-induced cell senescence	67%	7.8×10^{-3}
Positive regulation of cellular metabolic process	2%	2.4×10^{-2}
Positive regulation of MAPK cascade	3%	1.1×10^{-2}
Positive regulation of protein binding	7%	4.0×10^{-2}
Positive regulation of transcription, DNA-templated	2%	2.2×10^{-2}
Regulation of cell communication	2%	1.9×10^{-2}
Regulation of cellular response to drug	67%	7.8×10^{-3}
Regulation of developmental process	2%	1.1×10^{-4}
Regulation of histone H3-K27 acetylation	67%	7.8×10^{-3}
Regulation of phospholipase activity	9%	3.4×10^{-2}
Response to calcium ion	6%	8.1×10^{-3}
KEGG Pathway	*% Functional genes*	Adj. *p*-value
Basal cell carcinoma	11%	1.7×10^{-3}
Hippo signaling pathway	6%	3.4×10^{-3}
Melanogenesis	6%	4.5×10^{-2}
Pathways in cancer	3%	3.1×10^{-2}

4.3.3. miRNA profiles and putative functions

A total of 1,890 miRNAs were identified with two or more raw read counts in at least one of the samples analyzed. The median number of miRNA identified in conventional and inflammatory exosomes were 557 and 426, respectively. In the case of conventional and inflammatory MDSC, the median number of miRNAs identified were 721 and 574, respectively (see Appendix 4.1). Seventy-five miRNAs, listed in Appendix 4.5, were present in all 14 libraries with at least five raw read counts. After performing low read count filtering, a total of 1,453 miRNAs confident identifications were used for the four relative quantitation comparisons under analysis. For the purpose of this study, we will focus on those miRNA that were found in greater abundance in exosomes and the exosomal miRNAs that have abundance differences between inflammation conditions. The number of miRNAs found to be in greater or lower abundance for each comparison is shown in Table 4.4. An important observation is that 43% (~624) of all the miRNAs found were predicted by Ensembl. Predicted miRNAs as defined by Ensembl (http://useast.ensembl.org/info/genome/genebuild/ncrna.html) entail the determination of similar regions in the *Mus musculus* genome using BLAST (Basic Local Alignment Search Tool) and comparing against miRBase sequences from all species. Additionally, Ensembl requires that the aligned nucleotide sequence can form a hairpin structure. Hence, this study offers experimental evidence of the existence of these miRNA. Additionally, approximately half of these predicted miRNAs were found to be in greater abundance in inflammatory exosomes.

Table 4.4. Number of miRNA found to have statistically significant differences in abundance.

Comparison	# miRNAs greater in abundance (*miRBase, Ensembl predicted*)	# miRNAs lower in abundance (*miRBase, Ensembl predicted*)
Infl. MDSC vs. Conv. MDSC	53 *(23, 30)*	110 *(92, 18)*
Infl. exosomes vs. Conv. exosomes	41 *(13, 28)*	59 *(50, 9)*
Conv. exosomes vs. Conv. MDSC	199 *(86, 113)*	106 *(102, 4)*
Infl. exosomes vs. Infl. MDSC	499 *(198, 301)*	84 *(78, 6)*

MicroRNAs carried by exosomes could be delivered to MDSCs (autocrine signaling) or other host cells (paracrine signaling) present in the tumor microenvironment, where they may perform their repressing activity.[145–149,209] In order to better understand the putative biological processes that could be repressed in the receiver cell, mRNA targets were predicted for the top 5 most enriched miRNAs for each comparison of interest. Prediction tools propose putative mRNA targets based on prior knowledge on conserved seed sequences by which a miRNA binds to one or various mRNA transcripts with different degrees of sequence complementarity.[143,210,211] A major disadvantage of using prediction tools is that based on the current knowledge on miRNAs and miRNA-mRNA interactions and the fact that partial complementarity is considered, these tools provide a large number of false positives (~70%), rendering functional analyses difficult to interpret.[210,212]

In this study, the miRWalk database was used for target prediction as it combines the output of 11 different prediction algorithms and a database of validated targets.[177,178] In order to be conservative, we required that at least 2 of the predictive tools must agree on a predicted mRNA target for it to be considered for further analysis. Nevertheless, in this study the data processing of the small RNA provided a list of identified immature miRNAs instead of mature miRNAs, not being able to differentiate

between the -3p and -5p strands. Thus both the -3p and -5p strand were considered when performing target prediction, albeit only one may be present. This could exacerbate the number of false positive and false negative mRNA targets predicted as different strands can have different targets. As an example, miRWalk predicts that miR-146a-3p and miR-146a-5p have 418 and 532 targets, respectively; but only 18 of those targets are actually shared. Aiming to compensate for the limited sensitivity and specificity of predicting mRNA targets, we also based our analysis on thorough literature search to include the most up-to-date validated targets related to our biological study model.

The GO biological processes annotated for the predicted mRNA targets of the top 5 most enriched miRNAs in each comparison is provided in Table 4.5. The GO categories obtained for the conventional vs. inflammatory exosomes comparison were similar between conditions, though a few more categories were found to be significantly enriched under inflammatory conditions. The biological processes targeted in both conditions included: protein phosphorylation, DNA transcription and the transport of molecules. Additionally, apoptosis process, cell adhesion and vesicle-mediated transport were found enriched under inflammatory conditions. In the case of those top 5 most enriched miRNA in exosomes when compared to their parental cells, similar GO biological processes were found to be enriched independently of their inflammation conditions, including again phosphorylation, cell cycle and DNA transcription. Interestingly, protein ubiquitination and blood vessel remodeling were found enriched in inflammatory exosomes.

94

Table 4.5. GO biological processes annotated for the predicted mRNA targets of the top 5 most enriched miRNAs.

Top 5 miRNAs	GO Biological Process	% Fraction of genes	Adj. *p*-value
Exosomes Infl. vs. Conv. - Greater abundance in Conv.			
miR-122	Nervous system development	2.8	3.4×10^{-2}
miR-125b-1	Phosphorylation	4.3	3.6×10^{-3}
miR-143	Protein phosphorylation	3.9	3.8×10^{-2}
miR-9-2	Protein transport	4.0	2.7×10^{-2}
miR-10b	Regulation of transcription, DNA-templated	13.3	6.0×10^{-3}
	Small GTPase mediated signal transduction	2.0	1.2×10^{-2}
	Transcription, DNA-templated	11.3	3.1×10^{-3}
	Transport	10.5	4.0×10^{-2}
Exosomes Infl. vs. Conv. - Greater abundance in Infl.			
miR-6481	Apoptotic process	3.9	3.5×10^{-3}
miR-5627	Homophilic cell adhesion via plasma membrane adhesion molecules	1.5	2.5×10^{-3}
miR-7062	Phosphorylation	4.0	3.8×10^{-2}
miR-3075	Regulation of transcription, DNA-templated	12.8	2.8×10^{-2}
miR-704	Transport	11.2	9.1×10^{-6}
	Vesicle-mediated transport	1.8	7.1×10^{-3}
Exosomes vs. MDSC (Conv.). - Greater abundance in Exosomes Conv.			
miR-467b	Brain development	2.0	3.1×10^{-2}
miR-3470a	Cell cycle	4.4	2.9×10^{-2}
miR-6538	Phosphorylation	4.4	3.2×10^{-2}
miR-592	Positive regulation of transcription from RNA polymerase II promoter	6.8	7.2×10^{-3}
miR-2137	Positive regulation of transcription, DNA-templated	4.7	1.3×10^{-4}
	Regulation of transcription from RNA polymerase II promoter	3.2	2.6×10^{-2}
	Regulation of transcription, DNA-templated	14.5	3.3×10^{-5}
	Transcription, DNA-templated	12.2	8.8×10^{-5}
Exosomes vs. MDSC (Infl.). - Greater abundance in Exosomes Infl.			
miR-6538	Blood vessel remodeling	0.6	2.8×10^{-2}
miR-2137	Cell cycle	4.3	4.4×10^{-3}
miR-467b	Multicellular organism development	6.7	4.7×10^{-3}
miR-146a	Phosphorylation	4.6	1.6×10^{-4}
miR-6367	Positive regulation of transcription from RNA polymerase II promoter	6.7	4.0×10^{-4}
	Positive regulation of transcription, DNA-templated	4.5	2.4×10^{-5}
	Protein phosphorylation	4.3	5.1×10^{-4}
	Protein polyubiquitination	1.2	4.7×10^{-3}
	Regulation of cell cycle	1.2	2.4×10^{-2}
	Regulation of transcription from RNA polymerase II promoter	3.0	6.6×10^{-3}
	Regulation of transcription, DNA-templated	14.5	7.7×10^{-8}
	Transcription, DNA-templated	12.6	7.3×10^{-9}

Many of the miRNAs found have information on validated targets in MDSC and/or other cells in the tumor microenvironment. A list of selected miRNAs with previously validated functions is provided in Table 4.6. In general, processes similar to those seen for the top 5 most enriched miRNAs are targeted, such as DNA transcription, apoptosis and angiogenesis. Notably, miR-126a, miR-146a, miR-155, miR-690 and miR-9-2 were found in greater abundance (2.8 to 18.3-fold) in inflammatory exosomes when compared to their parental cells. Among these miRNAs, mir-146a showed the highest fold change of 18.3-fold. This miRNA binds to the 3'-UTR of TNF receptor-associated factor 6 (TRAF6) and IL-1 receptor-associated kinase 1 (IRAK1), both players in the NF-κB pathway. The repression of these targets by miR-146a negatively regulates NF-κB activation, imposing a stop to inflammation[213] and, in other cases, reducing myeloproliferation, thus suppressing the development of malignant tumors.[214]

MiR-494, miR-223 and miR-690 are relevant for MDSC suppressive function in the tumor microenvironment, as they are capable of affecting the cell cycle, suppressing the differentiation of myeloid cells and increasing MDSC proliferation. In our study, only miR-690 was found to be in greater abundance in exosomes irrespective of their inflammation condition.[12,215] MiR-17 and miR-20a were not enriched in exosomes, these miRNAs affect the release of reactive oxygen species, which is an important mechanism for MDSC suppression of T-cell function.[1] Hence, their transfer from exosomes to other MDSCs would not be desirable. A key miRNA found to be enriched in exosomes is miR-155. If delivered to MDSC, miR-155 is known to cause

Table 4.6. Selected miRNAs found in this study with previously reported functions related to MDSC and/or the tumor microenvironment, including their validated targets. Differences in abundance ($\log_2$(fold-changes)) are shown and significant results (adjusted p-value ≤ 0.05) are marked in **bold**.

miRNA	Putative function	Reported targets	Exo vs. MDSC (Infl.)		Exo vs. MDSC (Conv.)		Exosomes Infl. vs. Conv.		MDSC Infl. vs. Conv.		Ref.
			$\log_2$ (FC)	Adj. p-value	$\log_2$ (FC)	Adj. p-value	$\log_2$ (FC)	Adj. p-value	$\log_2$ (FC)	Adj. p-value	
miR-126a	Promotes angiogenesis	SPRED-1	**1.5**	**4.1×10^{-2}**	**2.8**	**1.2×10^{-3}**	**-1.7**	**2.7×10^{-2}**	-0.4	5.0×10^{-1}	[216]
miR-146a	Regulates inflammation	TRAF6, IRAK1, STAT1	**4.2**	**5.0×10^{-3}**	2.7	8.5×10^{-2}	0.2	8.9×10^{-1}	-1.3	2.8×10^{-1}	[213,214,217,218]
miR-155	Promotes MDSC expansion, T_{Reg} cell survival and production of T_H1. Evades apoptosis	SHIP1, PTEN SOCS1, MAF PU.1, FADD	**1.8**	**1.9×10^{-2}**	**1.8**	**3.2×10^{-2}**	-0.2	7.7×10^{-1}	-0.2	8.0×10^{-1}	[219–224]
miR-17	Reduces MDSC suppressive functions	STAT3	-0.3	5.7×10^{-1}	**-1.6**	**2.3×10^{-2}**	1.0	1.4×10^{-1}	-0.3	5.7×10^{-1}	[225]
miR-199a-1	Reduces NF-κB activity and cytokine production	IKKβ	0.9	3.6×10^{-1}	1.7	6.6×10^{-2}	**-4.2**	**1.7×10^{-4}**	**-3.3**	**4.2×10^{-4}**	[226]
miR-199a-2			**2.0**	**1.1×10^{-2}**	0.2	8.0×10^{-1}	0.5	5.9×10^{-1}	-1.2	1.3×10^{-1}	
miR-20a	Reduces MDSC suppressive functions	STAT3	-0.7	2.4×10^{-1}	-1.4	7.5×10^{-2}	1.0	2.0×10^{-1}	0.3	5.4×10^{-1}	[225]
miR-21a	Promotes MDSC expansion and tumor angiogenesis. Induces IL-10	PTEN, PDCD4	0.1	7.5×10^{-1}	-0.7	1.8×10^{-1}	0.5	3.3×10^{-1}	-0.3	4.0×10^{-1}	[220,227–230]
miR-223	Suppress myeloid cell differentiation, promotes monocyte differentiation	MEF2C, NFIA	0.5	2.3×10^{-1}	-0.6	2.5×10^{-1}	0.9	8.2×10^{-2}	-0.2	6.0×10^{-1}	[231,232]
miR-494	Promotes MDSC accumulation	PTEN	-1.0	1.6×10^{-1}	0.8	4.6×10^{-1}	-0.3	7.5×10^{-1}	1.5	7.3×10^{-2}	[233]
miR-690	Suppress myeloid cell differentiation and promotes MDSC expansion	C/EBPα	**2.0**	**4.9×10^{-2}**	**3.2**	**1.3×10^{-2}**	0.0	9.7×10^{-1}	1.2	1.4×10^{-1}	[215]
miR-9-2	Promotes angiogenesis	NF-κB1, RUNX1	**2.3**	**3.0×10^{-2}**	-0.8	4.4×10^{-1}	**-5.1**	**8.2×10^{-5}**	**-8.2**	**2.1×10^{-9}**	[234]
miR-98	Evades apoptosis	Fas, c-Myc	-0.1	9.0×10^{-1}	0.9	2.5×10^{-1}	1.0	1.7×10^{-1}	**1.9**	**2.0×10^{-3}**	[235,236]

Note: C/EBPβ, CCAAT/enhancer binding protein-β; FADD, Fas-Associated protein with Death Domain; IKKβ, Inhibitor of nuclear factor kappa-B kinase subunit beta.; IRAK, IL-1R-associated kinase; MAF, macrophage-activating factor; MEF2C, myeloid ELF1-like factor 2C; NFIA, nuclear factor I/A; NF-κB, nuclear factor-κB subunit 1; PDCD4, programmed cell death 4; PTEN, phosphatase and tensin homologue; RUNX1, runt-related transcription factor 1; SPRED-1, Sprouty Related EVH1 Domain Containing 1; STAT, Signal transducer and transcription activator; TRAF6, TNFR-associated factor 6; SHIP1, SH2-domain-containing inositol-5-phosphatase 1.

MDSC expansion and production of IL-10.[220] The increased production of IL-10 by MDSC has been associated with the polarization of macrophages to a tumor-promoting phenotype (M2) and the induction of regulator T cells (T_{Reg}).[1,237] The induction of T_{Reg} is important as they can also suppress tumor immunity.[238] If miR-155 was delivered to T_{Reg}, it has been reported that it targets SOCS-1 increasing cell survival.[238,239] Therefore, exosomes carry a selective miRNA cargo that could be important in tumor microenvironment cell-cell communication if transferred.

4.3.4. Protein profiles and functional analyses

The protein cargo of MDSCs and exosomes shed by MDSCs from tumor-bearing mice under two inflammation conditions was interrogated and the effect of inflammation on protein abundance determined, as previously reported.[3,84] The aim of carrying out these quantitative comparisons is to deepen our knowledge on the identity and abundance differences of proteins carried by MDSC-derived exosomes and their parental cells by extending our inventory using the orbitrap Fusion Lumos, a state-of-the-art mass spectrometer. Additionally, this work includes the first global relative quantitation between MDSC-exosomes and their parental cells providing novel information. Three biological replicates per inflammation condition were analyzed by LC-MS/MS. Five technical replicates per sample were injected in order to achieve in-depth protein identifications. A total of 60 data files were analyzed in PepArML using the search engines OMSSA, X!Tandem and MSGF+, as stated in the "Materials and Methods" section.

4.3.4.1. MDSC-derived exosome protein cargo

Exosomes shed from MDSC were found to carry 814 (6,030 unshared peptides) and 1,189 (7,702 unshared peptides) proteins identified with protein FDR of at most 1% for conventional and inflammatory exosomes, respectively. These analyses amounted to a combined total of 1,249 proteins, with 61% overlap between conditions (see Figure 4.6). A complete list of proteins identified is provided in Appendix 4.1. Interestingly, 754 (93%) of the proteins identified in conventional exosomes were also present in inflammatory exosomes. However, 435 (35%) of the total proteins identified were found exclusively in inflammatory exosomes. Since the same amount of total protein was injected per condition, differences in protein identifications can be due to the actual absence of some proteins in one of the conditions or, more likely due to differences in protein abundances between conditions.
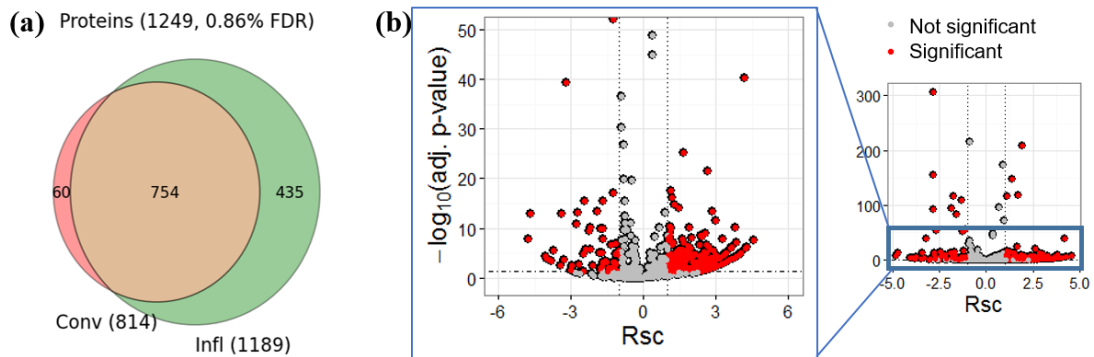
Figure 4.6. Characterization of exosomal proteins. (a) Venn diagram comparing the proteins identified in conventional and inflammatory exosomes, showing a large overlap between conditions. (b) Differences in abundance are shown by plotting $-\log_{10}$(adj. *p*-value) vs. protein ratio from spectral counts (Rsc). Statistically significant results present a $-\log_{10}$(adj. *p*-value) $\geq$ 1.3 corresponding to an adjusted *p*-value $\leq$ 0.05 (FDR$\leq$5%), and are marked by an horizontal dashed line. Two fold-change in abundance (Rsc between 1 and -1) are marked by dotted vertical lines.

Several proteins that are usually regarded as exosome markers were identified including CD9; heat shock proteins Hsp70 protein-4, Hsp cognate 71KDa, Hsp90α and Hsp90β; MHC II molecules; and components of the ESCRT endosomal sorting machinery such as Alix (programmed cell death 6-interacting protein) and vacuolar protein sorting-associated proteins Vps25, Vps4B and Vps37B.[89,91,92] Additional proteins observed included annexins (A1-A7 and A11), flotillin-1, and small GTP-ases-Rab proteins. A suite of integrins, tetraspanins and other surface proteins were also identified and are listed in Table 4.7. Many of these proteins have been previously reported to be present in exosomes shed by MDSC,[3,85,240] including 5 glycoproteins that Chauhan *et al.* identified to be on the surface of exosomes using a chemical method that selectively enriches for exosome surface N-linked glycoproteins.[240]

Out of the 1,249 proteins identified 353 were assigned to the GO category "nucleic acid binding", including several histone isoforms. Twenty-four ribosomal proteins were identified in exosomes, 12 proteins belonging to the 40S subunit and 12 proteins to the 60S subunit of the ribosome. Almost half of them were found in inflammatory exosomes only. Considering that exosomes do not carry rRNA, as shown in the previous section, the presence of a whole ribosome is not expected. However, many proteins related to translation, including 21 eukaryotic translation initiation factors, 7 elongation factors and 19 aminoacyl-tRNA synthetases were found. In the cell, the function of aminoacyl-tRNA synthetases is to attach a specific amino acid to tRNA for mRNA translation. In our case, exosomes were found to contain aminoacyl-tRNA synthetases for all the 20 amino acid residues except Q in inflammatory

exosomes, and Q, M and I in conventional exosomes. Strikingly, proteins that have been proposed to sort miRNA into extracellular vesicles, such as heterogeneous ribonucleoproteins (hnRNP) and protein argonaute-2 (AGO-2),[151,241] were also carried by exosomes. The role of these proteins could be related to cell-cell communication, sorting a set of miRNAs into the exosome and stabilizing them for future delivery into receiver cells, or simply a sorting mechanism to remove unwanted miRNAs from the cell.

Furthermore, several proteins related to proteolysis were identified. Thirty-four of the proteins found were part of the proteasome including: 7 α-subunits and 6 β-subunits of the 20S core, immunoproteasome subunits (proteasome subunit beta type-9, -8 and -10), proteasome regulators (14 proteins of the 19S regulatory particle, proteasome-associated protein ECM29, proteasome activator complex subunit 1 (PA28α) and subunit 2 (PA28β)).[242,243] In the cell, the immunoproteasome and PA28αβ are in part responsible for producing MHC class I antigen peptides.[244,245] Adams has recently reported the presence of 30 proteasome proteins in inflammatory exosomes, from which 9 were found to be ubiquitinated.[246] Lai *et al.* reported the presence of the 20S proteasome and immunoproteasome inside exosomes and associated to exosomes shed from mesenchymal stem cells collected from plasma and demonstrated that it was functional in intact exosomes.[247]

Table 4.7. List of selected surface proteins identified in exosomes including their ratio from spectral counts (Rsc). Statistically significant differences in abundance correspond to observed fold-change ≥ 2 (Rsc ≥ 1 or Rsc ≤ -1) and Fisher's Exact test FDR $\leq 5\%$, are shown in **bold**.

Protein Accession #	Description	Rsc (Infl. vs. Conv)	FDR
	Tetraspanins		
P40240	CD9 antigen	-0.1	9.0×10^{-1}
Q8QZY6	Tetraspanin-14	-0.2	1.0
Q8R2S8	CD177 antigen	0.0	8.8×10^{-1}
	Integrins		
E9Q604	Integrin alpha-M (CD11b)	-0.8	4.0×10^{-16}
O54890	Integrin beta-3 (CD61)	-0.3	6.6×10^{-1}
P09055	Integrin beta-1 (CD29)	**-1.3**	**1.9×10^{-2}**
P11835	Integrin beta-2 (CD18)	-0.8	3.3×10^{-13}
P24063	Integrin alpha-L (CD11a)	0.8	8.2×10^{-2}
Q61739	Integrin alpha-6 (CD49f)	-1.4	2.9×10^{-1}
Q62469	Integrin alpha-2 (CD49b)	-0.3	1.0
Q9QUM0	Integrin alpha-IIb (CD41)	-0.6	2.2×10^{-5}
	Other Surface Proteins		
A8E0Y8	Immunoglobulin superfamily member 2 (CD101)	-2.1	5.2×10^{-1}
O09126	Semaphorin-4D (CD100)	-1.1	4.6×10^{-1}
O35598	Disintegrin and metalloproteinase domain-containing protein 10 (CD156c)	-0.6	6.0×10^{-1}
O35930	Platelet glycoprotein Ib alpha chain (CD42b)	**-1.7**	**3.2×10^{-16}**
O54990	Prominin-1 (CD133)	1.5	3.1×10^{-1}
P06800	Receptor-type tyrosine-protein phosphatase C (CD45)	0.4	2.0×10^{-1}
P09581	Macrophage colony-stimulating factor 1 receptor (CD115)	-1.0	1.4×10^{-1}
P0CW02	Lymphocyte antigen 6C1 (Ly6C)	0.8	3.4×10^{-1}
P10810	Monocyte differentiation antigen CD14	0.0	7.5×10^{-1}
P10852	4F2 cell-surface antigen heavy chain (CD98)	**3.1**	**4.0×10^{-3}**
P18337	L-selectin (CD62L)	0.1	7.7×10^{-1}
P18572	Isoform 2 of Basigin (CD147)	-1.7	5.8×10^{-2}
P19973	Lymphocyte-specific protein 1	-0.8	1.5×10^{-1}
P27931	Interleukin-1 receptor type 2 (CD121b)	**2.3**	**2.4×10^{-7}**
P35343	C-X-C chemokine receptor type 2 (CD182)	0.7	4.6×10^{-1}
P35441	Thrombospondin-1 (TSP-1)	**-1.3**	**9.4×10^{-111}**
P35461	Lymphocyte antigen 6G (Ly6G)	-0.3	1.0
P97484	Leukocyte immunoglobulin-like receptor subfamily B member 3 (PIR-B)	2.2	7.2×10^{-2}
Q01102	P-selectin (CD62P)	**-1.8**	**8.8×10^{-3}**
Q2VLH6	Scavenger receptor cysteine-rich type 1 protein M130 (CD163)	**2.9**	**1.0×10^{-2}**
Q60767	Lymphocyte antigen 75 (CD205)	0.2	5.1×10^{-1}
Q64455	Receptor-type tyrosine-protein phosphatase eta (CD148)	-1.3	5.8×10^{-2}
Q9QWK4	CD5 antigen-like (CD5L)	**-2.1**	**2.8×10^{-3}**
Q9QZU3	Platelet glycoprotein V (Fragment) (CD42d)	**-2.2**	**9.6×10^{-11}**

The differences in protein abundance related to heightened inflammation were determined by spectral counting. Spectra were counted using an in-house software and a total of 98,714 and 120,844 spectral counts were obtained for conventional and inflammatory exosomes, respectively. The median spectral counts per protein observed for exosomes was ~20, with ~200 proteins presenting more than 100 counts. The distribution of counts observed per sample type is shown in Appendix 4.6. Ratios from spectral counts (Rsc) were estimated as stated in the "Materials and Methods" section and statistically significant differences in abundance were defined as cases that presented a fold-change ≥ 2 (Rsc ≥ 1 or Rsc ≤ -1) and a Fisher's Exact test FDR ≤ 5%. Rsc varied from -4.8 to 4.6, corresponding to fold-changes from 28-fold decrease to 24-fold increase in inflammation, with a median Rsc value of 0.6. Out of the 1,249 proteins identified, 69 (6%) proteins were found to be in greater abundance in conventional exosomes (Rsc ≤-1) and 347 (28%) in inflammatory exosomes (Rsc ≥1). The relative quantitation results are visualized in a volcano plot in Figure 4.6. For each condition, functions and pathways of the enriched proteins were investigated based on the generic GO and PIR GO slims, and KEGG and Reactome databases.

In the cases of proteins with greater abundance in inflammatory exosomes, the GO molecular function "nucleotide binding" (6.5-fold, p-value $= 5.7 \times 10^{-6}$) was significantly enriched. Sixteen ribosomal proteins were found to be 2 to 16-fold more abundant under inflammation conditions. Many proteins known to participate in mRNA translation were also in higher abundance under inflammation conditions. Regarding proteins that may be related to miRNA sorting, the abundances of 6 out of

the 15 hnRNP were found to be 2.3 to 8-fold higher under inflammation conditions. AGO-2, a protein which is part of the RNA-induced silencing complex,[248] was found 2.8-fold higher in abundance. Additionally, proteins with protease activity, caspase-1, caspase-6, cathepsin B, cathepsin G and 3 regulatory subunits of the proteasome were found to be 2.1 to 13-fold in greater abundance in inflammatory exosomes. Interestingly, the thiol protease inhibitor Stefin-2 was also found 2.1-fold more abundant in inflammation. This protein inhibits cysteine cathepsins, such as cathepsin B, [249] also found to be enriched in inflammatory exosomes.

In the case of conventional exosomes, the proteins that were enriched were significantly represented in the KEGG pathway "complement and coagulation cascades" (6.7-fold, p-value = 0.007). This pathway has been linked to inflammation and cancer, and can assist in tumor angiogenesis and metastasis.[250,251] The statistically significant GO molecular function "oxygen binding" (18-fold, p-value = 5.6×10^{-4}) was found to be enriched in conventional exosomes, due to the identification of hemoglobin subunits. Although these proteins are not expected in MDSC, their presence could be explained by the surface protein CD163, which is known for its ability to bind and internalize complexes of hemoglobin-haptoglobin.[252] The surface proteins CD42b, CD42d and CD62P, found to be enriched in conventional exosomes, are platelet surface markers previously reported to be abundant in MDSC from 4T1 tumor-bearing mice.[83]

Relative quantitation results were summarized in Table 4.8 for a selected set of proteins, which are of particular importance as they have been previously reported to

increase MDSC accumulation, expansion, and/or promote MDSC suppressive activity. Functional redundancy is observed, as different proteins may produce a similar outcome through diverse mechanisms. Additionally, note that biologically active proteins such as S100A8, S100A9 and HMGB1 did not present differences in abundance. The lack of difference in abundances for S100A8 and S100A9 in exosomes from animals with more and less inflammation does not match our results at the proteoform level shown in the previous chapter. However, this type of behavior can be explained by the fact that our bottom-up proteomics approach do not take into account modified peptides, other than those carrying an oxidation on Met. Hence, the resultant relative quantitation by bottom-up analysis could be considered an "average" of all existing proteoforms. Unfortunately, in top-down proteomics not all proteoforms present are always identified, as a result estimating an "average" relative quantitation for comparison purposes can be challenging. Additionally, in the case of S100A8, it is clear that one proteoform (loss of initial Met), which showed no differences in abundance by top-down proteomics, corresponded to the majority of the TIC signal. Thus, the effect of minor proteoforms may not be reflected on our bottom-up analysis. This observation demonstrates that both approaches offer relevant and complementary information.

Table 4.8. List of proteins found in exosomes known to play a role on MDSC accumulation, expansion and/or suppression activity in the tumor microenvironment. Statistically significant (adjusted *p*-value ≤ 0.05) estimated fold-changes when comparing inflammatory and conventional exosomes are shown.

Protein Accession #	Description	Fold-change	Reported function	Ref.
Greater abundance in inflammatory exosomes				
Q61176	Arginase1 (ARG1)	13.2	MDSC suppression of T cells	[1]
P10852	CD98	8.6	Membrane receptor that interacts with galectin-3, present in MDSC and their exosomes. Chemoattractant for MDSCs and macrophages.	[253, 254]
Q2VLH6	CD163	7.5	Marker of tumor-associated macrophages (M2). Associated to tumor progression and poor prognosis in breast and colorectal cancer.	[255,256]
P34884	Macrophage migration inhibitor factor (MIF)	2.8	Cytokine. Increases MDSC expansion, promotes tumor growth.	[257,258]
Greater abundance in conventional exosomes				
Q9Z126	Platelet factor 4 (PF-4)	2.3	Chemokine, chemoattractant for monocytes and neutrophils.	[259]
P35441	Thrombospondin-1 (TSP-1)	2.5	Receptor to leukocyte surface antigen CD47. Chemotactic for MDSC. Promotes tumor progression.	[240]
P09055	CD29	2.4	Suppress Tcell proliferation, acting together with CD11b and CD18.	[260]
No significant difference in abundances				
P27005	S100A8	1.1	Chemotactic for MDSC. Stimulate MDSCs accumulation and suppressive activity.	[3]
P31725	S100A9	1.6	Chemotactic for MDSC. Stimulate MDSCs accumulation and suppressive activity.	[3]
P63158	High-mobility-group box 1 (HMGB1)	1.5	Pro-inflammatory protein secreted by MDSCs. Increase MDSC production, accumulation and suppressive activity, promoting tumor progression.	[261]

4.3.4.2. MDSC protein cargo

The protein content of parental MDSCs that were matched to their exosomes was interrogated in order to offer a complete context for further comparisons. A combined total of 1,423 proteins were identified from which 1,141 proteins (7,162 unshared peptides) were found in conventional MDSC and 1,255 (7,556 unshared peptides) in inflammatory, with protein FDR of at most 1%. A significant overlap

(68%) in proteins identified between conditions was observed as shown in Figure 4.7. A complete list of proteins identified is provided in Appendix 4.1. Several characteristic MDSC surface markers (CD11b, Ly6C and Ly6G) and an intracellular marker (ARGI) were identified as expected.[1,262] Additionally, several cell-surface markers were identified. Unsurprisingly many of the identified proteins are shared with those found in their exosomes (see Table 4.9). Out of the 28 CD proteins listed in Table 4.9, 19 were previously reported to be found on the surface of MDSC by Chauhan *et al*.[240] supporting the conclusion that our analysis was able to capture part of the surface proteins.

As stated previously, relative quantitation was performed by spectral counting. A total of 103,677 and 116,751 spectral counts were obtained for conventional and inflammatory MDSCs, respectively. The median spectral count observed for MDSC was also ~20, with ~220 proteins presenting more than 100 counts (see Appendix 4.6.). Rsc varied from -6.8 to 5.4, corresponding to a 111-fold decrease and a 42-fold increase in inflammatory MDSC, with a median Rsc value of 0.2. Based on our quantitation criteria, 123 (9%) proteins were found in greater abundance in conventional MDSC (Rsc ≤-1) and 231 (16%) in inflammatory MDSC (Rsc ≥1), as shown in Figure 4.7. Note that a smaller number of proteins exhibited differences in abundance when compared to the cargo of inflammatory or conventional exosomes.

Table 4.9. List of selected cell-surface proteins identified in MDSCs. Statistically significant differences in abundance present fold-change ≥ 2 (Rsc ≥ 1 or Rsc $\leq$ -1) with Fisher's Exact test FDR $\leq 5\%$ and are shown in **bold**.

Protein Accession #	Description	Rsc (Infl. vs. Conv.)	FDR
	Tetraspanin		
Q8R2S8	CD177 antigen	**-1.3**	**1.9×10^{-42}**
	Integrins		
E9Q604	Integrin alpha-M (CD11b)	**-1.1**	**1.1×10^{-36}**
O54890	Integrin beta-3 (CD61)	-0.4	1.0
P11835	Integrin beta-2 (CD18)	**-1.5**	**5.4×10^{-26}**
P24063	Integrin alpha-L (CD11a)	**-1.6**	**3.0×10^{-10}**
Q9QUM0	Integrin alpha-IIb (CD41)	**-1.1**	**1.1×10^{-11}**
	Other surface proteins		
A2AFG7	Neural cell adhesion molecule L1 (CD171)*	**-5.0**	**1.1×10^{-10}**
A8E0Y8	Immunoglobulin superfamily member 2 (CD101)	-1.5	6.9×10^{-1}
O35930	Platelet glycoprotein Ib alpha chain (CD42b)	**1.2**	**3.6×10^{-2}**
P06800	Receptor-type tyrosine-protein phosphatase C (CD45)	**-1.3**	**1.6×10^{-23}**
P0CW02	Lymphocyte antigen 6C1 (Ly6C)	**-5.7**	**2.1×10^{-17}**
P10810	Monocyte differentiation antigen CD14	-0.6	2.9×10^{-1}
P10852	4F2 cell-surface antigen heavy chain (CD98)	0.0	1.0
P13595	Neural cell adhesion molecule 1 (CD56)	-1.9	4.0×10^{-1}
P17047	Lysosome-associated membrane glycoprotein 2 (CD107b)	-1.3	2.2×10^{-1}
P18572	Isoform 2 of Basigin (CD147)*	**-3.9**	**1.8×10^{-4}**
P19973	Lymphocyte-specific protein 1	-0.6	3.1×10^{-7}
P27931	Interleukin-1 receptor type 2 (IL-1r2, CD121b)	**3.3**	**2.3×10^{-3}**
P31809	Carcinoembryonic antigen-related cell adhesion molecule 1 (CD66a)*	**-3.2**	**1.0×10^{-2}**
P35441	Thrombospondin-1 (TSP-1)	-0.2	5.0×10^{-1}
P35461	Lymphocyte antigen 6G (Ly6G)	**-1.8**	**6.8×10^{-6}**
P40223	Granulocyte colony-stimulating factor receptor (CD114)	-1.9	4.0×10^{-1}
P55772	Ectonucleoside triphosphate diphosphohydrolase 1 (CD39)*	**-3.7**	**7.2×10^{-4}**
P97370	Sodium/potassium-transporting ATPase subunit beta-3 (CD298)	-1.1	7.8×10^{-1}
P97797	Tyrosine-protein phosphatase non-receptor type substrate 1 (CD172a/Sirp-α)*	**-3.5**	**2.6×10^{-3}**
Q01102	P-selectin (CD62P)	-0.6	7.1×10^{-1}
Q60767	Lymphocyte antigen 75 (CD205)	-0.5	7.2×10^{-1}
Q64277	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2 (CD157)*	**-3.9**	**1.8×10^{-4}**
Q64455	Receptor-type tyrosine-protein phosphatase eta (CD148)	-0.9	1.5×10^{-1}
Q6SJQ7	CMRF35-like molecule 1 (CD300f)	-0.3	9.9×10^{-1}
Q9QUN7	Toll-like receptor 2 (TLR-2, CD282)	-1.9	4.0×10^{-1}
Q9QZU3	Platelet glycoprotein V (Fragment) (CD42d)	**3.1**	**6.5×10^{-3}**

Note: * Identified only in conventional MDSC
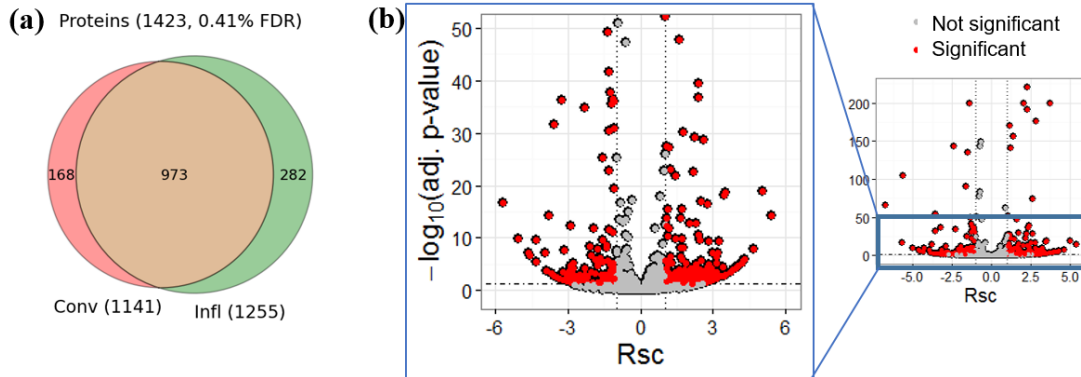
Figure 4.7. Characterization of MDSC proteins. (a) Venn diagram comparing the proteins identified in conventional and inflammatory MDSC. (b) Differences in abundance are shown by plotting $-\log_{10}$(adj. *p*-value) vs. protein ratio of spectral counts (Rsc). Statistically significant results present $-\log_{10}$(adj. *p*-value) ≥ 1.3, corresponding to an adjusted *p*-value ≤ 0.05, and are marked by an horizontal dashed line. Two fold-change in abundance (Rsc between 1 and -1) are marked by dotted vertical lines.

In the case of proteins with greater abundance in inflammatory MDSC, GO molecular functions that take part in metabolic processes "transferase activity" (2.6-fold, *p*-value = 0.045), "isomerase activity" (2-fold, *p*-value = 0.030) and "ligase activity" (2.5-fold, *p*-value = 0.0003) were significantly enriched. This is expected as more MDSCs are produced under inflammatory conditions.[5] The largest fold-change observed in inflammatory MDSCs was for thiol protease inhibitor Stefin-3, with a 42-fold increase in abundance. Regarding the chemotactic S100A8/A9 heterodimer known to regulate MDSC accumulation and suppressive activity, S100A8 was found in 4.2-fold greater abundance in inflammatory MDSC, but S100A9 showed no significant changes in abundance with a marginal 1.8-fold increase (*p*-value = 8×10^{-63}). Additionally, the cytokine MIF, which as stated previously could be in part responsible for the accumulation of MDSC in the tumor microenvironment, was found to be in 13-fold greater abundance. Proteins known to enhance MDSC suppressive activity,[1] such

as ARG1 and STAT3 (signal transducer and activator of transcription 3) were also found to be in greater abundance when compared to conventional MDSC, observing a 8- and 2.4-fold increase, respectively.

Many proteins found in greater abundance in conventional MDSCs were cell surface proteins, with the highest fold changes of 52-fold estimated for the MDSC marker LyC6. The metalloproteinase-9 (MMP9), known to increase MDSC production and promote angiogenesis,[263] was found 3-fold in greater abundance in conventional MDSC.

4.3.4.3. Comparing the protein cargo of exosomes and their parental cells

Differences between MDSC-derived exosomes and their parental cells were evaluated by performing global parsimony of all exosome and all MDSC samples, irrespective of their inflammation condition. This comparison provides novel information as it represents the first quantitative comparison between the protein cargo of MDSC-derived exosomes and their parental cells. A total of 1,726 proteins were identified with protein FDR of at most 1%, from which 1,256 corresponded to exosomes and 1,434 to MDSC. Figure 4.8 shows that more than half of the proteins (58%) were identified in both MDSC and exosomes. A complete list of proteins identified is presented in Appendix 4.1. Based on extensive literature we expect exosomes to contain a selective cargo. In order to investigate the major differences between the protein cargo of exosomes and parental cells, relative quantitation was performed and functional categories and pathways were assigned to those proteins

found to be in greater abundance in exosomes. A total of 371 proteins (21%) were found in greater abundance in exosomes (Rsc ≤-1) and 612 proteins (35%) were found in greater abundance in MDSC (Rsc ≥1) shown as a volcano plot in Figure 4.8. Based on the GO cellular compartment category assignments, proteins found in greater abundance in exosomes were localized to the "extracellular region", "extracellular space" and "cell surface". Significantly enriched GO molecular functions included "antigen binding", "signal transducer" and "peptidase activity" (see Table 4.10).
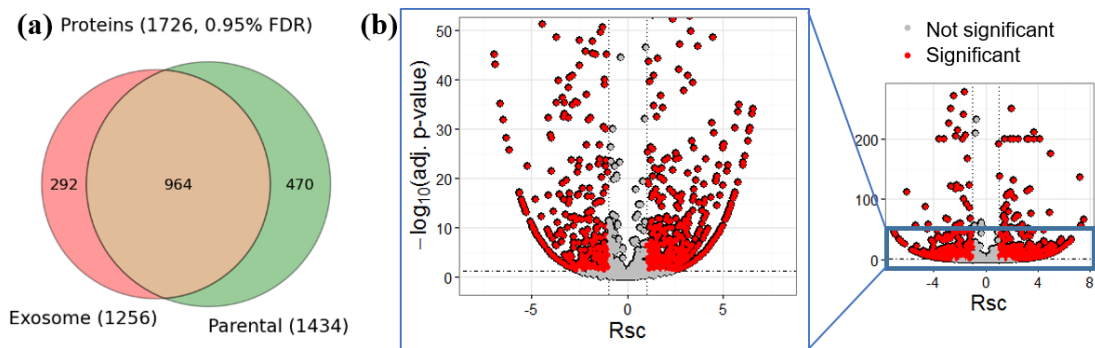
Figure 4.8. (a) Comparison of proteins identified in MDSC vs. MDSC-derived exosomes shown in a Venn diagram. (b) Differences in abundance are shown by plotting –$\log_{10}$(adj. *p*-value) vs. protein ratio of spectral counts (Rsc). Statistically significant results have –$\log_{10}$(adj. *p*-value) ≥ 1.3 corresponding to an adjusted *p*-value ≤ 0.05, and are marked by an horizontal dashed line. Two fold-change in abundance (Rsc between 1 and -1) are marked by dotted vertical lines.

Table 4.10. GO categories and KEGG pathways enriched for proteins in greater abundance in exosomes when compared to their parental cells.

Database	Pathway	Number of proteins	Fold Enrichment	Adjusted *p*-value
GO Cellular Compartment	Extracellular region	216	1.3	2.9×10^{-8}
	Extracellular space	64	1.6	3.5×10^{-5}
	Cell surface	35	1.5	1.7×10^{-2}
GO Molecular Function	Antigen binding	11	2.4	3.7×10^{-2}
	Signal transducer activity	30	1.7	3.7×10^{-2}
	Peptidase activity	55	1.6	6.3×10^{-3}
	Transferase activity	86	1.3	4.4×10^{-2}
KEGG	Amino sugar and nucleotide sugar metabolism	11	2.8	1.8×10^{-2}
	Complement and coagulation cascades	14	3.1	1.0×10^{-3}
	Cysteine and methionine metabolism	8	3.1	4.0×10^{-2}
	Pentose phosphate pathway	11	2.7	2.8×10^{-2}
	Proteasome	19	2.6	1.0×10^{-3}
	Purine metabolism	18	2.2	1.8×10^{-2}

Immunoglobulins were identified among the "antigen binding" proteins. Out of the 8 proteins found, 5 were detected exclusively in exosomes (see Table 4.11). The protein Ig mu chain C region was found to be 64-fold higher in abundance in exosomes. Note that immunoglobulins have been previously identified in MDSC-derived exosomes.[3] The enrichment for molecular functions related to proteolysis in the exosomes may correlate with the higher abundance of proteins belonging to the proteasome. Additionally, caspase-3 and -6 were found to be 2.2 and 3.4-fold more abundant in exosomes, respectively. Sixteen cell surface proteins shown in Table 4.12 were found to be enriched in exosomes, including the exosome marker CD9 which was found to be 89-fold more abundant. The "complement and coagulation cascade" remains an enriched pathway (see Table 4.10) when comparing exosomes and their parental cells.

Table 4.11. Immunoglobulins found enriched in exosomes. Note: * Identified only in exosomes

Protein Accession #	Description	Rsc MDSC vs. Exo	FDR
P01878	Ig alpha chain C region*	-3.7	1.7×10^{-4}
P06336	Ig epsilon chain C region*	-3.9	4.5×10^{-5}
P01868	Ig gamma-1 chain C region secreted form	-4.1	1.3×10^{-9}
P01812	Ig heavy chain V region MOPC 173*	-3.0	8.1×10^{-3}
P01786	Ig heavy chain V region MOPC 47A*	-3.6	3.2×10^{-4}
P01837	Ig kappa chain C region	-2.9	9.7×10^{-13}
P01639	Ig kappa chain V-V region MOPC 41*	-3.8	8.6×10^{-5}
P01872	Ig mu chain C region	-6.0	2.3×10^{-112}

Table 4.12. Surface proteins found enriched in exosomes.

Protein Accession #	Description	Rsc MDSC vs. Exo	FDR
Q9QWK4	CD5 antigen-like	-4.4	1.1×10^{-7}
P40240	CD9 antigen	-6.5	1.7×10^{-32}
O35598	Disintegrin and metalloproteinase domain-containing protein 10 (CD156c)	-5.5	7.3×10^{-16}
Q61739	Integrin alpha-6 (CD49f)	-2.2	8.6×10^{-3}
P09055	Integrin beta-1 (CD29)	-3.5	3.2×10^{-10}
O54890	Integrin beta-3 (CD61)	-3.2	2.7×10^{-20}
P27931	Interleukin-1 receptor type 2 (CD121b)	-2.3	6.1×10^{-10}
P18337	L-selectin (CD62L)	-3.4	1.2×10^{-3}
Q60767	Lymphocyte antigen 75 (CD205)	-1.4	1.1×10^{-4}
P09581	Macrophage colony-stimulating factor 1 receptor (CD115)	-5.0	3.7×10^{-11}
O35930	Platelet glycoprotein Ib alpha chain (CD42b)	-2.6	5.4×10^{-36}
Q9QZU3	Platelet glycoprotein V (CD42d)	-3.0	9.6×10^{-18}
Q2VLH6	Scavenger receptor cysteine-rich type 1 protein M130 (CD163)	-3.2	4.2×10^{-3}
O09126	Semaphorin-4D (CD100)	-3.7	1.6×10^{-4}
Q8QZY6	Tetraspanin-14	-5.4	1.4×10^{-15}
P35441	Thrombospondin-1 (TSP-1)	-3.5	~ 0

The pro-inflammatory proteins S100A8 and S100A9 were 5.8 and 3.6-fold more abundant in MDSC, respectively. HMGB1 showed no difference in abundance. The cytokine MIF and the chemokine PF-4 were found 2.2 and 4.8-fold more abundant in exosomes, respectively. The transforming growth factor-β-1 (TGF-β1) showed no difference in abundance in MDSC or MDSC-derived exosomes when comparing inflammation conditions. However, when comparing all exosomes and all MDSC, TGF-β1 was found 4.3-fold higher in abundance in exosomes (p-value = 8.4×10^{-14}). Tumor cells and MDSC have been reported to secrete TGF-β1 into the tumor microenvironment. TGF-β1 secreted by MDSC together with IL-10 or IL-6 can induce T_{Reg} or Th17 cells, respectively.[1] Additionally, Xiang *et al.* demonstrated that exosomes from 4T1 tumors that were enriched in TGF-β1 and PGE$_2$ were able to induce MDSC accumulation in the tumor microenvironment.[264]

4.3.5. An example: integrating the miRNA, mRNA and protein cargo profiles from MDSC-derived exosomes to propose putative effects on signaling pathways related to cancer.

The KEGG pathway "pathways in cancer", shown in Figure 4.9, was selected to visualize the putative combined effect of the exosome cargo. This pathway integrates 10 signaling pathways (MAPK, cAMP, Wnt, VEGF, TGF-β, p53, mTOR, PI3K/Akt, JAK-STAT and PPAR signaling pathways), cell-cycle and apoptosis, many of which were previously discussed to be important for MDSC suppressive activity and tumor progression. The most enriched miRNAs, mRNAs and proteins in exosomes compared

to their parental cells were marked over the pathway map (Figure 4.9). The difference in coverage depth between our proteomics and transcriptomics analyses is quickly noted in Figure 4.9, as a larger number of mRNA transcripts than proteins were identified. This can be explained by the current bias of discovery shotgun proteomics towards the identification of abundant proteins. As observed, the mRNA and protein cargos often overlap, which could be seen as a redundant approach to target the pathway of interest. Additionally, most of the enriched miRNAs present are not accompanied by their target mRNA or their encoded protein. In most cases the effect that the exosome cargo would have over the pathways is not contradictory, which means that the mRNAs, miRNAs and proteins carried are not often activating and suppressing the same pathway.

In Figure 4.9, it can be observed that the MAPK, VEGF and Wnt pathways are activated and could promote cell proliferation, angiogenesis and tissue invasion. Additionally, miR-98 and miR-155 could be in part responsible for evading cell apoptosis. Several miRNAs, mRNA and proteins that would ultimately block cell differentiation are also present. All this information taken together supports the observation that MDSC-derived exosomes carry a selective protein, mRNA and miRNA cargo and that cell-cell communication through exosomes could contribute significantly to tumor progression and metastasis.
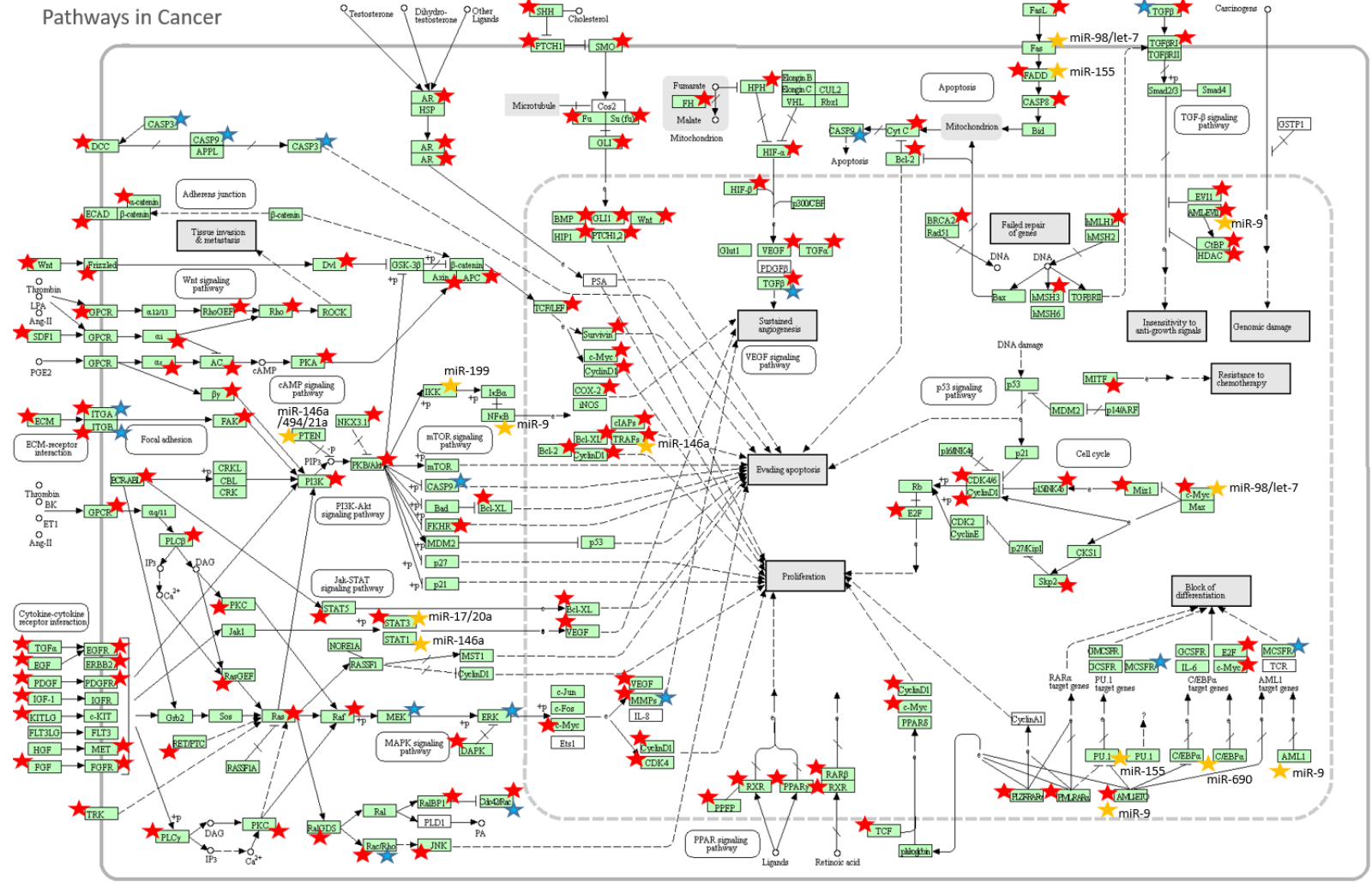
Figure 4.9. Map of "pathways in cancer" (KEGG, Kanehisa Laboratories, http://www.genome.jp/kegg-bin/show_pathway?mmu05200) overlaid with the mRNA (red stars), miRNA (yellow stars) and proteins (blue stars) found to be enriched in exosomes vs. parental cells. A detailed list of the map is found in Appendix 4.7. Boxes colored in green are genes present in the genome.

4.4. Summary

The MDSCs and MDSC-derived exosome protein, miRNAs and mRNAs cargoes were successfully interrogated using bottom-up proteomics and next-generation sequencing. We first demonstrated that exosomes shed from MDSC carry mRNAs and miRNAs. All cargoes (mRNA, miRNA and proteins) were found to be selectively enriched when compared to their parental cells. Strikingly, this study provided experimental evidence of ~624 Ensembl predicted miRNAs; many of them present in higher abundance in exosomes than in their parental cells. Additionally, relative quantitation of the protein, mRNA and miRNA cargoes revealed quantitative differences between inflammation conditions.

Several biological processes such as promoting cell proliferation, blocking cell differentiation and promoting angiogenesis were found to be enriched under heightened inflammation. However, a limitation of performing functional analysis on miRNA target genes, is that miRNAs can have more than one target and repression will most likely happen to those mRNA transcripts present in higher abundance in the receiver cell. Hence, it is difficult to predict the combined effect of the exosome cargo in the tumor microenvironment, requiring future biological assays possibly considering different types of receiver cells.

Chapter 5: Conclusion and prospectus

Myeloid derived suppressor cells (MDSC) are an immunosuppressive heterogeneous population of immature monocytes present in most cancer patients[1] known to release extracellular vesicles, called exosomes, into the tumor microenvironment. Exosomes act as a mean of communication with other cells (paracrine activity) and stimulate the expansion and accumulation of MDSC in the tumor microenvironment (autocrine activity).[3] Unfortunately, the presence of MDSCs is known hinder immunotherapy as they suppress both adaptive and innate immune responses.[1] Moreover, many cancers are accompanied by inflammation, a processes that further intensify MDSC suppressive activity causing an aggressive tumor progression and metastasis.[4–7]

This thesis work addressed the cargo of MDSC and their immunosuppressive exosomes, aiming to gather knowledge that can offer insights on the mechanisms by which MDSCs cause immune suppression, focusing on the role of exosomes as intercellular communication mediators in the tumor microenvironment. In order to achieve our objective a well-established mouse model based on conventional mammary carcinoma (4T1 cells) and heightened inflammation mammary carcinoma (4T1/IL-1β) was used. This work centered on the analysis of intact proteins using advanced mass spectrometric instrumentation for top-down mass spectrometry analysis. We successfully interrogated the protein cargo of MDSC-derived exosomes, identifying more than 200 low-mass proteoforms from 21 distinct proteins using extensive protein

fractionation through a 2D fractionation approach, combining GELFrEE and reversed-phase LC. Several proteoforms of the pro-inflammatory S100A8 and S100A9 proteins, which are chemotactic for MDSC, were identified and many fully characterized.

Since post-translational modifications can define protein function and their participation in different cellular processes, we were also interested in obtaining quantitative information that could help understand if there were differences in abundance at the proteoform level related to inflammation. However, approaches for proteoform relative quantitation by top-down analysis are not well-established. Therefore, we evaluated the applicability of spectral counting for top-down proteomics workflows and compare its performance to the most commonly used label-free strategies based on chromatographic peak areas and peak intensities. Through carefully planned spiking experiments and robust statistical analysis, we demonstrated that spectral counting was simple to apply and offered fairly accurate proteoform ratio estimates, with similar or better sensitivity the often used chromatographic approaches. Spectral counting was subsequently applied to the exosome samples, producing the first relative quantitation analysis of MDSC-derived exosomal proteins at the proteoform level. Differences in abundance related to the presence of inflammation were observed for some of the S100A8 and S100A9 potentially active proteoforms. Future biological assays for these proteoforms could help define if the observed changes have any biological relevance.

Under the same premise that inflammation is an important factor on tumor progression and recognizing that exosomes in other biological settings are known to carry and transfer mRNAs and miRNAs;[11] we decided to study the protein, mRNA and miRNA contents of parental MDSCs and MDSC-derived exosomes shed under conventional and heightened inflammation conditions. In this case, shotgun proteomics was performed due to the higher proteome coverage offered when compared to top-down mass spectrometry. In this study we provided evidence that MDSC-derived exosomes carry mRNA and miRNA. Relative quantitation demonstrated quantitative differences between the exosome cargo and the cargo of their parental cells, supporting the hypothesis that selective loading into the exosomes is possible. Additionally, quantitative and functional analyses of the exosome cargo generated under conventional and heightened inflammation conditions are consistent with clinical observations that inflammation is linked to cancer development.

Appendices

Appendix 2.1. List of identified intact proteins in MDSC-derived exosomes. Proteins with similar masses were grouped in colored boxes and counted once.[100]

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
HMG-17	P09602	9417.07	9285.97	-131.10	2.8E-09	Loss of initial Met
MT-4	P47945	6269.31	5621.02	-648.29	4.9E-06	
S100 A6	P14069	10044.29	9955.27	-89.02	3.4E-08	Loss of initial Met + acetylation Ala2
S100 A8	P27005	10288.07	10137.93	-150.13	7.4E-10	
S100 A8	P27005	10288.07	10139.09	-148.98	3.2E-09	Modification localized to first 3 amino acids
S100 A8	P27005	10288.07	10147.11	-140.95	4.9E-09	
S100 A8	P27005	10288.07	10155.96	-132.11	2.6E-13	
S100 A8	P27005	10288.07	10156.94	-131.13	1.6E-17	Loss of initial Met
S100 A8	P27005	10288.07	10159.88	-128.19	5.4E-17	
S100 A8	P27005	10288.07	10170.02	-118.04	4.9E-09	
S100 A8	P27005	10288.07	10172.04	-116.03	7.6E-12	
S100 A8	P27005	10288.07	10172.99	-115.08	6.7E-27	Loss of initial Met + oxidation of Met37
S100 A8	P27005	10288.07	10177.11	-110.95	6.0E-09	
S100 A8	P27005	10288.07	10178.93	-109.14	9.7E-12	
S100 A8	P27005	10288.07	10184.03	-104.04	1.4E-09	
S100 A8	P27005	10288.07	10185.03	-103.04	2.4E-10	
S100 A8	P27005	10288.07	10187.95	-100.11	2.7E-08	
S100 A8	P27005	10288.07	10189.05	-99.02	2.2E-16	
S100 A8	P27005	10288.07	10190.02	-98.05	5.0E-10	
S100 A8	P27005	10288.07	10195.10	-92.97	2.0E-09	
S100 A8	P27005	10288.07	10195.83	-92.23	6.7E-09	
S100 A8	P27005	10288.07	10199.06	-89.01	1.9E-06	Loss of initial Met + acetylation Pro2
S100 A8	P27005	10288.07	10200.03	-88.04	3.6E-16	
S100 A8	P27005	10288.07	10200.92	-87.14	4.4E-13	
S100 A8	P27005	10288.07	10201.95	-86.12	5.0E-10	
S100 A8	P27005	10288.07	10203.05	-85.02	3.1E-09	
S100 A8	P27005	10288.07	10211.01	-77.05	2.4E-10	
S100 A8	P27005	10288.07	10212.08	-75.99	2.0E-16	
S100 A8	P27005	10288.07	10212.98	-75.08	3.8E-05	
S100 A8	P27005	10288.07	10216.08	-71.99	4.0E-05	
S100 A8	P27005	10288.07	10221.97	-66.09	3.7E-06	
S100 A8	P27005	10288.07	10222.98	-65.09	2.7E-05	
S100 A8	P27005	10288.07	10244.60	-43.47	1.3E-05	
S100 A8	P27005	10288.07	10266.90	-21.17	1.6E-12	
S100 A8	P27005	10288.07	10268.98	-19.09	1.2E-25	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
S100 A8	P27005	10288.07	10287.06	-1.01	1.5E-14	
S100 A8	P27005	10288.07	10288.03	-0.04	1.9E-32	
S100 A8	P27005	10288.07	10289.08	1.02	7.8E-20	
S100 A8	P27005	10288.07	10309.11	21.04	1.9E-32	
S100 A8	P27005	10288.07	10309.96	21.90	1.9E-09	
S100 A8	P27005	10288.07	10310.98	22.91	1.2E-27	
S100 A8	P27005	10288.07	10330.07	42.01	4.1E-12	Acetylation of initial Met
S100 A8	P27005	10288.07	10339.96	51.89	1.5E-30	
S100 A8	P27005	10288.07	10340.96	52.90	1.9E-23	
S100 A8	P27005	10288.07	10366.21	78.15	7.5E-07	
S100 A8	P27005	10288.07	10405.15	117.08	2.0E-07	
S100 A8	P27005	10288.07	10570.30	282.23	1.5E-05	
S100 A8	P27005	10288.07	10571.17	283.11	1.2E-14	
S100 A8	P27005	10288.07	10572.23	284.16	7.9E-09	
S100 A8	P27005	10288.07	10587.13	299.06	1.1E-05	
S100 A8	P27005	10288.07	10593.16	305.09	6.1E-07	
S100 A8	P27005	10288.07	10595.14	307.08	2.2E-06	
S100 A8	P27005	10288.07	10985.39	697.32	5.5E-05	
S100 A8	P27005	10288.07	11004.01	715.94	1.1E-08	
S100 A8	P27005	10288.07	12330.95	2042.88	1.7E-07	
S100 A8	P27005	10288.07	12331.96	2043.89	6.3E-05	
S100 A9	P31725	13041.29	12965.14	-76.15	9.3E-19	Loss of initial Met + acetylation Ala2+ methylation H107
H2A.1	P22752	14126.95	14020.80	-106.15	5.7E-05	
H2A.1	P22752	14126.95	14037.98	-88.97	1.7E-12	Loss of initial Met + acetylation Ser2
H2A.1	P22752	14126.95	14063.95	-63.00	1.8E-08	
H2A.1	P22752	14126.95	14080.84	-46.11	3.3E-05	
H2A.1	P22752	14126.95	14818.31	691.36	1.9E-06	
H2A.X	P27661	15133.44	15044.42	-89.02	1.8E-07	Loss of initial Met + acetylation Ser2
H2A.2A	Q6GSS7	14086.89	13285.25	-801.64	6.9E-09	
H2A.2A	Q6GSS7	14086.89	13997.90	-88.99	6.3E-09	Loss of initial Met + acetylation Ser2
H2A.3	Q8BFU2	14112.93	12497.86	-1615.08	6.3E-08	
H2A.3	Q8BFU2	14112.93	14020.80	-92.13	6.0E-05	
H2A.3	Q8BFU2	14112.93	14037.98	-74.95	1.0E-08	Loss of initial Met + 2 methylations (any K or R residues from 2 - 24) + 2 methylations (any K or R residues from 24-50)
H2A.3	Q8BFU2	14112.93	14041.01	-71.92	1.3E-09	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H2A.3	Q8BFU2	14112.93	14061.84	-51.09	5.0E-09	Loss of initial Met + phosphorylation Ser2
H2A.3	Q8BFU2	14112.93	14063.88	-49.06	1.6E-05	
H2A.3	Q8BFU2	14112.93	14080.84	-32.09	1.4E-05	
H2A.3	Q8BFU2	14112.93	14818.31	705.38	1.8E-08	
H2B.1A	P70696	14227.76	12024.56	-2203.20	7.6E-08	
H2B.1B	Q64475	13943.56	12228.71	-1714.85	1.5E-14	
H2B.1B	Q64475	13943.56	13077.81	-865.75	1.3E-05	
H2B.1B	Q64475	13943.56	13765.41	-178.15	2.6E-05	
H2B.1B	Q64475	13943.56	13766.40	-177.16	1.9E-06	
H2B.1B	Q64475	13943.56	13767.32	-176.24	2.1E-09	
H2B.1B	Q64475	13943.56	13778.42	-165.15	1.9E-08	
H2B.1B	Q64475	13943.56	13779.53	-164.04	3.9E-08	
H2B.1B	Q64475	13943.56	13780.41	-163.16	2.8E-06	
H2B.1B	Q64475	13943.56	13780.41	-163.16	1.2E-06	
H2B.1B	Q64475	13943.56	13780.44	-163.12	3.8E-09	
H2B.1B	Q64475	13943.56	13780.47	-163.09	4.5E-06	
H2B.1B	Q64475	13943.56	13781.54	-162.02	3.1E-05	
H2B.1B	Q64475	13943.56	13794.51	-149.05	4.3E-06	Loss of initial Met + dehydration (any S or Y residues from 7-95)
H2B.1B	Q64475	13943.56	13795.48	-148.08	2.0E-06	Loss of initial Met + loss of ammonia (on N-terminus or any N or Q) from residue 2-50
H2B.1B	Q64475	13943.56	13795.49	-148.07	2.8E-11	
H2B.1B	Q64475	13943.56	13796.39	-147.18	2.8E-11	
H2B.1B	Q64475	13943.56	13796.42	-147.14	4.4E-15	
H2B.1B	Q64475	13943.56	13796.42	-147.14	1.5E-12	
H2B.1B	Q64475	13943.56	13796.42	-147.14	6.7E-10	
H2B.1B	Q64475	13943.56	13796.47	-147.09	1.3E-10	
H2B.1B	Q64475	13943.56	13797.41	-146.15	1.5E-12	
H2B.1B	Q64475	13943.56	13797.51	-146.05	1.0E-10	
H2B.1B	Q64475	13943.56	13797.54	-146.02	2.7E-10	
H2B.1B	Q64475	13943.56	13810.47	-133.09	3.6E-17	
H2B.1B	Q64475	13943.56	13811.54	-132.03	3.4E-05	
H2B.1B	Q64475	13943.56	13812.55	-131.02	8.6E-06	Loss of initial Met
H2B.1B	Q64475	13943.56	13814.47	-129.10	7.6E-05	
H2B.1B	Q64475	13943.56	13816.34	-127.22	8.9E-07	
H2B.1B	Q64475	13943.56	13820.46	-123.10	1.0E-05	
H2B.1B	Q64475	13943.56	13837.48	-106.09	3.7E-05	
H2B.1B	Q64475	13943.56	13838.41	-105.16	2.2E-06	
H2B.1B	Q64475	13943.56	13841.42	-102.15	8.3E-05	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H2B.1B	Q64475	13943.56	13894.38	-49.18	9.8E-05	
H2B.1B	Q64475	13943.56	14563.48	619.92	4.3E-07	
H2B.1C	Q6ZWY9	13897.56	13139.36	-758.20	3.8E-09	
H2B.1C	Q6ZWY9	13897.56	13765.53	-132.03	3.2E-08	
H2B.1C	Q6ZWY9	13897.56	13766.54	-131.02	1.9E-07	Loss of initial Met
H2B.1C	Q6ZWY9	13897.56	13767.42	-130.14	7.0E-13	
H2B.1C	Q6ZWY9	13897.56	13768.39	-129.17	7.2E-07	
H2B.1C	Q6ZWY9	13897.56	13779.48	-118.08	1.1E-07	
H2B.1C	Q6ZWY9	13897.56	13780.53	-117.03	1.2E-06	Loss of initial Met + methylation Pro2
H2B.1C	Q6ZWY9	13897.56	13781.48	-116.07	9.1E-11	
H2B.1C	Q6ZWY9	13897.56	13785.41	-112.15	4.9E-07	
H2B.1C	Q6ZWY9	13897.56	13795.42	-102.14	5.8E-12	
H2B.1C	Q6ZWY9	13897.56	13796.39	-101.17	7.2E-12	
H2B.1C	Q6ZWY9	13897.56	13799.34	-98.21	3.0E-05	
H2B.1C	Q6ZWY9	13897.56	13811.56	-86.00	1.5E-05	
H2B.1C	Q6ZWY9	13897.56	13813.51	-84.05	1.8E-05	
H2B.1C	Q6ZWY9	13897.56	13822.44	-75.12	7.2E-13	
H2B.1C	Q6ZWY9	13897.56	13841.42	-56.14	1.3E-10	
H2B.1C	Q6ZWY9	13897.56	14577.92	680.36	2.4E-07	
H2B.1F	P10853	13927.57	13077.81	-849.76	2.6E-05	
H2B.1F	P10853	13927.57	13122.99	-804.58	2.6E-05	
H2B.1F	P10853	13927.57	13778.52	-149.05	9.2E-05	Loss of initial Met + dehydration S65 Modification localized to first 3 aminoacids
H2B.1F	P10853	13927.57	13779.52	-148.05	2.6E-05	
H2B.1F	P10853	13927.57	13780.41	-147.16	6.2E-11	
H2B.1F	P10853	13927.57	13781.54	-146.03	1.2E-06	
H2B.1F	P10853	13927.57	13794.51	-133.06	1.6E-14	
H2B.1F	P10853	13927.57	13795.48	-132.09	3.8E-09	
H2B.1F	P10853	13927.57	13796.53	-131.04	6.6E-07	Loss of initial Met
H2B.1F	P10853	13927.57	13797.54	-130.03	2.6E-05	
H2B.1F	P10853	13927.57	13801.39	-126.18	3.4E-05	
H2B.1F	P10853	13927.57	13811.49	-116.08	8.6E-06	
H2B.1F	P10853	13927.57	13816.34	-111.22	7.6E-05	
H2B.1F	P10853	13927.57	13838.41	-89.16	1.0E-05	Loss of initial Met + acetylation (any K residues from 56-62)
H2B.1F	P10853	13927.57	13841.42	-86.15	1.6E-07	
H2B.1F	P10853	13927.57	13894.38	-33.18	6.1E-08	
H2B.1F	P10853	13927.57	14468.56	540.99	2.5E-05	
H2B.1F	P10853	13927.57	14504.23	576.66	7.8E-05	
H2B.1F	P10853	13927.57	14543.35	615.78	6.6E-07	
H2B.1F	P10853	13927.57	14563.48	635.92	9.8E-05	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H2B.1H	Q64478	13911.57	13139.36	-772.22	8.4E-09	
H2B.1H	Q64478	13911.57	13765.53	-146.05	5.6E-11	
H2B.1H	Q64478	13911.57	13766.54	-145.03	1.0E-14	
H2B.1H	Q64478	13911.57	13768.39	-143.18	5.7E-08	
H2B.1H	Q64478	13911.57	13779.48	-132.09	7.2E-07	
H2B.1H	Q64478	13911.57	13780.47	-131.10	3.4E-13	Loss of initial Met
H2B.1H	Q64478	13911.57	13781.48	-130.09	8.2E-07	
H2B.1H	Q64478	13911.57	13796.56	-115.02	1.4E-07	Loss of initial Met + oxidation of Met60 or Met63
H2B.1H	Q64478	13911.57	13798.38	-113.19	3.4E-05	
H2B.1H	Q64478	13911.57	13799.34	-112.23	4.8E-08	
H2B.1H	Q64478	13911.57	13816.34	-95.23	1.2E-09	
H2B.1M	P10854	13927.57	13139.36	-788.21	3.2E-08	
H2B.1M	P10854	13927.57	13748.56	-179.01	6.1E-09	
H2B.1M	P10854	13927.57	13765.42	-162.14	1.0E-11	
H2B.1M	P10854	13927.57	13766.54	-161.03	7.2E-07	
H2B.1M	P10854	13927.57	13768.39	-159.18	3.4E-13	
H2B.1M	P10854	13927.57	13779.53	-148.04	1.2E-07	Modification localized to first 8 aminoacids
H2B.1M	P10854	13927.57	13779.60	-147.97	2.5E-05	
H2B.1M	P10854	13927.57	13781.48	**-146.09**	1.8E-10	
H2B.1M	P10854	13927.57	13785.47	-142.10	1.1E-05	
H2B.1M	P10854	13927.57	13796.56	-131.01	4.8E-08	Loss of initial Met
H2B.1M	P10854	13927.57	13799.34	-128.22	5.8E-12	
H2B.1M	P10854	13927.57	13812.47	-115.10	7.2E-12	Loss of initial Met + oxidation of Met60
H2B.1M	P10854	13927.57	14577.92	650.35	1.0E-08	
H2B.2B	Q64525	13911.57	13139.36	-772.22	8.4E-09	
H2B.2B	Q64525	13911.57	13765.53	-146.05	5.6E-11	
H2B.2B	Q64525	13911.57	13766.54	-145.03	7.2E-07	
H2B.2B	Q64525	13911.57	13768.39	-143.18	1.1E-07	
H2B.2B	Q64525	13911.57	13779.49	-132.09	1.9E-06	
H2B.2B	Q64525	13911.57	13780.56	-131.01	8.2E-07	Loss of initial Met
H2B.2B	Q64525	13911.57	13781.48	-130.09	4.8E-08	
H2B.2B	Q64525	13911.57	13796.56	-115.02	1.4E-07	Loss of initial Met + oxidation of Met60 or Met63
H2B.2B	Q64525	13911.57	13798.38	-113.19	5.8E-08	
H2B.2B	Q64525	13911.57	13799.34	-112.23	3.4E-05	
H2B.2B	Q64525	13911.57	13815.45	-96.12	4.8E-08	
H2B.2B	Q64525	13911.57	13816.34	-95.23	1.2E-09	
H3.1	Q6LBF0	15394.48	13071.10	-2323.38	6.4E-05	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H3.1	Q6LBF0	15394.48	15305.56	-88.92	1.5E-07	Loss of initial Met + acetylation (any K residues from 16-48)
H3.1	Q6LBF0	15394.48	15318.56	-75.93	5.2E-07	
H3.1	Q6LBF0	15394.48	15393.48	-1.00	7.2E-05	
H3.2	P84228	15378.51	13011.17	-2367.33	1.0E-09	Loss of 22 residues from N-terminus
H3.2	P84228	15378.51	13024.22	-2354.28	9.7E-11	
H3.2	P84228	15378.51	13025.13	-2353.38	9.9E-08	
H3.2	P84228	15378.51	13026.15	-2352.36	8.1E-08	Loss of 22 residues + methylation (any K or R residues from 23-37)
H3.2	P84228	15378.51	13039.22	-2339.28	2.7E-05	Loss of 22 residues + methylation (any K or R residues from 23-37) + methylation (any K or R residues from 38-90)
H3.2	P84228	15378.51	13043.19	-2335.31	5.3E-06	
H3.2	P84228	15378.51	13054.25	-2324.26	6.7E-07	
H3.2	P84228	15378.51	13056.11	-2322.40	7.8E-08	
H3.2	P84228	15378.51	13098.18	-2280.33	7.4E-06	Loss of 21 residues + methylation (any K or R residues from 22-90)
H3.2	P84228	15378.51	13709.39	-1669.11	5.8E-05	
H3.2	P84228	15378.51	13724.14	-1654.37	1.1E-05	
H3.2	P84228	15378.51	14027.99	-1350.52	1.5E-07	
H3.2	P84228	15378.51	15291.43	-87.08	3.0E-06	
H3.2	P84228	15378.51	15305.49	-73.02	4.4E-10	Loss of M N-terminus + acetylation (any K residues from 16-48) + oxidation of Met91
H3.2	P84228	15378.51	15319.42	-59.09	4.4E-07	
H3.2	P84228	15378.51	15332.52	-45.99	1.8E-07	
H3.3	P84244	15318.50	13011.17	-2307.33	1.0E-09	
H3.3	P84244	15318.50	13024.22	-2294.28	9.7E-11	Loss of 21 residues from N-terminus
H3.3	P84244	15318.50	13026.15	-2292.35	2.0E-09	
H3.3	P84244	15318.50	13035.10	-2283.40	9.9E-08	
H3.3	P84244	15318.50	13038.18	-2280.32	8.1E-08	
H3.3	P84244	15318.50	13039.22	-2279.28	1.3E-08	
H3.3	P84244	15318.50	13040.10	-2278.40	5.0E-10	
H3.3	P84244	15318.50	13041.09	-2277.41	1.7E-19	
H3.3	P84244	15318.50	13043.19	-2275.31	5.0E-21	
H3.3	P84244	15318.50	13054.25	-2264.25	2.4E-13	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H3.3	P84244	15318.50	13056.11	-2262.39	2.7E-05	
H3.3	P84244	15318.50	13065.16	-2253.34	3.0E-05	Loss of 21 residues + acetylation (any K residues from 59-89)
H3.3	P84244	15318.50	13069.20	-2249.31	5.3E-06	
H3.3	P84244	15318.50	13083.08	-2235.42	7.8E-08	
H3.3	P84244	15318.50	13098.18	-2220.32	7.4E-06	
H3.3	P84244	15318.50	13709.39	-1609.11	5.8E-05	
H3.3	P84244	15318.50	13724.14	-1594.37	1.1E-05	
H3.3	P84244	15318.50	15275.44	-43.06	1.5E-07	
H3.3	P84244	15318.50	15291.45	-27.05	5.0E-07	
H3.3	P84244	15318.50	15305.56	-12.94	1.7E-06	
H3.3	P84244	15318.50	15318.56	0.05	8.7E-05	
H3.3	P84244	15318.50	15319.47	0.96	6.7E-05	
H3.3	P84244	15318.50	15332.56	14.06	4.6E-05	Methylation (any K or R residues from 51-73)
H3.3	P84244	15318.50	15334.45	15.94	8.9E-06	
H3.3	P84244	15318.50	15347.48	28.98	7.0E-06	Oxidation of initial Met + methylation (any K or R residues from 53-84)
H3.C	P02301	15305.50	13011.17	-2294.32	5.3E-13	Loss of 22 residues + methylation (any K or R residues from 51-57) + methylation (any K or R residues from 58-91)
H3.C	P02301	15305.50	13024.22	-2281.27	2.5E-12	
H3.C	P02301	15305.50	13026.15	-2279.35	1.7E-12	
H3.C	P02301	15305.50	13035.10	-2270.40	1.9E-12	
H3.C	P02301	15305.50	13038.18	-2267.31	1.0E-09	
H3.C	P02301	15305.50	13039.22	-2266.27	9.7E-11	
H3.C	P02301	15305.50	13040.10	-2265.39	2.0E-09	
H3.C	P02301	15305.50	13041.09	-2264.41	1.6E-08	
H3.C	P02301	15305.50	13043.19	-2262.30	9.9E-08	
H3.C	P02301	15305.50	13054.25	-2251.25	8.1E-08	Loss of 21 residues from N-terminus
H3.C	P02301	15305.50	13056.11	-2249.39	1.3E-08	
H3.C	P02301	15305.50	13065.16	-2240.34	5.0E-10	
H3.C	P02301	15305.50	13070.17	-2235.33	1.7E-19	
H3.C	P02301	15305.50	13083.08	-2222.41	2.7E-05	
H3.C	P02301	15305.50	13098.18	-2207.32	3.0E-05	
H3.C	P02301	15305.50	13709.39	-1596.10	5.3E-06	
H3.C	P02301	15305.50	13724.14	-1581.36	7.8E-08	
H3.C	P02301	15305.50	13791.95	-1513.55	7.4E-06	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H3.C	P02301	15305.50	13807.65	-1497.85	5.8E-05	
H3.C	P02301	15305.50	13823.67	-1481.83	1.1E-05	
H3.C	P02301	15305.50	13837.46	-1468.03	1.8E-06	
H3.C	P02301	15305.50	15290.51	-14.98	1.5E-07	
H3.C	P02301	15305.50	15304.51	-0.99	1.8E-09	
H3.C	P02301	15305.50	15305.56	0.07	6.2E-08	
H3.C	P02301	15305.50	15306.43	0.93	1.8E-08	
H3.C	P02301	15305.50	15316.43	10.93	7.7E-07	
H3.C	P02301	15305.50	15331.37	25.87	4.4E-06	
H3.C	P02301	15305.50	15347.48	41.99	6.8E-06	Acetylation any S or K before residue 75
H3.C	P02301	15305.50	16018.36	712.86	7.0E-06	
H4	P62806	11360.38	9043.97	-2316.41	6.3E-07	Loss of 23 residues + oxidation of Met85
H4	P62806	11360.38	9059.95	-2300.43	1.3E-13	Loss of 23 residues from N-terminus
H4	P62806	11360.38	9079.86	-2280.52	3.8E-07	
H4	P62806	11360.38	9584.41	-1775.97	3.2E-07	Loss of 18 residues + 2 methylathions (any K or R residues from 20-53) + oxidation of Met85
H4	P62806	11360.38	9705.37	-1655.01	2.7E-08	Loss of 18 residues + 2 methylations (any K or R residues from 19-23)
H4	P62806	11360.38	11285.26	-75.12	2.7E-08	Loss of initial Met + dimethylation R4 + 2 methylations (any K or R residues from 21-40)
H4	P62806	11360.38	11287.29	-73.09	1.6E-06	
H4	P62806	11360.38	11299.32	-61.07	1.4E-11	
H4	P62806	11360.38	11300.32	-60.06	8.1E-15	
H4	P62806	11360.38	11301.32	-59.06	7.0E-08	
H4	P62806	11360.38	11302.29	-58.09	5.3E-06	
H4	P62806	11360.38	11303.18	-57.21	3.6E-12	
H4	P62806	11360.38	11304.24	-56.14	7.3E-07	
H4	P62806	11360.38	11325.20	-35.19	1.0E-06	
H4	P62806	11360.38	11329.34	-31.05	5.3E-10	
H4	P62806	11360.38	11343.21	-17.17	1.2E-05	Loss of ammonia (on N-terminus, or any N or Q residues from 1-40)
H4	P62806	11360.38	11344.43	-15.95	1.8E-05	
H4	P62806	11360.38	11353.38	-7.00	4.1E-08	
H4	P62806	11360.38	11359.35	-1.03	1.8E-19	

Protein variant	Ascension #	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Putative post-translational modification
H4	**P62806**	**11360.38**	**11365.22**	**4.84**	**7.4E-38**	
H4	P62806	11360.38	11381.28	20.90	2.1E-12	
H4	P62806	11360.38	11384.44	23/24.059	3.3E-09	
H4	P62806	11360.38	12314.91	954.53	6.7E-05	
H4	P62806	11360.38	13765.57	2405.19	1.2E-10	
H4	P62806	11360.38	13782.57	2422.19	3.2E-10	
H4	P62806	11360.38	13796.41	2436.02	2.8E-05	

Appendix 3.1. Linear regression plots of observed and expected protein ratios without log$_2$ transformation obtained for (a) normalized area, (b) normalized intensity, (c) spectral counts.[118]

(a) Normalized areas
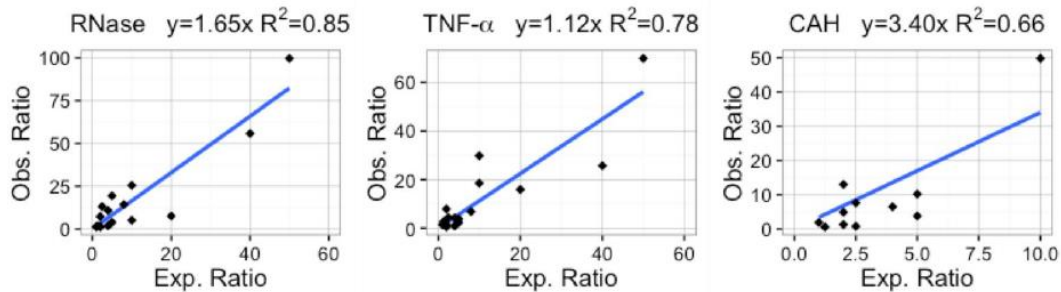
(b) Normalized intensities
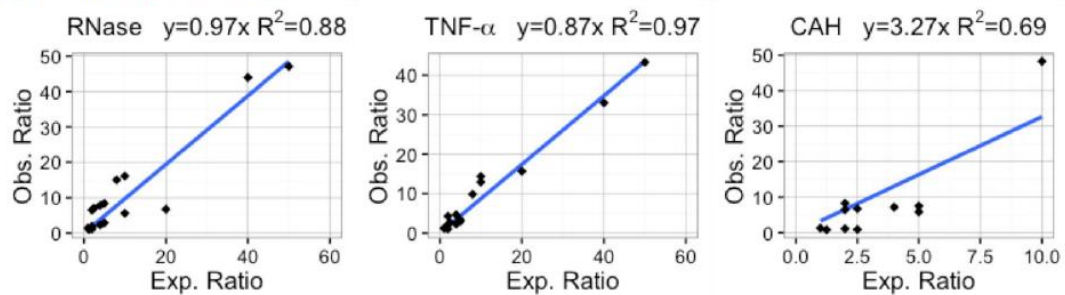
(c) Spectral counts
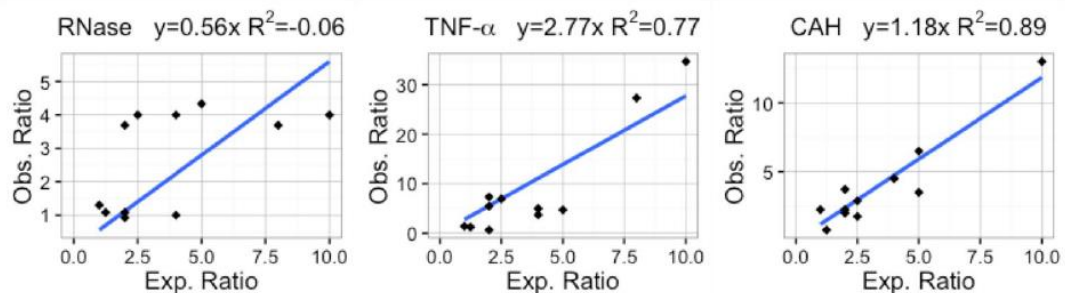
Appendix 3.2. Complete list of proteoforms identified and comparison of differential abundances calculated by spectral counting, normalized intensity and area.[118]

Note: All identifications required < 1% FDR against a decoy database. PrSM counts shown were corrected by global median normalization
* Residue coverage obtained for those proteoforms that were fully characterized. Coverages were determined by combining the fragment ions obtained from 2 spectra one obtained by CID and one by EThcD fragmentation.
\# mass differences (ppm) were provided for those proteoforms that were fully characterized. Unknown mass additions are stated in the Putative PTM section.
Grey highlight means that the proteoform had < 3 counts or was not present in 50% of the biological replicates
MW – Mann-Whitney test

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P31786	9905.15	Loss of initial M	-0.71	49	2.3	2.1×10^{-1}	1.8×10^{-1}	7.9×10^{-2}				
Q9CY02	11693.83	Loss of initial M	0.43	15	-0.6	7.9×10^{-1}	7.1×10^{-1}	5.2×10^{-1}				
Q9CPW5	16189.30	Loss of initial M + acetylation S2	0.25	13	1.0	9.7×10^{-1}	5.6×10^{-1}	3.6×10^{-1}				
Q9CPW5	16308.28	Loss of initial M + acetylation S2 + cysteinylation C45	-1.37	54	-1.6	1.4×10^{-8}	1.2×10^{-1}	6.2×10^{-8}	-2.6	1.6×10^{-1}	-2.0	4.0×10^{-2}
Q9CPW5	17016.69	Loss of initial M + 827.39 Da	-	-	1.0	9.7×10^{-1}	5.6×10^{-1}	3.6×10^{-1}				
Q9CPW5	17049.05	Loss of initial M + 859.76 Da	-	-	-1.4	2.6×10^{-1}	4.1×10^{-1}	1.4×10^{-1}				
Q9CPW5	17084.31	Loss of initial M + 895.01 Da	-	-	-1.6	1.8×10^{-1}	3.0×10^{-1}	9.4×10^{-2}				

130

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
Q9CPW5	17123.18	Loss of initial M + 933.89 Da	-	-	-2.3	1.9×10^{-1}	4.7×10^{-1}	7.9×10^{-2}				
Q9CPW5	17789.69	Loss of initial M + 1600.39 Da	-	-	-2.0	3.4×10^{-1}	4.7×10^{-1}	1.3×10^{-1}				
Q9CPW5	17861.28	Loss of initial M + 1671.98 Da	-	-	-1.6	5.6×10^{-1}	6.1×10^{-1}	2.1×10^{-1}				
P62204	16764.84	Loss of initial M + acetylation A2 + 27.03 Da	-	-	1.0	9.7×10^{-1}	1.9×10^{-4}	3.6×10^{-1}				
P62204	16779.85	Loss of initial M + acetylation A2 + trimethylation K 116	-0.34	46	-1.6	1.1×10^{-4}	2.3×10^{-1}	1.2×10^{-4}	-2.5	2.0×10^{-1}	-1.2	7.3×10^{-1}
P62204	16795.82	Loss of initial M + acetylation A2 + trimethylation K 116 + oxidation between residues 64 to 136	-1.82	27	-1.6	5.4×10^{-4}	1.2×10^{-1}	5.6×10^{-4}	-2.3	7.1×10^{-1}	-1.5	7.3×10^{-1}

131

					Spectral counting				Peak Areas		Peak Intensities	
Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P62204	16811.91	Loss of initial M + acetylation A2 + trimethylation K116 + 32.05 Da	-	-	-2.6	1.2×10^{-1}	3.0×10^{-1}	5.2×10^{-2}				
Q64433	10866.88	Loss of initial M + acetylation A2	3.79	72	-3.3	1.3×10^{-2}	2.8×10^{-1}	8.9×10^{-3}				
P35175	10982.50	-16.52 Da	-	-	-1.4	1.3×10^{-1}	2.8×10^{-1}	7.9×10^{-2}				
P35173	11661.93	Loss of initial M + acetylation S2 + cysteinylation C69	0.34	29	0.0	1.0	9.6×10^{-1}	9.9×10^{-1}				
P29391	20698.51	Loss of initial M + acetylation T2 -1.95 Da	-	-	2.8	6.3×10^{-2}	6.3×10^{-2}	3.2×10^{-2}				
P22752	12354.95	Loss of initial M + acetylation S2 + loss of C-term after V114	1.71	63	1.2	7.1×10^{-3}	1.4×10^{-1}	5.6×10^{-3}	1.3	1.2×10^{-1}	1.4	7.5×10^{-2}
P22752	14038.9	Loss of initial M + acetylation S2 + citrullination R4	-0.26	13	-5.0	1.4×10^{-8}	5.0×10^{-1}	5.9×10^{-7}				

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
Q6GSS7	12315.85	Loss of initial M + acetylation S2 + citrullination R4 + loss of C-term after V114	0.31	59	3.2	2.1×10^{-2}	1.8×10^{-2}	1.3×10^{-2}				
P10853	13796.53	Loss of initial M	0.11	33	2.0	3.6×10^{-1}	2.6×10^{-1}	1.3×10^{-1}				
P62806	9044.01	Loss until L23	-2.26	92	3.3	1.2×10^{-2}	6.3×10^{-2}	9.0×10^{-3}				
P62806	9060.04	Loss until L23 + oxidation M85	1.72	76	2.8	6.3×10^{-2}	6.3×10^{-2}	3.2×10^{-2}				
P62806	9706.44	-	-	-	2.0	3.6×10^{-1}	6.3×10^{-2}	1.3×10^{-1}				
P62806	11286.33	Loss of initial M + acetylation S2 + citrullination R4 + methylation K21 or K32	-2.03	57	1.0	9.7×10^{-1}	3.4×10^{-1}	3.6×10^{-1}				
P62806	11301.34	Loss of initial M + acetylation S2 + citrullination R4 and R18 + dimethylation K21	-0.97	65	0.7	3.8×10^{-1}	5.9×10^{-1}	2.6×10^{-1}				

					Spectral counting				Peak Areas		Peak Intensities	
Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P62806	11317.32	Loss of initial M + acetylation S2 + citrullination R4 and R18 + dimethylation K21 + oxidation M85	-1.92	56	-1.5	1.1×10^{-1}	6.3×10^{-1}	5.7×10^{-2}				
P62806	11342.35	Loss of initial M + acetylation S2 + citrullination R4 + acetylation K17 + dimethylation K21	-2.69	81	4.2	4.8×10^{-5}	6.3×10^{-2}	1.2×10^{-4}				
P62806	11360.35	Unmodified	-2.90	43	-1.6	1.2×10^{-1}	4.4×10^{-1}	6.2×10^{-2}				
P01942	14907.63	Loss of initial M + cysteinylation C105 + loss of C-term R	0.80	45	-3.2	2.2×10^{-2}	2.8×10^{-1}	1.3×10^{-2}				
P01942	14921.63	Loss of initial M + AAS S69T + cysteinylation C105 + loss of C-term R	-0.56	40	-2.6	1.2×10^{-1}	3.0×10^{-1}	5.2×10^{-2}				
P01942	14944.73	Loss of initial M	0.40	26	1.5	1.2×10^{-2}	1.4×10^{-1}	9.0×10^{-3}				

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P01942	14958.73	Loss of initial M + AAS S69T	-0.21	61	2.2	4.8×10^{-5}	6.3×10^{-1}	6.7×10^{-5}				
P01942	15063.71	Loss of initial M + cysteinylation C105	-0.62	47	-0.6	1.9×10^{-1}	6.3×10^{-1}	1.3×10^{-1}	-1.6	3.9×10^{-1}	-1.0	4.9×10^{-1}
P01942	15077.81	Loss of initial M + AAS S69T + cysteinylation C105	-1.26	64	-0.6	2.8×10^{-1}	6.3×10^{-1}	2.0×10^{-1}	-1.7	3.6×10^{-1}	-0.9	4.9×10^{-1}
P02088	15697.09	Loss of initial M + disulfide bond C14-C94	-0.33	30	2.3	2.1×10^{-1}	6.3×10^{-2}	7.9×10^{-2}				
P09602	9286.02	Loss of initial M	-0.91	59	0.1	9.3×10^{-1}	6.8×10^{-1}	6.8×10^{-1}	0.0	2.1×10^{-2}	0.4	4.8×10^{-2}
P09602	9157.94	Loss of initial M + loss of C-term K	0.68	34	2.5	3.9×10^{-2}	1.9×10^{-1}	0.017785				
P99027	11739.78	Oxidation M109 + phosphorylation S86	-1.05	13	-3.2	2.2×10^{-2}	2.8×10^{-1}	1.3×10^{-2}				
P99027	11659.84	Oxidation M109	1.13	20	-2.5	3.0×10^{-2}	2.8×10^{-1}	1.7×10^{-2}				
P99027	11643.82	Unmodified	-0.40	10	-1.6	5.6×10^{-1}	4.7×10^{-1}	2.1×10^{-1}				

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P14069	10074.27	Loss of initial M + acetylation A2 + cysteinylation C3	0.36	80	-0.7	1.3×10^{-1}	2.8×10^{-1}	9.5×10^{-2}	0.0	3.1×10^{-2}	0.3	3.2×10^{-2}
P31725	12963.26	Loss of initial M + acetylation A2 + methylation H107 + disulfide bond C91-C111	0.41	60	2.4	3.0×10^{-12}	7.5×10^{-3}	6.5×10^{-11}	4.6	5.8×10^{-4}	3.8	3.0×10^{-2}
P31725	12979.33	Loss of initial M + acetylation A2 + methylation H107 + disulfide bond C91-C111 + oxidation M81	6.21	65	2.0	3.6×10^{-1}	6.3×10^{-2}	1.3×10^{-1}				
P31725	13146.28	Loss of initial M + acetylation A2 + methylation H107 + disulfide bond C91-C111 + 183.03 Da	-	-	2.8	6.3×10^{-1}	1.1×10^{-1}	3.2×10^{-2}				
P31725	13203.28	Loss of initial M + acetylation A2 + 252.02 Da	-	-	2.3	2.1×10^{-1}	6.4×10^{-2}	7.9×10^{-2}				

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P31725	13683.93	Loss of initial M + acetylation A2 + 732.68 Da	-	-	1.0	9.7×10^{-1}	1.9×10^{-4}	3.6×10^{-1}				
P08228	15841.80	Loss of initial M + acetylation A2 + disulfide bond C57-C146	2.73	13	0.6	9.7×10^{-1}	6.0×10^{-1}	6.1×10^{-1}				
P27005	9675.83	Loss of initial M + loss of C-term residues (S86 - E89)	2.97	70	0.4	9.7×10^{-1}	8.4×10^{-1}	6.9×10^{-1}				
P27005	10101.98	Loss of N-terminus (M1-P2) + acetylation S3	-0.60	85	0.3	8.7×10^{-1}	3.4×10^{-1}	5.9×10^{-1}	0.5	1.5×10^{-2}	1.0	2.6×10^{-2}
P27005	10140.03	-148.04 Da	-	-	2.8	4.8×10^{-5}	1.9×10^{-4}	7.4×10^{-5}	3.0	5.1×10^{-3}	2.0	2.5×10^{-2}
P27005	10157.05	Loss of initial M	2.35	97	0.3	1.3×10^{-10}	9.2×10^{-2}	4.4×10^{-6}	0.6	2.2×10^{-3}	0.0	2.5×10^{-2}
P27005	10172.99	Loss of initial M + cysteine sulfenic acid C42	-2.84	90	-1.2	2.8×10^{-5}	3.5×10^{-2}	4.1×10^{-5}	-0.2	1.5×10^{-2}	1.1	2.6×10^{-2}
P27005	10179.05	Loss of initial M + 22.03 Da	-	-	0.1	9.7×10^{-1}	8.8×10^{-1}	7.7×10^{-1}	0.2	8.0×10^{-3}	0.6	2.9×10^{-2}

					Spectral counting				Peak Areas		Peak Intensities	
Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P27005	10189.07	Loss of initial M + cysteine sulfinic acid C42	4.96	89	-2.6	9.9×10^{-10}	3.5×10^{-2}	7.2×10^{-9}	-1.4	4.7×10^{-1}	-0.4	6.8×10^{-1}
P27005	10194.98	Loss of initial M + 37.95 Da	-	-	3.2	6.9×10^{-4}	1.9×10^{-4}	7.2×10^{-4}				
P27005	10199.01	Loss of initial M + acetylation P2 or S3	-2.10	45	2.0	9.1×10^{-3}	1.1×10^{-3}	6.6×10^{-3}				
P27005	10205.02	Loss of initial M + cysteine sulfonic acid C42	1.44	74	-3.5	3.5×10^{-7}	1.3×10^{-2}	2.6×10^{-6}	-2.1	7.2×10^{-1}	-0.6	7.5×10^{-1}
P27005	10212.54	Loss of initial M + 55.52 Da	-	-	-0.8	4.9×10^{-1}	5.1×10^{-1}	3.2×10^{-1}	-0.7	9.6×10^{-1}	1.8	6.6×10^{-1}
P27005	10221.02	Loss of initial M + oxidation M37 + cysteine sulfonic acid C42	1.41	80	-2.4	8.0×10^{-12}	1.3×10^{-2}	7.6×10^{-11}	-2.7	7.1×10^{-1}	-2.2	7.3×10^{-1}
P27005	10228.59	Loss of initial M + 71.56 Da	-	-	-1.6	5.6×10^{-1}	4.1×10^{-1}	2.1×10^{-1}				
P27005	10235.02	Loss of initial M + 78.00 Da	-	-	-2.1	3.2×10^{-7}	2.8×10^{-1}	9.9×10^{-7}				
P27005	10276.01	Loss of initial M + cysteinylation C42	-1.77	78	-0.4	5.1×10^{-1}	4.1×10^{-1}	3.6×10^{-1}	0.0	5.2×10^{-2}	1.4	4.0×10^{-2}

Accession n°	Proteoform mass (Da)	Putative PTM	Mass diff. (ppm) #	% res.*	Spectral counting				Peak Areas		Peak Intensities	
					log2 (Ratio) Counts	Fisher Exact's FDR	Perm. FDR	χ^2 test FDR	log2 (I/C) Area	MW adj. p-value	log2 (I/C) Int.	MW adj. p-value
P27005	10288.07	Unmodified	-0.62	91	-0.2	4.1×10^{-1}	5.1×10^{-1}	3.6×10^{-1}	1.4	5.8×10^{-4}	0.2	2.5×10^{-2}
P27005	10304.04	Oxidation M1	-1.70	81	-1.3	2.3×10^{-4}	6.3×10^{-2}	2.5×10^{-4}	0.3	9.6×10^{-3}	-0.4	2.4×10^{-1}
P27005	10340.04	Loss of initial M + aminoethyl benzene sulfonylation C42	-1.66	86	0.0	9.7×10^{-1}	9.4×10^{-1}	9.8×10^{-1}	0.3	5.8×10^{-4}	0.4	2.5×10^{-2}
P27005	10366.05	+77.98Da	-	-	-1.9	3.0×10^{-2}	4.4×10^{-1}	1.8×10^{-2}				
P27005	10377.12	Loss of initial M + 220.09 Da	-	-	3.0	3.7×10^{-2}	6.3×10^{-2}	2.0×10^{-2}				
P27005	10423.15	Loss of initial M + 266.12 Da	-	-	4.7	7.5×10^{-7}	2.6×10^{-1}	5.2×10^{-6}				
P27005	10455.19	Loss of initial M + 298.17 Da	-	-	2.3	2.1×10^{-1}	1.8×10^{-1}	7.9×10^{-2}				
P27005	10469.21	Loss of initial M + 312.19 Da	-	-	3.6	1.5×10^{-5}	1.1×10^{-1}	3.4×10^{-5}				
P27005	10483.21	Loss of initial M + 326.19 Da	-	-	2.7	9.4×10^{-4}	5.3×10^{-2}	8.7×10^{-4}				
P27005	10554.22	+266.15 Da	-	-	1.0	9.7×10^{-1}	3.4×10^{-1}	3.6×10^{-1}				
P27005	10689.35	Loss of initial M + 532.32 Da	-	-	1.0	9.7×10^{-1}	1.9×10^{-1}	3.6×10^{-1}				

Appendix 4.1. Results of data processing for the analysis of the mRNA, miRNA and protein cargo are attached in separate Appendix 4.1 pdf file.

Appendix 4.2. Sequencing quality assessment by FastQC showed that in the first 67 base pairs assignments a Phred quality score > 30 was obtained, which corresponds to an accuracy > 99.9% in assigning the correct base pair.
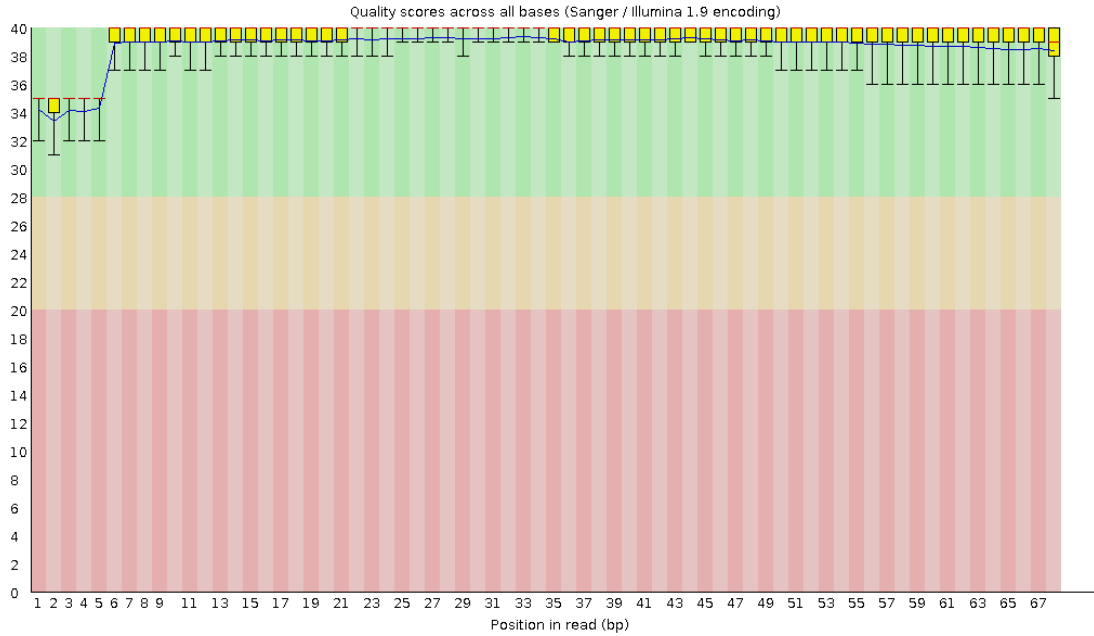
Appendix 4.3. Heatmap of hierarchical clustering by Euclidean distance for (a) miRNA and (b) mRNA.
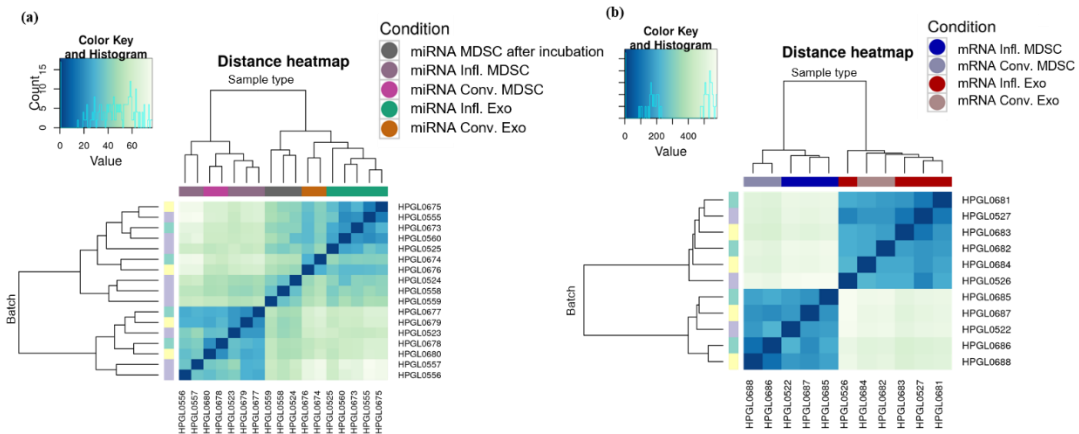
Appendix 4.4. *mRNA* transcripts found in greater abundance in exosomes compared to MDSC that translate into surface proteins, as annotated by GO cellular compartment categories: "cell surface", "cell periphery" and "plasma membrane". Statistically significant fold-changes (FC ≥ 2, with adjusted p-value ≤ 0.05) for the comparisons of exosomes vs. MDSC under conventional and inflammatory conditions are shown.

Ensembl Transcript ID (ENSMUST#)	Uniprot Accession (Swiss-Prot)	Uniprot Accession (TrEMBL)	Description	FC Infl.	FC Conv.
00000000161	Q3UV74		Integrin beta 2-like	6.6	-
00000000808	Q60837		Interleukin-12 receptor subunit beta-1 (CD212)	7.8	12.0
00000000834	P41047	Q544E9	CD95L (FasL)	5.2	5.9
00000001040	Q9ERM2	Q14AB0	Intercellular adhesion molecule 4 (CD242)	48.9	84.4
00000001548	Q62470		Integrin alpha-3 (CD49c)	124.3	-
00000002379	Q9Z1P5		CD320 antigen	170.5	85.4
00000006101	Q60677		Integrin alpha E, epithelial-associated	8.7	-
00000003061	Q9R069		Basal cell adhesion molecule (CD239)	150.9	3.0
00000005678	P20693		Low affinity immunoglobulin epsilon Fc receptor (CD23)	24.1	10.3
00000006749	P04919	Q53ZN9	Band 3 anion transport protein (CD233)	15.7	-
00000009058	P06795	B2RUR3	Multidrug resistance protein 1B (CD243)	5.3	-
00000009705	Q63961		Endoglin (CD105)	42.2	164.8
00000012847	Q91ZX1		CD209a antigen	13.3	-
00000015460	Q9QUM4	Q544K1	Signaling lymphocytic activation molecule (CD150)	5.0	-
00000015540	O88324		CD83 antigen	6.5	-
00000015998	Q9QWK4		CD5 antigen-like	68.0	-
00000016640	Q9EP73	Q3U472	CD274 antigen	6.7	-
00000017453	Q1ERP8		CD300 antigen like family member G	9.8	10.2
00000017799	P27512		CD40 antigen	23.8	16.9
00000019248		Q3UP36	CD22 antigen	46.5	33.6
00000019633	O55237	Q05A52	CD70 antigen	101.0	44.9
00000020579	Q00547		Hyaluronan mediated motility receptor (CD168)	10.4	10.5
00000022592	O35235		Tumor necrosis factor ligand superfamily member 11 (CD254)	9.8	15.7
00000022663	Q9QZM4		Tumor necrosis factor receptor superfamily member 10B (CD262)	30.0	60.1
00000022871	P43407		Syndecan-2 (CD362)	29.0	-
00000023335	Q9JLB9		Nectin-3 (CD113)	736.7	233.5
00000023341	E9QMY1		CD200 antigen	-	56.7
00000023464	Q9R0R1	Q544J8	Melanotransferrin (CD228)	259.9	312.6

Ensembl Transcript ID (ENSMUST#)	Uniprot Accession (Swiss-Prot)	Uniprot Accession (TrEMBL)	Description	FC Infl.	FC Conv.
00000024044	P06332	Q3TSV7	CD4 antigen	5.9	-
00000024721	Q9QUT0		Ammonium transporter Rh type A (CD241)	22.8	48.9
00000025166	P15116		Cadherin-2 (CD antigen CD325)	513.8	189.9
00000026360		Q0VBD0	Integrin beta 8	206.5	32.6
00000026917	P97333		Neuropilin-1 (CD304)	3.9	-
00000027165	P31041		CD28 antigen	6.2	-
00000027241	P13504	Q32MH0	Interleukin-1 receptor type 1 (CD121a)	60.8	13.3
00000027507	Q02242	Q544F3	Programmed cell death protein 1 (CD279)	69.5	-
00000027559	O35305		Tumor necrosis factor receptor superfamily member 11A (CD265)	66.6	33.2
00000028024	P43488	B6DXE3	Tumor necrosis factor ligand superfamily member 4 (CD252)	95.8	6.9
00000028045	Q61830	Q2HZ94	Macrophage mannose receptor 1 (CD206)	3.6	-
00000028106	A2ARA8		Integrin alpha 8	133.3	76.9
00000028111	P01590	Q544I2	Interleukin-2 receptor subunit alpha (CD25)	53.3	44.2
00000028328	O55026	Q921R1	Ectonucleoside triphosphate diphosphohydrolase 2 (CD39 antigen-like 1)	50.1	88.5
00000028735	Q9QXX0	Q3UVN4	Protein jagged-1 (CD339)	4.5	5.6
00000029744		E9Q6R1	Integrin, alpha 10	17.9	-
00000031899	Q9EQF2		Kell blood group glycoprotein homolog (CD238)	89.1	159.3
00000032217	Q61790		Lymphocyte activation gene 3 protein (CD223)	44.3	58.5
00000032234	Q2VLH6		CD163 antigen	47.8	-
00000032341	Q9CRA0		Ecto-ADP-ribosyltransferase 4 (CD297)	16.3	12.3
00000032472	Q99JB4		Killer cell lectin-like receptor subfamily B member 1B allele B (CD161b)	60.8	7.9
00000032486	P41272	Q3U4X0	CD27 antigen	11.6	8.7
00000032968	P25918		CD19 antigen	3.2	-
00000033300	Q60935	Q545U9	GPI-linked NAD(P)(+)-arginine ADP-ribosyltransferase 1 (CD296)	133.6	131.1
00000033051		E9PXZ7	Integrin, alpha D	14.9	75.1
00000033888	Q91ZW7	Q3KNN8	CD209e antigen	63.4	-
00000034056	Q99MB1		Toll-like receptor 3 (CD283)	85.7	-
00000034339	P55284		Cadherin-5 (CD144)	71.6	22.7
00000034510	Q9JKF6		Nectin-1 (CD111)	12.6	11.9
00000034774		A0A0B4J1F0	Integrin alpha 11	408.7	169.4
00000035203	Q62190		Macrophage-stimulating protein receptor (CD136)	164.3	-

Ensembl Transcript ID (ENSMUST#)	Uniprot Accession (Swiss-Prot)	Uniprot Accession (TrEMBL)	Description	FC Infl.	FC Conv.
00000037261	Q9Z2J6		Prostaglandin D2 receptor 2 (CD294)	252.8	78.7
00000037882	Q8VBX4		CD207 antigen	32.3	18.5
00000039788	Q8VE98	A6MDC5	CD276 antigen	7.2	5.1
00000043059	Q9QUR8		Semaphorin-7A (CD108)	5.1	6.2
00000044060	Q91ZW9		CD209c antigen	45.1	54.8
00000044165	B8JK39		Integrin alpha 9	272.7	6.5
00000044228		B1ARJ9	CD79B antigen	4.8	4.2
00000047036	Q76KJ5		CD3E antigen, epsilon polypeptide associated protein	2.2	-
00000049681	Q8VDV0		Integrin, beta-like 1	22.3	92.6
00000053577	Q99JW5		Epithelial cell adhesion molecule (CD326)	19.6	-
00000054414	P51680		C-C chemokine receptor type 4 (CD194)	55.4	15.6
00000056117	Q62469		Integrin alpha 2	14.8	11.2
00000058755	Q61088	Q3V1B2	Frizzled-4 (CD344)	23.5	19.5
00000059888	Q9Z0T9		Integrin beta 6	118.6	348.7
00000061673	Q3V3R4		Integrin alpha 1	3.6	8.1
00000062572	Q9R216		Frizzled-9 (CD349)	307.8	79.2
00000065248	P10300	Q3TEK8	CD8 antigen, beta chain 1	3.8	-
00000075017	P13372	Q545E0	Immunoglobulin iota chain (CD179a)	82.3	82.1
00000070630	Q91V98		CD248 antigen, endosialin	568.9	24.8
00000081880	Q7TST5		Lysosome-associated membrane glycoprotein 3 (CD208)	48.0	25.5
00000084086	Q8CJ91	Q3UTX5	CD209b antigen	83.6	78.5
00000093812	Q8R422	A6MDD3	CD109 antigen	10.9	12.5
00000095412	O54709		NKG2-D type II integral membrane protein (CD314)	12.6	-
00000099112		G3X9Q1	Integrin alpha 7	44.8	18.6
00000099972		Q792F9	Integrin alpha 4	-	2.3
00000100335	Q64449		C-type mannose receptor 2 (CD280)	106.5	29.1
00000102537		Q8BS01	Integrin alpha E, epithelial-associated	10.9	12.8
00000102694	Q9WV91		Prostaglandin F2 receptor negative regulator (CD315)	4.8	-
00000102805	Q6XJV6		Cell surface glycoprotein CD200 receptor 2	663.5	168.0
00000102827	Q9WVS0		Inducible T-cell costimulator (CD278)	4.1	3.8
00000105672	Q8R037		Secreted CD137 antigen	22.3	-
00000106237	Q3V0T4		Integrin, alpha D	135.8	190.3
00000106458	A2A863		Integrin beta 4	6.8	-
00000106460	A2A863		Integrin beta 4	482.7	299.0
00000106461	A2A863		Integrin beta 4	16.1	-

| Ensembl Transcript ID (ENSMUST#)	Uniprot Accession (Swiss-Prot)	Uniprot Accession (TrEMBL)	Description	FC Infl.	FC Conv.
00000106796	Q08481		Platelet endothelial cell adhesion molecule (CD31)	40.1	-
00000107164	Q1ERP8		CD300 antigen like family member G	115.5	-
00000107739	Q62470		Integrin alpha 3	-	33.3
00000108551	O35930	A2CFB8	Platelet glycoprotein Ib alpha chain (CD42b)	63.9	54.2
00000111131		A2BI28	CD59b antigen	12.5	19.9
00000111132		Q6PBG1	CD59b antigen	210.4	230.6
00000112101	Q61739		Integrin alpha 6	98.0	-
00000112477	Q2TB54		High affinity immunoglobulin alpha and immunoglobulin mu Fc receptor (CD351)	119.4	100.1
00000112517		B1AYG6	Integrin beta 6	86.8	93.5
00000112576	Q9WUL5	Q3U304	Programmed cell death 1 ligand 2 (CD273)	29.9	27.2
00000114586		I7HJQ8	CD99 antigen-like 2	106.7	130.7
00000117102	Q8BKG4	Q149J3	Frizzled-10 (CD350)	325.8	127.6
00000117301	Q03146		Epithelial discoidin domain-containing receptor 1 (CD167a)	185.1	264.5
00000119365		E9PXZ3	Integrin, alpha 10	60.6	37.8
00000120375	Q62470		Integrin alpha 3	8.3	-
00000135386			CD84 antigen	3.4	-
00000136687	D3YZA5		Integrin-linked protein kinase (Fragment)	9.7	-
00000142615		Q3TWL5	Integrin alpha FG-GAP repeat containing 2	122.3	36.3
00000149150		F7DB81	Integrin alpha 9	22.3	146.6
00000155249		F6VSK8	Integrin alpha 6	352.6	60.1
00000163230		Q80VX2	CD200 antigen	219.6	13.5
00000163494	P16297		Interleukin-2 receptor subunit beta (CD122)	37.0	56.2
00000165365	Q8VE98	A6MDC5	CD276 antigen	-	40.9
00000166731	E9Q7T2		Cell surface glycoprotein CD200 receptor 3	44.3	125.5
00000168071	Q5MPX5		Activated leukocyte cell adhesion molecule soluble isoform (CD166)	51.7	60.3
00000168162	P26618		Platelet-derived growth factor receptor alpha (CD140a)	502.5	370.9
00000168176	O55186	A2BI31	CD59a antigen	222.4	-
00000170051	Q08857	Q3UAI3	CD36 antigen	12.3	-
00000179828	Q04683		C-X-C chemokine receptor type 5 (CD185)	99.9	-
00000180093	Q9WUT7		C-C chemokine receptor type 9 (CDw199)	717.6	97.5
00000189718		Q3UP36	CD22 antigen	140.8	70.0
00000194134	Q61391		CD10 antigen	166.9	378.5

Ensembl Transcript ID (ENSMUST#)	Uniprot Accession (Swiss-Prot)	Uniprot Accession (TrEMBL)	Description	FC Infl.	FC Conv.
00000194690	Q62371		CD167b antigen	1782.9	31.2
00000195656	Q9ET39		SLAM family member 6 (CD352)	34.7	12.1

Appendix 4.5. List of miRNAs found in all MDSC and exosomes samples

miRNAs					
let-7d	mir-16-1	mir-26a-2	mir-340	mir-6240	Gm23632
let-7g	mir-181a-2	mir-26b	mir-3473f	mir-652	Gm27512
let-7i	mir-186	mir-27a	mir-351	mir-709	Gm25190
mir-10a	mir-192	mir-27b	mir-423	mir-92a-1	Gm25442
mir-1198	mir-1928	mir-28a	mir-425	Gm23935	Gm28017
mir-140	mir-1954	mir-3057	mir-451a	Gm24270	Gm27468
mir-142b	mir-1957a	mir-30a	mir-451b	Gm24715	Gm27864
mir-146a	mir-1983	mir-30d	mir-484	Gm24245	Gm27773
mir-146b	mir-21a	mir-30e	mir-486a	Gm22655	Gm27399
mir-148a	mir-22	mir-30f	mir-5099	Gm25960	Gm27284
mir-148b	mir-221	mir-320	mir-5100	Gm25732	
mir-151	mir-223	mir-328	mir-582	Gm24083	
mir-15a	mir-24-2	mir-339	mir-6236	Gm25301	

Appendix 4.6. Distribution of spectral counts obtained for MDSC and MDSC-derived exosomes from tumor-bearing mice under conventional and inflammatory condition.

145
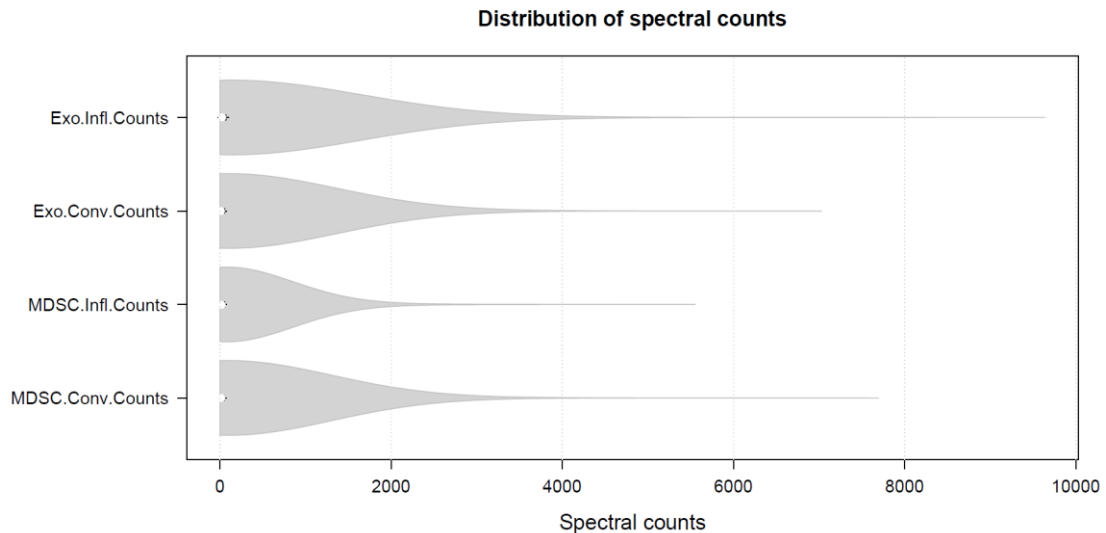
Appendix 4.7. List of genes that correspond to the KEGG "pathways in cancer", including the mRNA, miRNA and proteins found in greater abundance in exosomes compared to MDSC, irrespective of their inflammation condition. Species that are marked as Y were found and N were not.

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 1(Appl1)	N	N	
Adenomatosis polyposis coli 2(Apc2)	Y	N	
Adenomatosis polyposis coli(Apc)	N	N	
Adenylate cyclase 1(Adcy1)	Y	N	
Adenylate cyclase 2(Adcy2)	Y	N	
Adenylate cyclase 3(Adcy3)	Y	N	
Adenylate cyclase 4(Adcy4)	N	N	
Adenylate cyclase 5(Adcy5)	Y	N	
Adenylate cyclase 6(Adcy6)	Y	N	
Adenylate cyclase 7(Adcy7)	N	N	
Adenylate cyclase 8(Adcy8)	Y	N	
Adenylate cyclase 9(Adcy9)	N	N	
Androgen receptor(Ar)	Y	N	
Angiotensin II receptor, type 1a(Agtr1a)	Y	N	
Angiotensin II receptor, type 1b(Agtr1b)	Y	N	
Araf proto-oncogene, serine/threonine kinase(Araf)	N	N	
Aryl hydrocarbon receptor nuclear translocator 2(Arnt2)	Y	N	
Aryl hydrocarbon receptor nuclear translocator(Arnt)	N	N	
Axin 1(Axin1)	N	N	
Axin 2(Axin2)	Y	N	
B cell leukemia/lymphoma 2(Bcl2)	Y	N	
Baculoviral IAP repeat-containing 2(Birc2)	N	N	
Baculoviral IAP repeat-containing 3(Birc3)	Y	N	
Baculoviral IAP repeat-containing 5(Birc5)	Y	N	
Baculoviral IAP repeat-containing 7 (livin)(Birc7)	N	N	
BCL2-associated agonist of cell death(Bad)	N	N	
BCL2-associated X protein(Bax)	N	N	
BCL2-like 1(Bcl2l1)	Y	N	
BH3 interacting domain death agonist(Bid)	N	N	
Bone morphogenetic protein 2(Bmp2)	Y	N	
Bone morphogenetic protein 4(Bmp4)	Y	N	
Bradykinin receptor, beta 1(Bdkrb1)	Y	N	
Bradykinin receptor, beta 2(Bdkrb2)	Y	N	
Braf transforming gene(Braf)	N	N	
Breakpoint cluster region(Bcr)	N	N	
Breast cancer 2, early onset(Brca2)	Y	N	
C-abl oncogene 1, non-receptor tyrosine kinase(Abl1)	Y	N	
Cadherin 1(Cdh1)	N	N	
Casitas B-lineage lymphoma b(Cblb)	N	N	
Casitas B-lineage lymphoma c(Cblc)	N	N	
Casitas B-lineage lymphoma(Cbl)	N	N	
Caspase 3(Casp3)	N	Y	
Caspase 8(Casp8)	Y	N	
Caspase 9(Casp9)	N	Y	

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Catenin (cadherin associated protein), alpha 1(Ctnna1)	N	N	
Catenin (cadherin associated protein), alpha 2(Ctnna2)	Y	N	
Catenin (cadherin associated protein), alpha 3(Ctnna3)	Y	N	
Catenin (cadherin associated protein), beta 1(Ctnnb1)	N	N	
CCAAT/enhancer binding protein (C/EBP), alpha(Cebpa)	N	N	miR-690
CDC28 protein kinase 1b(Cks1b)	N	N	
CDC28 protein kinase regulatory subunit 2(Cks2)	N	N	
Cell division cycle 42(Cdc42)	N	Y	
Chemokine (C-X-C motif) ligand 12(Cxcl12)	Y	N	
Chemokine (C-X-C motif) receptor 4(Cxcr4)	N	N	
Coagulation factor II (thrombin) receptor(f2r)	Y	N	
Coagulation factor II (thrombin) receptor-like 3(f2rl3)	Y	N	
Coiled-coil domain containing 6(Ccdc6)	Y	N	
Collagen, type IV, alpha 1(Col4a1)	Y	N	
Collagen, type IV, alpha 2(Col4a2)	Y	N	
Collagen, type IV, alpha 3(Col4a3)	Y	N	
Collagen, type IV, alpha 4(Col4a4)	Y	N	
Collagen, type IV, alpha 5(Col4a5)	Y	N	
Collagen, type IV, alpha 6(Col4a6)	Y	N	
Colony stimulating factor 1 receptor(Csf1r)	N	Y	
Colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)(Csf2ra)	N	N	
Colony stimulating factor 3 receptor (granulocyte)(Csf3r)	N	N	
Conserved helix-loop-helix ubiquitous kinase(Chuk)	N	N	
CREB binding protein(Crebbp)	N	N	
C-terminal binding protein 1(Ctbp1)	N	N	
C-terminal binding protein 2(Ctbp2)	Y	N	
Cullin 2(Cul2)	N	N	
Cyclin D1(Ccnd1)	Y	N	
Cyclin E1(Ccne1)	N	N	
Cyclin E2(Ccne2)	N	N	
Cyclin-dependent kinase 2(Cdk2)	N	N	
Cyclin-dependent kinase 4(Cdk4)	Y	N	
Cyclin-dependent kinase 6(Cdk6)	Y	N	
Cyclin-dependent kinase inhibitor 1A (P21)(Cdkn1a)	N	N	
Cyclin-dependent kinase inhibitor 1B(Cdkn1b)	N	N	
Cyclin-dependent kinase inhibitor 2A(Cdkn2a)	N	N	
Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)(Cdkn2b)	Y	N	
Cytochrome c, somatic(Cycs)	Y	N	
Cytochrome c, testis(Cyct)	N	N	
Death associated protein kinase 1(Dapk1)	Y	N	
Death-associated protein kinase 2(Dapk2)	N	N	
Death-associated protein kinase 3(Dapk3)	Y	N	
Deleted in colorectal carcinoma(Dcc)	Y	N	
Dishevelled segment polarity protein 1(Dvl1)	N	N	
Dishevelled segment polarity protein 2(Dvl2)	Y	N	
Dishevelled segment polarity protein 3(Dvl3)	N	N	
E1A binding protein p300(Ep300)	N	N	
E2F transcription factor 1(E2f1)	Y	N	

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
E2F transcription factor 2(E2f2)	N	N	
E2F transcription factor 3(E2f3)	N	N	
Egl-9 family hypoxia-inducible factor 1(Egln1)	N	N	
Egl-9 family hypoxia-inducible factor 2(Egln2)	Y	N	
Egl-9 family hypoxia-inducible factor 3(Egln3)	N	N	
Endothelial PAS domain protein 1(Epas1)	Y	N	
Endothelin receptor type A(Ednra)	Y	N	
Endothelin receptor type B(Ednrb)	Y	N	
Epidermal growth factor receptor(Egfr)	Y	N	
Epidermal growth factor(Egf)	Y	N	
Erb-b2 receptor tyrosine kinase 2(Erbb2)	Y	N	
Fas (TNF receptor superfamily member 6)(Fas)	N	N	miR-98/let-7
Fas (TNFRSF6)-associated via death domain(Fadd)	Y	N	miR-155
Fas ligand (TNF superfamily, member 6)(Fasl)	Y	N	
FBJ osteosarcoma oncogene(Fos)	N	N	
Fibroblast growth factor 1(Fgf1)	Y	N	
Fibroblast growth factor 10(Fgf10)	Y	N	
Fibroblast growth factor 11(Fgf11)	Y	N	
Fibroblast growth factor 12(Fgf12)	Y	N	
Fibroblast growth factor 13(Fgf13)	Y	N	
Fibroblast growth factor 14(Fgf14)	Y	N	
Fibroblast growth factor 15(Fgf15)	Y	N	
Fibroblast growth factor 16(Fgf16)	Y	N	
Fibroblast growth factor 17(Fgf17)	Y	N	
Fibroblast growth factor 18(Fgf18)	Y	N	
Fibroblast growth factor 2(Fgf2)	Y	N	
Fibroblast growth factor 20(Fgf20)	Y	N	
Fibroblast growth factor 21(Fgf21)	Y	N	
Fibroblast growth factor 22(Fgf22)	Y	N	
Fibroblast growth factor 23(Fgf23)	Y	N	
Fibroblast growth factor 3(Fgf3)	Y	N	
Fibroblast growth factor 4(Fgf4)	Y	N	
Fibroblast growth factor 5(Fgf5)	Y	N	
Fibroblast growth factor 6(Fgf6)	Y	N	
Fibroblast growth factor 7(Fgf7)	Y	N	
Fibroblast growth factor 8(Fgf8)	Y	N	
Fibroblast growth factor 9(Fgf9)	Y	N	
Fibroblast growth factor receptor 1(Fgfr1)	Y	N	
Fibroblast growth factor receptor 2(Fgfr2)	Y	N	
Fibroblast growth factor receptor 3(Fgfr3)	N	N	
Fibronectin 1(Fn1)	N	Y	
FMS-like tyrosine kinase 3 ligand(Flt3l)	N	N	
FMS-like tyrosine kinase 3(Flt3)	N	N	
Forkhead box O1(Foxo1)	Y	N	
Frizzled class receptor 1(Fzd1)	Y	N	
Frizzled class receptor 10(Fzd10)	Y	N	
Frizzled class receptor 2(Fzd2)	Y	N	
Frizzled class receptor 3(Fzd3)	Y	N	
Frizzled class receptor 4(Fzd4)	Y	N	

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Frizzled class receptor 5(Fzd5)	Y	N	
Frizzled class receptor 6(Fzd6)	Y	N	
Frizzled class receptor 7(Fzd7)	Y	N	
Frizzled class receptor 8(Fzd8)	Y	N	
Frizzled class receptor 9(Fzd9)	Y	N	
Fumarate hydratase 1(Fh1)	Y	N	
GLI-Kruppel family member GLI1(Gli1)	Y	N	
GLI-Kruppel family member GLI2(Gli2)	Y	N	
GLI-Kruppel family member GLI3(Gli3)	Y	N	
Glycogen synthase kinase 3 beta(Gsk3b)	N	N	
GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus(Gnas)	Y	N	
Growth factor receptor bound protein 2(Grb2)	N	N	
Guanine nucleotide binding protein (G protein), alpha inhibiting 1(Gnai1)	Y	N	
Guanine nucleotide binding protein (G protein), alpha inhibiting 2(Gnai2)	N	N	
Guanine nucleotide binding protein (G protein), alpha inhibiting 3(Gnai3)	N	N	
Guanine nucleotide binding protein (G protein), beta 1(Gnb1)	N	N	
Guanine nucleotide binding protein (G protein), beta 2(Gnb2)	N	N	
Guanine nucleotide binding protein (G protein), beta 3(Gnb3)	Y	N	
Guanine nucleotide binding protein (G protein), beta 4(Gnb4)	Y	N	
Guanine nucleotide binding protein (G protein), beta 5(Gnb5)	N	N	
Guanine nucleotide binding protein (G protein), gamma 10(Gng10)	N	N	
Guanine nucleotide binding protein (G protein), gamma 11(Gng11)	N	N	
Guanine nucleotide binding protein (G protein), gamma 12(Gng12)	N	N	
Guanine nucleotide binding protein (G protein), gamma 13(Gng13)	Y	N	
Guanine nucleotide binding protein (G protein), gamma 2(Gng2)	N	N	
Guanine nucleotide binding protein (G protein), gamma 3(Gng3)	Y	N	
Guanine nucleotide binding protein (G protein), gamma 4(Gng4)	Y	N	
Guanine nucleotide binding protein (G protein), gamma 5(Gng5)	N	N	
Guanine nucleotide binding protein (G protein), gamma 7(Gng7)	N	N	
Guanine nucleotide binding protein (G protein), gamma 8(Gng8)	Y	N	
Guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 1(Gngt1)	Y	N	
Guanine nucleotide binding protein (G protein), gamma transducing activity polypeptide 2(Gngt2)	N	N	
Guanine nucleotide binding protein, alpha 11(Gna11)	N	N	
Guanine nucleotide binding protein, alpha 12(Gna12)	N	N	
Guanine nucleotide binding protein, alpha 13(Gna13)	N	N	
Guanine nucleotide binding protein, alpha q polypeptide(Gnaq)	N	N	
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5(LOC102641276)	N	N	
Harvey rat sarcoma virus oncogene(Hras)	Y	N	
Heat shock protein 90 alpha (cytosolic), class B member 1(Hsp90ab1)	N	N	

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Heat shock protein 90, alpha (cytosolic), class A member 1(Hsp90aa1)	N	N	
Heat shock protein 90, beta (Grp94), member 1(Hsp90b1)	N	N	
Hedgehog-interacting protein(Hhip)	Y	N	
Hepatocyte growth factor(Hgf)	N	N	
Histone deacetylase 1(Hdac1)	N	N	
Histone deacetylase 2(Hdac2)	Y	N	
Hypoxia inducible factor 1, alpha subunit(Hif1a)	N	N	
Inhibitor of kappab kinase beta(Ikbkb)	N	N	miR-155, miR-199
Inhibitor of kappab kinase gamma(Ikbkg)	N	N	
Insulin-like growth factor 1(Igf1)	Y	N	
Insulin-like growth factor I receptor(Igf1r)	N	N	
Integrin alpha 2(Itga2)	Y	N	
Integrin alpha 2b(Itga2b)	N	Y	
Integrin alpha 3(Itga3)	Y	N	
Integrin alpha 6(Itga6)	Y	Y	
Integrin alpha V(Itgav)	N	N	
Integrin beta 1 (fibronectin receptor beta)(Itgb1)	N	Y	
Interleukin 6(Il6)	N	N	
Janus kinase 1(Jak1)	N	N	
Jun proto-oncogene(Jun)	N	N	
Junction plakoglobin(Jup)	N	N	
Kirsten rat sarcoma viral oncogene homolog(Kras)	N	N	
Kit ligand(Kitl)	Y	N	
Kit oncogene(Kit)	N	N	
Laminin B1(Lamb1)	Y	N	
Laminin gamma 3(Lamc3)	Y	N	
Laminin, alpha 1(Lama1)	Y	N	
Laminin, alpha 2(Lama2)	Y	N	
Laminin, alpha 3(Lama3)	Y	N	
Laminin, alpha 4(Lama4)	Y	N	
Laminin, alpha 5(Lama5)	Y	N	
Laminin, beta 2(Lamb2)	Y	N	
Laminin, beta 3(Lamb3)	Y	N	
Laminin, gamma 1(Lamc1)	Y	N	
Laminin, gamma 2(Lamc2)	Y	N	
Lymphoid enhancer binding factor 1(Lef1)	Y	N	
Lysophosphatidic acid receptor 1(Lpar1)	Y	N	
Lysophosphatidic acid receptor 2(Lpar2)	N	N	
Lysophosphatidic acid receptor 3(Lpar3)	Y	N	
Lysophosphatidic acid receptor 4(Lpar4)	Y	N	
Lysophosphatidic acid receptor 5(Lpar5)	Y	N	
Lysophosphatidic acid receptor 6(Lpar6)	Y	N	
Matrix metallopeptidase 1a (interstitial collagenase)(Mmp1a)	N	N	
Matrix metallopeptidase 1b (interstitial collagenase)(Mmp1b)	N	N	
Matrix metallopeptidase 2(Mmp2)	Y	N	
Matrix metallopeptidase 9(Mmp9)	N	Y	
Max protein(Max)	N	N	
MDS1 and EVI1 complex locus(Mecom)	Y	N	

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Mechanistic target of rapamycin (serine/threonine kinase)(Mtor)	N	N	
Met proto-oncogene(Met)	Y	N	
Microphthalmia-associated transcription factor(Mitf)	Y	N	
Mitogen-activated protein kinase 1(Mapk1)	N	Y	
Mitogen-activated protein kinase 10(Mapk10)	Y	N	
Mitogen-activated protein kinase 3(Mapk3)	N	Y	
Mitogen-activated protein kinase 8(Mapk8)	N	N	
Mitogen-activated protein kinase 9(Mapk9)	Y	N	
Mitogen-activated protein kinase kinase 1(Map2k1)	N	N	
Mitogen-activated protein kinase kinase 2(Map2k2)	N	N	
Mutl homolog 1(Mlh1)	Y	N	
Muts homolog 2(Msh2)	N	N	
Muts homolog 3(Msh3)	Y	N	
Muts homolog 6(Msh6)	N	N	
Myelocytomatosis oncogene(Myc)	Y	N	miR-98/let-7
Neuroblastoma ras oncogene(Nras)	N	N	
Neurotrophic tyrosine kinase, receptor, type 1(Ntrk1)	Y	N	
Nitric oxide synthase 2, inducible(Nos2)	N	N	
NK-3 transcription factor, locus 1 (Drosophila)(Nkx3-1)	Y	N	
Nuclear factor of kappa light polypeptide gene enhancer in B cells 1, p105(Nfkb1)	N	N	miR-9
Nuclear factor of kappa light polypeptide gene enhancer in B cells 2, p49/p100(Nfkb2)	N	N	
Nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor, alpha(Nfkbia)	N	N	
Nuclear receptor coactivator 4(Ncoa4)	N	N	
Paired box 8(Pax8)	Y	N	
Patched 1(Ptch1)	Y	N	
Patched 2(Ptch2)	Y	N	
Peroxisome proliferator activated receptor gamma(Pparg)	Y	N	
Peroxisome proliferator activator receptor delta(Ppard)	N	N	
Phosphatase and tensin homolog(Pten)	N	N	miR-494 miR-21a miR-155
Phosphatidylinositol 3 kinase, regulatory subunit, polypeptide 3 (p55)(Pik3r3)	Y	N	
Phosphatidylinositol 3-kinase catalytic delta polypeptide(Pik3cd)	Y	N	
Phosphatidylinositol 3-kinase, catalytic, alpha polypeptide(Pik3ca)	Y	N	
Phosphatidylinositol 3-kinase, catalytic, beta polypeptide(Pik3cb)	Y	N	
Phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)(Pik3r1)	Y	N	
Phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 2 (p85 beta)(Pik3r2)	N	N	
Phosphoinositide-3-kinase, catalytic, gamma polypeptide(Pik3cg)	N	N	
Phosphoinositide-3-kinase, regulatory subunit 5, p101(Pik3r5)	N	N	
Phospholipase C, beta 1(Plcb1)	Y	N	
Phospholipase C, beta 2(Plcb2)	N	N	
Phospholipase C, beta 3(Plcb3)	N	N	

151

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Phospholipase C, beta 4(Plcb4)	Y	N	
Phospholipase C, gamma 1(Plcg1)	Y	N	
Phospholipase C, gamma 2(Plcg2)	N	N	
Placental growth factor(Pgf)	Y	N	
Platelet derived growth factor receptor, alpha polypeptide(Pdgfra)	Y	N	
Platelet derived growth factor receptor, beta polypeptide(Pdgfrb)	Y	N	
Platelet derived growth factor, alpha(Pdgfa)	Y	N	
Platelet derived growth factor, B polypeptide(Pdgfb)	Y	N	
Pleckstrin homology domain containing, family G (with rhogef domain) member 5(Plekhg5)	Y	N	
Predicted gene 15776(Gm15776)	N	N	
Predicted gene 5741(Gm5741)	N	N	
Predicted gene 9840(Gm9840)	N	N	
Promyelocytic leukemia(Pml)	Y	N	
Prostaglandin E receptor 1 (subtype EP1)(Ptger1)	N	N	
Prostaglandin E receptor 2 (subtype EP2)(Ptger2)	N	N	
Prostaglandin E receptor 3 (subtype EP3)(Ptger3)	N	N	
Prostaglandin E receptor 4 (subtype EP4)(Ptger4)	N	N	
Prostaglandin-endoperoxide synthase 2(Ptgs2)	Y	N	
Protein inhibitor of activated STAT 2(Pias2)	Y	N	
Protein kinase C, alpha(Prkca)	Y	N	
Protein kinase C, beta(Prkcb)	N	N	
Protein kinase C, gamma(Prkcg)	Y	N	
Protein kinase, camp dependent, catalytic, alpha(Prkaca)	N	N	
Protein kinase, camp dependent, catalytic, beta(Prkacb)	Y	N	
PTK2 protein tyrosine kinase 2(Ptk2)	Y	N	
RAD51 recombinase(Rad51)	N	N	
Ral guanine nucleotide dissociation stimulator(Ralgds)	Y	N	
Rala binding protein 1(Ralbp1)	Y	N	
Ras association (ralgds/AF-6) domain family member 1(Rassf1)	N	N	
Ras association (ralgds/AF-6) domain family member 5(Rassf5)	N	N	
RAS guanyl releasing protein 1(Rasgrp1)	Y	N	
RAS guanyl releasing protein 4(Rasgrp4)	N	N	
Ras homolog family member A(Rhoa)	N	Y	
RAS, guanyl releasing protein 2(Rasgrp2)	Y	N	
RAS, guanyl releasing protein 3(Rasgrp3)	Y	N	
RAS-related C3 botulinum substrate 1(Rac1)	N	Y	
RAS-related C3 botulinum substrate 2(Rac2)	N	Y	
RAS-related C3 botulinum substrate 3(Rac3)	Y	Y	
Ret proto-oncogene(Ret)	Y	N	
Retinoblastoma 1(Rb1)	N	N	
Retinoic acid receptor, alpha(Rara)	N	N	
Retinoic acid receptor, beta(Rarb)	Y	N	
Retinoid X receptor alpha(Rxra)	Y	N	
Retinoid X receptor beta(Rxrb)	N	N	
Retinoid X receptor gamma(Rxrg)	Y	N	
Rho guanine nucleotide exchange factor (GEF) 1(Arhgef1)	N	N	
Rho guanine nucleotide exchange factor (GEF) 11(Arhgef11)	N	N	
Rho guanine nucleotide exchange factor (GEF) 12(Arhgef12)	Y	N	

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Rho-associated coiled-coil containing protein kinase 1(Rock1)	N	N	
Rho-associated coiled-coil containing protein kinase 2(Rock2)	N	N	
RIKEN cdna 4930544G11 gene(4930544G11Rik)	Y	Y	
Ring-box 1(Rbx1)	N	N	
Runt related transcription factor 1(Runx1)	N	N	miR-9
Runt-related transcription factor 1; translocated to, 1 (cyclin D-related)(Runx1t1)	Y	N	
Serine/threonine kinase 36(Stk36)	N	N	
Serine/threonine kinase 4(Stk4)	N	N	
Signal transducer and activator of transcription 1(Stat1)	N	N	miR-146a
Signal transducer and activator of transcription 3(Stat3)	Y	N	
Signal transducer and activator of transcription 5A(Stat5a)	Y	N	
Signal transducer and activator of transcription 5B(Stat5b)	N	N	
SMAD family member 2(Smad2)	N	N	
SMAD family member 3(Smad3)	N	N	
SMAD family member 4(Smad4)	N	N	
Smoothened, frizzled class receptor(Smo)	Y	N	
Solute carrier family 2 (facilitated glucose transporter), member 1(Slc2a1)	N	N	
Son of sevenless homolog 1 (Drosophila)(Sos1)	N	N	
Son of sevenless homolog 2 (Drosophila)(Sos2)	N	N	
Sonic hedgehog(Shh)	Y	N	
S-phase kinase-associated protein 2 (p45)(Skp2)	Y	N	
Spleen focus forming virus (SFFV) proviral integration oncogene(Spi1)	N	N	miR-155
Suppressor of fused homolog (Drosophila)(Sufu)	Y	N	
Thymoma viral proto-oncogene 1(Akt1)	Y	N	
Thymoma viral proto-oncogene 2(Akt2)	Y	N	
Thymoma viral proto-oncogene 3(Akt3)	Y	N	
TNF receptor associated factor 4(Traf4)	N	N	
TNF receptor-associated factor 1(Traf1)	Y	N	
TNF receptor-associated factor 2(Traf2)	N	N	
TNF receptor-associated factor 3(Traf3)	N	N	
TNF receptor-associated factor 5(Traf5)	N	N	
TNF receptor-associated factor 6(Traf6)	N	N	miR-146a
Transcription elongation factor B (SIII), polypeptide 1(Tceb1)	N	N	
Transcription elongation factor B (SIII), polypeptide 2(Tceb2)	N	N	
Transcription elongation factor B polypeptide 1(LOC102642819)	N	N	
Transcription factor 7 like 1 (T cell specific, HMG box)(Tcf7l1)	Y	N	
Transcription factor 7 like 2, T cell specific, HMG box(Tcf7l2)	Y	N	
Transcription factor 7, T cell specific(Tcf7)	N	N	
Transformation related protein 53(Trp53)	N	N	
Transformed mouse 3T3 cell double minute 2(Mdm2)	N	N	
Transforming growth factor alpha(Tgfa)	Y	N	
Transforming growth factor, beta 1(Tgfb1)	N	Y	
Transforming growth factor, beta 2(Tgfb2)	Y	N	
Transforming growth factor, beta 3(Tgfb3)	Y	N	
Transforming growth factor, beta receptor I(Tgfbr1)	Y	N	
Transforming growth factor, beta receptor II(Tgfbr2)	N	N	
Translocated promoter region, nuclear basket protein(Tpr)	N	N	

153

David Gene Name in Pathway	mRNA found?	Protein found?	Putative miRNA target
Trk-fused gene(Tfg)	N	N	
Tropomyosin 3, gamma(Tpm3)	N	N	
Vascular endothelial growth factor A(Vegfa)	Y	N	
Vascular endothelial growth factor B(Vegfb)	N	N	
Vascular endothelial growth factor C(Vegfc)	Y	N	
Vascular endothelial growth factor D(Vegfd)	N	N	
V-crk avian sarcoma virus CT10 oncogene homolog(Crk)	N	N	
V-crk avian sarcoma virus CT10 oncogene homolog-like(Crkl)	N	N	
Von Hippel-Lindau tumor suppressor(Vhl)	N	N	
V-raf-leukemia viral oncogene 1(Raf1)	Y	N	
V-ral simian leukemia viral oncogene A (ras related)(Rala)	N	N	
V-ral simian leukemia viral oncogene B(Ralb)	N	N	
V-rel reticuloendotheliosis viral oncogene homolog A (avian)(Rela)	N	N	
Wingless-type MMTV integration site family, member 1(Wnt1)	Y	N	
Wingless-type MMTV integration site family, member 10A(Wnt10a)	Y	N	
Wingless-type MMTV integration site family, member 10B(Wnt10b)	Y	N	
Wingless-type MMTV integration site family, member 11(Wnt11)	Y	N	
Wingless-type MMTV integration site family, member 16(Wnt16)	Y	N	
Wingless-type MMTV integration site family, member 2(Wnt2)	Y	N	
Wingless-type MMTV integration site family, member 2B(Wnt2b)	Y	N	
Wingless-type MMTV integration site family, member 3(Wnt3)	Y	N	
Wingless-type MMTV integration site family, member 3A(Wnt3a)	Y	N	
Wingless-type MMTV integration site family, member 4(Wnt4)	Y	N	
Wingless-type MMTV integration site family, member 5A(Wnt5a)	Y	N	
Wingless-type MMTV integration site family, member 5B(Wnt5b)	Y	N	
Wingless-type MMTV integration site family, member 6(Wnt6)	Y	N	
Wingless-type MMTV integration site family, member 7A(Wnt7a)	Y	N	
Wingless-type MMTV integration site family, member 7B(Wnt7b)	Y	N	
Wingless-type MMTV integration site family, member 8A(Wnt8a)	Y	N	
Wingless-type MMTV integration site family, member 8B(Wnt8b)	Y	N	
Wingless-type MMTV integration site family, member 9A(Wnt9a)	Y	N	
Wingless-type MMTV integration site family, member 9B(Wnt9b)	Y	N	
X-linked inhibitor of apoptosis(Xiap)	N	N	
Zinc finger and BTB domain containing 16(Zbtb16)	Y	N	

Bibliography

(1)   Parker, K. H.; Beury, D. W.; Ostrand-Rosenberg, S. In *Adv. Cancer Res.*;
         2015; Vol. 128, pp 95–139.

(2)   Balkwill, F. R.; Capasso, M.; Hagemann, T. *J. Cell Sci.* **2012**, *125* (23), 5591–
         5596.

(3)   Burke, M.; Choksawangkarn, W.; Edwards, N.; Ostrand-Rosenberg, S.;
         Fenselau, C. *J. Proteome Res.* **2014**, *13* (2), 836–843.

(4)   Bunt, S. K.; Sinha, P.; Clements, V. K.; Leips, J.; Ostrand-Rosenberg, S. *J.
         Immunol.* **2006**, *176* (1), 284–290.

(5)   Bunt, S. K.; Yang, L.; Sinha, P.; Clements, V. K.; Leips, J.; Ostrand-
         Rosenberg, S. *Cancer Res.* **2007**, *67* (20), 10019–10026.

(6)   Sinha, P.; Clements, V. K.; Fulton, A. M.; Ostrand-Rosenberg, S. *Cancer Res.*
         2007, *67* (9), 4507–4513.

(7)   Ezernitchi, A. V; Vaknin, I.; Cohen-Daniel, L.; Levy, O.; Manaster, E.; Halabi,
         A.; Pikarsky, E.; Shapira, L.; Baniyash, M. *J. Immunol.* **2006**, *177* (7), 4763–
         4772.

(8)   Deribe, Y. L.; Pawson, T.; Dikic, I. *Nat. Struct. Mol. Biol.* **2010**, *17* (6), 666–
         672.

(9)   Liu, J.; Qian, C.; Cao, X. *Immunity* **2016**, *45* (1), 15–30.

(10)   Karve, T. M.; Cheema, A. K. *J. Amino Acids* **2011**, *2011*, 1–13.

(11)   Colombo, M.; Raposo, G.; Théry, C. *Annu. Rev. Cell Dev. Biol* **2014**, *30*, 255–
         289.

(12)   Chen, S.; Zhang, Y.; Kuzel, T. M.; Zhang, B. *Cancer Cell Microenviron.* **2015**, *2* (1), e637.

(13)   Thind, A.; Wilson, C. *J. Extracell. Vesicles* **2016**, *5*, 31292.

(14)   EL Andaloussi, S.; Mäger, I.; Breakefield, X. O.; Wood, M. J. A. *Nat. Rev. Drug Discov.* **2013**, *12* (5), 347–357.

(15)   Tkach, M.; Théry, C. *Cell* **2016**, *164* (6), 1226–1232.

(16)   Marcotte, E. M. *Nat. Biotechnol.* **2007**, *25* (7), 755–757.

(17)   Garcia, B. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (2), 193–202.

(18)   Zhang, H.; Ge, Y. *Circ. Cardiovasc. Genet.* **2011**, *4* (6), 711.

(19)   Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24* (2), 168–200.

(20)   Yates, J. R.; Kelleher, N. L. *Anal. Chem.* **2013**, *85* (13), 6151–6151.

(21)   Dass, C. *Fundamentals of Contemporary Mass Spectrometry*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007.

(22)   Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Anal. Chem.* **2013**, *85* (1), 2–9.

(23)   Makarov, A. Mass spectrometer. U.S. Patent 5,588,346, 1999.

(24)   Makarov, A. *Anal. Chem.* **2000**, *72* (6), 1156–1162.

(25)   Perry, R. H.; Cooks, R. G.; Noll, R. J. *Mass Spectrom. Rev.* **2008**, *27* (6), 661–699.

(26)   Knight, R. D. *Appl. Phys. Lett.* **1981**, *38* (4), 221–223.

(27)   Gillig, K. J.; Bluhm, B. K.; Russell, D. H. *Int. J. Mass Spectrom. Ion Process.* **1996**, *157/158*, 129–147.

(28)   Zubarev, R.; Makarov, A. *Anal. Chem.* **2013**, *85* (11), 5288–5296.

156

(29) Makarov, A.; Denisov, E.; Lange, O. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1391–1396.

(30) Eliuk, S.; Makarov, A. *Annu. Rev. Anal. Chem.* **2015**, *8* (1), 61–80.

(31) Mullen, C.; Earley, L.; Weisbrod, C.; Syka, J. E. P.; Dunyach, J.; Scientific, T. F.; Jose, S. *Poster 63rd ASMS Conf.* **2015**, 2015.

(32) Rathore, D.; Aboufazeli, F.; Huang, Y.; Kolli, V.; Fernando, G. S.; Dodds, E. D. *Encyclopedia of Analytical Chemistry*. 2015, pp 1–26.

(33) Mitchell Wells, J.; McLuckey, S. A. In *Methods in enzymology*; 2005; Vol. 402, pp 148–185.

(34) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Biochim. Biophys. Acta* **2006**, *1764* (12), 1811–1822.

(35) Olsen, J. V; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4* (9), 709–712.

(36) McAlister, G. C.; Berggren, W. T.; Griep-Raming, J.; Horning, S.; Makarov, A.; Phanstiel, D.; Stafford, G.; Swaney, D. L.; Syka, J. E. P.; Zabrouskov, V.; Coon, J. J. *J. Proteome Res.* **2008**, *7* (8), 3127–3136.

(37) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. *Anal. Chem.* **2015**, *87* (8), 4152–4158.

(38) Nesvizhskii, A. I. *J. Proteomics* **2010**, *73* (11), 2092–2123.

(39) Kim, S.; Gupta, N.; Pevzner, P. A. *J. Proteome Res.* **2008**, *7* (8), 3354–3363.

(40) Kim, S.; Pevzner, P. A. *Nat. Commun.* **2014**, *5*, 5277.

(41)   Craig, R.; Beavis, R. C. *Bioinformatics* **2004**, *20* (9), 1466–1467.

(42)   MacLean, B.; Eng, J. K.; Beavis, R. C.; McIntosh, M. *Bioinformatics* **2006**, *22* (22), 2830–2832.

(43)   Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *J. Proteome Res.* **2004**, *3* (5), 958–964.

(44)   Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V; Mann, M. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.

(45)   Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. *J. Proteome Res.* **2014**, *13* (8), 3679–3684.

(46)   Tabb, D. L.; Fernando, C. G.; Chambers, M. C. *J. Proteome Res.* **2007**, *6* (2), 654–661.

(47)   Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20* (18), 3551–3567.

(48)   Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(49)   Benjamini, Y.; Hochberg, Y. *J. R. Stat. Soc. B*. 1995, pp 289–300.

(50)   Elias, J. E.; Gygi, S. P. *Nat Methods* **2007**, *4* (3), 207–214.

(51)   Nesvizhskii, A. I. *Mol. Cell. Proteomics* **2005**, *4* (10), 1419–1440.

(52)   Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. *Anal. Chem.* **2003**, *75* (17), 4646–4658.

(53)   Li, Y. F.; Arnold, R. J.; Li, Y.; Radivojac, P.; Sheng, Q.; Tang, H. *J. Comput. Biol.* **2009**, *16* (8), 1183–1193.

(54)   Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.;

Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. *Mol. Cell. Proteomics* **2009**, *8* (11), 2405–2417.

(55)   Edwards, N. J. *Curr. Protoc. Bioinforma.* **2013**, *44* ((Unit 13.23)), 1–23.

(56)   Allet, N.; Barrillat, N.; Baussant, T.; Boiteau, C.; Botti, P.; Bougueleret, L.; Budin, N.; Canet, D.; Carraud, S.; Chiappe, D.; Christmann, N.; Colinge, J.; Cusin, I.; Dafflon, N.; Depresle, B.; Fasso, I.; Frauchiger, P.; Gaertner, H.; Gleizes, A.; Gonzalez-Couto, E.; Jeandenans, C.; Karmime, A.; Kowall, T.; Lagache, S.; Mahé, E.; Masselot, A.; Mattou, H.; Moniatte, M.; Niknejad, A.; Paolini, M.; Perret, F.; Pinaud, N.; Ranno, F.; Raimondi, S.; Reffas, S.; Regamey, P.-O.; Rey, P.-A.; Rodriguez-Tomé, P.; Rose, K.; Rossellat, G.; Saudrais, C.; Schmidt, C.; Villain, M.; Zwahlen, C. *Proteomics* **2004**, *4* (8), 2333–2351.

(57)   Old, W. M. *Mol. Cell. Proteomics* **2005**, *4* (10), 1487–1502.

(58)   Liu, H.; Sadygov, R. G.; Yates, J. R. *Anal. Chem.* **2004**, *76* (14), 4193–4201.

(59)   Fu, X.; Gharib, S. A.; Green, P. S.; Aitken, M. L.; Frazer, D. A.; Park, D. R.; Vaisar, T.; Heinecke, J. W. *J. Proteome Res.* **2008**, *7* (3), 845–854.

(60)   Zybailov, B.; Coleman, M. K.; Florens, L.; Washburn, M. P. *Anal. Chem.* **2005**, *77* (19), 6218–6224.

(61)   Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Anal. Bioanal. Chem.* **2007**, *389* (4), 1017–1031.

(62)   Lundgren, D. H.; Hwang, S.-I.; Wu, L.; Han, D. K. *Expert Rev. Proteomics* **2010**, *7* (1), 39–53.

(63)   Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.;

Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. *Nucleic Acids Res.* **2007**, *35*, W701–W706.

(64)   D. LeDuc, R.; L. Kelleher, N. *Curr. Protoc. Bioinforma.* **2007**, *29* (Unit 13.6), 1–28.

(65)   Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. *Proteomics* **2015**, *15*, 1235–1238.

(66)   Liu, X.; Mammana, A.; Bafna, V. *Bioinformatics* **2012**, *28* (13), 1692–1697.

(67)   Kou, Q.; Xun, L.; Liu, X. *Bioinformatics* **2016**, 1–3.

(68)   Liu, X.; Sirotkin, Y.; Shen, Y.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2012**, *11* (6), M111.008524.

(69)   Kou, Q.; Zhu, B.; Wu, S.; Ansong, C.; Tolić, N.; Paša-Tolić, L.; Liu, X. *J. Proteome Res.* **2016**, *15* (8), 2422–2432.

(70)   Guner, H.; Close, P. L.; Cai, W.; Zhang, H.; Peng, Y.; Gregorich, Z. R.; Ge, Y. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (3), 464–470.

(71)   Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X.; Ge, Y. *Mol. Cell. Proteomics* **2016**, *15* (2), 703–714.

(72)   Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (4), 320–332.

(73)   Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2010**, *9* (12), 2772–2782.

(74)   Zabrouskov, V.; Senko, M. W.; Du, Y.; Leduc, R. D.; Kelleher, N. L. *J. Am.*

Soc. Mass Spectrom. **2005**, *16* (12), 2027–2038.

(75)   Kelleher, N. L.; Thomas, P. M.; Ntai, I.; Compton, P. D.; LeDuc, R. D. *Expert Rev. Proteomics* **2014**, *11* (6), 649–651.

(76)   Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2014**, *86* (10), 4961–4968.

(77)   Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y.-C.; Xu, Q.; Dong, X.; Guner, H.; Wang, S.; Kohmoto, T.; Young, K. H.; Moss, R. L.; Ge, Y. *J. Proteome Res.* **2011**, *10* (9), 4054–4065.

(78)   Ostrand-Rosenberg, S. *Cancer Immunol. Immunother.* **2010**, *59* (10), 1593–1600.

(79)   Gabrilovich, D. I.; Ostrand-Rosenberg, S.; Bronte, V. *Nat. Rev. Immunol.* **2012**, *12* (4), 253–268.

(80)   Gabrilovich, D. I.; Nagaraj, S. *Nat. Rev. Immunol.* **2009**, *9* (3), 162–174.

(81)   Kumar, V.; Patel, S.; Tcyganov, E.; Gabrilovich, D. I. *Trends Immunol.* **2016**, *37* (3), 208–220.

(82)   Sinha, P.; Okoro, C.; Foell, D.; Freeze, H. H.; Ostrand-Rosenberg, S.; Srikrishna, G. *J. Immunol.* **2008**, *181* (7), 4666–4675.

(83)   Boutté, A. M.; McDonald, W. H.; Shyr, Y.; Yang, L.; Lin, P. C. *PLoS One* **2011**, *6* (8), e22446.

(84)   Chornoguz, O.; Grmai, L.; Sinha, P.; Artemenko, K. A.; Zubarev, R. A.; Ostrand-Rosenberg, S. *Mol. Cell. Proteomics* **2011**, *10* (3), M110.002980.

(85)   Choksawangkarn, W.; Graham, L. M.; Burke, M.; Lee, S. B.; Ostrand-

Rosenberg, S.; Fenselau, C.; Edwards, N. J. *Proteomics* **2016**, *16* (13), 1881–1888.

(86) Mathivanan, S.; Ji, H.; Simpson, R. J. *J. Proteomics* **2010**, *73* (10), 1907–1920.

(87) Théry, C.; Amigorena, S.; Raposo, G.; Clayton, A.; Théry, C.; Amigorena, S.; Raposo, G.; Clayton, A. In *Current Protocols in Cell Biology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006; p 3.22.1-3.22.29.

(88) Raposo, G.; Nijman, H. W.; Stoorvogel, W.; Liejendekker, R.; Harding, C. V; Melief, C. J.; Geuze, H. J. *J. Exp. Med.* **1996**, *183* (3), 1161–1172.

(89) Raposo, G.; Stoorvogel, W. *J. Cell Biol.* **2013**, *200* (4), 373–383.

(90) Escola, J.-M.; Kleijmeer, M. J.; Stoorvogel, W.; Griffith, J. M.; Yoshie, O.; Geuze, H. J. *J. Biol. Chem.* **1998**, *273* (32), 20121–20127.

(91) Théry, C.; Zitvogel, L.; Amigorena, S. *Nat. Rev. Immunol.* **2002**, *2* (8), 569–579.

(92) Bobrie, A.; Théry, C. *Biochem. Soc. Trans.* **2013**, *41* (1), 263–267.

(93) Bobrie, A.; Colombo, M.; Krumeich, S.; Raposo, G.; Théry, C. *J. Extracell. Vesicles* **2012**, *1*, 18397.

(94) Lötvall, J.; Hill, A. F.; Hochberg, F.; Buzás, E. I.; Di Vizio, D.; Gardiner, C.; Song Gho, Y.; Kurochkin, I. V.; Mathivanan, S.; Quesenberry, P.; Sahoo, S.; Tahara, H.; Wauben, M. H.; Witwer, K. W.; Théry, C. *J. Extracell. Vesicles* **2014**, *3*, 26913.

(95) Gould, S. J.; Raposo, G. *J. Extracell. Vesicles* **2013**, *2*, 20389.

(96) Zitvogel, L.; Regnault, A.; Lozier, A.; Wolfers, J.; Flament, C.; Tenza, D.; Ricciardi-Castagnoli, P.; Raposo, G.; Amigorena, S. *Nat. Med.* **1998**, *4* (5),

594–600.

(97) Filipazzi, P.; Bürdek, M.; Villa, A.; Rivoltini, L.; Huber, V. *Semin. Cancer Biol.* **2012**, *22* (4), 342–349.

(98) Orgy, B. G.; Hung, M. E.; Breakefield, X. O.; Leonard, J. N. *Annu. Rev. Pharmacol. Toxicol* **2015**, *55*, 439–464.

(99) Burke, M. C.; Oei, M. S.; Edwards, N. J.; Ostrand-Rosenberg, S.; Fenselau, C. *J. Proteome Res.* **2014**, *13* (12), 5965–5972.

(100) Geis-Asteggiante, L.; Dhabaria, A.; Edwards, N.; Ostrand-Rosenberg, S.; Fenselau, C. *Int. J. Mass Spectrom.* **2015**, *378*, 264–269.

(101) Tran, J. C.; Doucette, A. A. *Anal. Chem.* **2008**, *80* (5), 1568–1573.

(102) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 683–693.

(103) Protein Discovery. *GELFrEE 8100 fractionation System Data Sheet*; 2009.

(104) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480* (7376), 254–258.

(105) Kellie, J. F.; Catherman, A. D.; Durbin, K. R.; Tran, J. C.; Tipton, J. D.; Norris, J. L.; Witkowski, C. E.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2012**, *84* (1), 209–215.

(106) Ahlf, D. R.; Compton, P. D.; Tran, J. C.; Early, B. P.; Thomas, P. M.; Kelleher, N. L. *J. Proteome Res.* **2012**, *11* (8), 4308–4314.

(107) Catherman, A. D.; Li, M.; Tran, J. C.; Durbin, K. R.; Compton, P. D.; Early, B.

P.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2013**, *85* (3), 1880–1888.

(108) Wessel, D.; Flügge, U. I. *Anal. Biochem.* **1984**, *138* (1), 141–143.

(109) Desrosiers, R.; Tanguay, R. M. *J. Biol. Chem.* **1988**, *263* (10), 4686–4692.

(110) Allis, C. D.; Allen, R. L.; Wiggins, J. C.; Chicoine, L. G.; Richman, R. *J. Cell Biol.* **1984**, *99* (5), 1669–1677.

(111) Duncan, E. M.; Muratore-Schroeder, T. L.; Cook, R. G.; Garcia, B. A.; Shabanowitz, J.; Hunt, D. F.; Allis, C. D. *Cell* **2008**, *135* (2), 284–294.

(112) Vossaert, L.; Meert, P.; Scheerlinck, E.; Glibert, P.; Van Roy, N.; Heindryckx, B.; De Sutter, P.; Dhaenens, M.; Deforce, D. *Stem Cell Res.* **2014**, *13* (1), 123–134.

(113) Duarte, L. F.; Young, A. R. J.; Wang, Z.; Wu, H.-A.; Panda, T.; Kou, Y.; Kapoor, A.; Hasson, D.; Mills, N. R.; Maayan, A.; Narita, M.; Bernstein, E. *Nat. Commun.* **2014**, *5*, 5210.

(114) Santos-Rosa, H.; Kirmizis, A.; Nelson, C.; Bartke, T.; Saksouk, N.; Cote, J.; Kouzarides, T. *Nat. Struct. Mol. Biol.* **2009**, *16* (1), 17–22.

(115) Tvardovskiy, A.; Wrzesinski, K.; Sidoli, S.; Fey, S. J.; Rogowska-Wrzesinska, A.; Jensen, O. N. *Mol. Cell. Proteomics* **2015**, *14* (12), 3142–3153.

(116) Duncan, E. M.; Allis, C. D. In *Epigenetics and Disease*; Gasser, S. M., Li, E., Eds.; Springer Basel: Basel, 2011; pp 69–90.

(117) Raftery, M. J.; Harrison, C. A.; Alewood, P.; Jones, A.; Geczy, C. L. *Biochem. J.* **1996**, *316*, 285–293.

(118) Geis-Asteggiante, L.; Ostrand-Rosenberg, S.; Fenselau, C.; Edwards, N. J. *Anal. Chem.* **2016**, In Press (DOI:10.1021/acs.analchem.6b02151).

164

(119) Smith, L. M.; Kelleher, N. L. *Nat. Methods* **2013**, *10* (3), 186–187.

(120) Toby, T. K.; Fornelli, L.; Kelleher, N. L. *Annu. Rev. Anal. Chem.* **2016**, *9* (1), 499–519.

(121) Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L. *Anal. Chem.* **2006**, *78* (3), 686–694.

(122) Waanders, L. F.; Hanke, S.; Mann, M. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (11), 2058–2064.

(123) Collier, T. S.; Sarkar, P.; Rao, B.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (6), 879–889.

(124) Mann, M. *Nat Rev Mol Cell Biol* **2006**, *7* (12), 952–958.

(125) Hung, C. W.; Tholey, A. *Anal. Chem.* **2012**, *84* (1), 161–170.

(126) Mazur, M. T.; Cardasis, H. L.; Spellman, D. S.; Liaw, A.; Yates, N. A.; Hendrickson, R. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (17), 7728–7733.

(127) Chen, Y.; Hoover, M. E.; Dang, X.; Shomo, A. A.; Guan, X.; Marshall, A. G.; Freitas, M. A.; Young, N. L. *Mol. Cell. Proteomics* **2016**, *15* (3), 818–833.

(128) Wu, S.; Brown, J. N.; Tolić, N.; Meng, D.; Liu, X.; Zhang, H.; Zhao, R.; Moore, R. J.; Pevzner, P.; Smith, R. D.; Paša-Tolić, L. *Proteomics* **2014**, *14*, 1211–1222.

(129) Ntai, I.; LeDuc, R. D.; Fellers, R. T.; Erdmann-Gilmore, P.; Davies, S. R.; Rumsey, J.; Early, B. P.; Thomas, P. M.; Li, S.; Compton, P. D.; Ellis, M. J. C.; Ruggles, K. V; Fenyö, D.; Boja, E. S.; Rodriguez, H.; Townsend, R. R.; Kelleher, N. L. *Mol. Cell. Proteomics* **2016**, *15* (1), 45–56.

(130) Durbin, K. R.; Fornelli, L.; Fellers, R. T.; Doubleday, P. F.; Narita, M.;

Kelleher, N. L. *J. Proteome Res.* **2016**, *15* (3), 976–982.

(131) Zhang, B.; VerBerkmoes, N. C.; Langston, M. A.; Uberbacher, E.; Hettich, R. L.; Samatova, N. F. *J. Proteome Res.* **2006**, *5* (11), 2909–2918.

(132) Jiang, H.; Salzman, J. *Biometrika* **2012**, *99* (4), 973–980.

(133) Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. *Anal. Chem.* **2009**, *81* (15), 6317–6326.

(134) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2011**, *83* (17), 6868–6874.

(135) Quackenbush, J. *Nat. Genet.* **2002**, *32* (Supp), 496–501.

(136) Karpievitch, Y. V; Dabney, A. R.; Smith, R. D. *BMC Bioinformatics* **2012**, *13* (Suppl 16), S5.

(137) Listgarten, J.; Emili, A. *Mol. Cell. Proteomics* **2005**, *4*, 419–434.

(138) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W. J.; Webb-Robertson, B. J. M.; Smith, R. D.; Lipton, M. S. *J. Proteome Res.* **2006**, *5* (2), 277–286.

(139) Raftery, M. J.; Yang, Z.; Valenzuela, S. M.; Geczy, C. L. *J. Biol. Chem.* **2001**, *276* (36), 33393–33401.

(140) Lim, S. Y.; Raftery, M. J.; Goyette, J.; Hsu, K.; Geczy, C. L. *J. Leukoc. Biol.* **2009**, *86* (3), 577–587.

(141) Lim, S. Y.; Raftery, M. J.; Geczy, C. L. *Antioxid. Redox Signal.* **2011**, *15* (8), 2235–2248.

(142) Raftery, M. J.; Harrison, C. a; Alewood, P.; Jones, A.; Geczy, C. L. *Biochem. J.* **1996**, *316*, 285–293.

(143) Bartel, D. P. *Cell* **2009**, *136* (2), 215–233.

(144) He, L.; Hannon, G. J. *Nat. Rev. Genet.* **2004**, *5* (7), 522–531.

(145) Montecalvo, A.; Larregina, A. T.; Shufesky, W. J.; Stolz, D. B.; Sullivan, M. L. G.; Karlsson, J. M.; Baty, C. J.; Gibson, G. A.; Erdos, G.; Wang, Z.; Milosevic, J.; Tkacheva, O. A.; Divito, S. J.; Jordan, R.; Lyons-Weiler, J.; Watkins, S. C.; Morelli, A. E. *Blood* **2012**, *119* (3), 756–766.

(146) Valadi, H.; Ekström, K.; Bossios, A.; Sjöstrand, M.; Lee, J. J.; Lötvall, J. O. *Nat. Cell Biol.* **2007**, *9* (6), 654–659.

(147) Ekström, K.; Valadi, H.; Sjöstrand, M.; Malmhäll, C.; Bossios, A.; Eldh, M.; Lötvall, J. *J. Extracell. Vesicles* **2012**, *1*, 18389.

(148) Kogure, T.; Lin, W.-L.; Yan, I. K.; Braconi, C.; Patel, T. *Hepatology* **2011**, *54* (4), 1237–1248.

(149) Pegtel, D. M.; Cosmopoulos, K.; Thorley-Lawson, D. A.; van Eijndhoven, M. A. J.; Hopmans, E. S.; Lindenberg, J. L.; de Gruijl, T. D.; Wurdinger, T.; Middeldorp, J. M. *Proc. Natl. Acad. Sci.* **2010**, *107* (14), 6328–6333.

(150) Chiba, M.; Kimura, M.; Asari, S. *Oncol. Rep.* **2012**, *28*, 1551–1558.

(151) Zhang, J.; Li, S.; Li, L.; Li, M.; Guo, C.; Yao, J.; Mi, S. *Genomics. Proteomics Bioinformatics* **2015**, *13* (1), 17–24.

(152) Janas, T.; Janas, M. M.; Sapoń, K.; Janas, T. *FEBS Lett.* **2015**, *589* (13), 1391–1398.

(153) Li, M.; Zeringer, E.; Barta, T.; Schageman, J.; Cheng, A.; Vlassov, A. V. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369*, 20130502.

(154) Fabbri, M.; Paone, A.; Calore, F.; Galli, R.; Gaudio, E.; Santhanam, R.; Lovat,

F.; Fadda, P.; Mao, C.; Nuovo, G. J.; Zanesi, N.; Crawford, M.; Ozer, G. H.; Wernicke, D.; Alder, H.; Caligiuri, M. A.; Nana-Sinkam, P.; Perrotti, D.; Croce, C. M. *Proc. Natl. Acad. Sci.* **2012**, *109* (31), E2110–E2116.

(155) Fabbri, M.; Paone, A.; Calore, F.; Galli, R.; Croce, C. M. *RNA Biol.* **2013**, *10* (2), 169–174.

(156) Skog, J.; Würdinger, T.; van Rijn, S.; Meijer, D. H.; Gainche, L.; Curry, W. T.; Carter, B. S.; Krichevsky, A. M.; Breakefield, X. O. *Nat. Cell Biol.* **2008**, *10* (12), 1470–1476.

(157) Ridder, K.; Sevko, A.; Heide, J.; Dams, M.; Rupp, A.-K.; Macas, J.; Starmann, J.; Tjwa, M.; Plate, K. H.; Sültmann, H.; Altevogt, P.; Umansky, V.; Momma, S. *Oncoimmunology* **2015**, *4* (6), e1008371.

(158) Tomasoni, S.; Longaretti, L.; Rota, C.; Morigi, M.; Conti, S.; Gotti, E.; Capelli, C.; Introna, M.; Remuzzi, G.; Benigni, A. *Stem Cells Dev.* **2013**, *22* (5), 772–780.

(159) Andrews, S. 2010, p FastQC: A Quality Control tool for High Throughput. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

(160) Bolger, A. M.; Lohse, M.; Usadel, B. *Bioinformatics* **2014**, *30* (15), 2114–2120.

(161) Martin, M. *EMBnet.journal* **2011**, *17* (1), 10–12.

(162) Langmead, B. *Curr. Protoc. Bioinforma.* **2010**, *32* (Unit 11.7), 11.7.1-11.7.14.

(163) Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D. R.; Pimentel, H.; Salzberg, S. L.; Rinn, J. L.; Pachter, L. *Nat. Protoc.* **2012**, *7* (3), 562–578.

(164) Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L. *Nat. Biotechnol.* **2016**, *34* (5), 525–527.

(165) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. *Bioinformatics* **2009**, *25* (16), 2078–2079.

(166) Anders, S.; Pyl, P. T.; Huber, W. *Bioinformatics* **2015**, *31* (2), 166–169.

(167) Kozomara, A.; Griffiths-Jones, S. *Nucleic Acids Res.* **2014**, *42*, D68–D73.

(168) Griffiths-Jones, S. *Nucleic Acids Res.* **2004**, *32* (90001), D109–D111.

(169) Dillies, M.-A.; Rau, A.; Aubert, J.; Hennequet-Antier, C.; Jeanmougin, M.; Servant, N.; Keime, C.; Marot, G.; Castel, D.; Estelle, J.; Guernec, G.; Jagla, B.; Jouneau, L.; Laloe, D.; Le Gall, C.; Schaeffer, B.; Le Crom, S.; Guedj, M.; Jaffrezic, F. *Brief. Bioinform.* **2013**, *14* (6), 671–683.

(170) Robinson, M. D.; Oshlack, A. *Genome Biol.* **2010**, *11*, R25.

(171) Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. *Bioinformatics* **2009**, *26* (1), 139–140.

(172) Bullard, J. H.; Purdom, E.; Hansen, K. D.; Dudoit, S. *BMC Bioinformatics* **2010**, *11*, 94.

(173) Anders, S.; Huber, W. *Genome Biol.* **2010**, *11* (10), R106.

(174) Leek, J. T.; Storey, J. D. *PLoS Genet.* **2007**, *3* (9), e161.

(175) Risso, D.; Ngai, J.; Speed, T. P.; Dudoit, S. *Nat. Biotechnol.* **2014**, *32* (9), 896–902.

(176) Ritchie, M. E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C. W.; Shi, W.; Smyth, G. K. *Nucleic Acids Res.* **2015**, *43* (7), e47.

(177) Dweep, H.; Gretz, N. *Nat. Methods* **2015**, *12* (8), 697–697.

(178) Dweep, H.; Sticht, C.; Pandey, P.; Gretz, N. *J. Biomed. Inform.* **2011**, *44* (5), 839–847.

(179) Paraskevopoulou, M. D.; Georgakilas, G.; Kostoulas, N.; Vlachos, I. S.; Vergoulis, T.; Reczko, M.; Filippidis, C.; Dalamagas, T.; Hatzigeorgiou, A. G. *Nucleic Acids Res.* **2013**, *41*, W169–W173.

(180) Krek, A.; Grün, D.; Poy, M. N.; Wolf, R.; Rosenberg, L.; Epstein, E. J.; MacMenamin, P.; da Piedade, I.; Gunsalus, K. C.; Stoffel, M.; Rajewsky, N. *Nat. Genet.* **2005**, *37* (5), 495–500.

(181) Enright, A. J.; John, B.; Gaul, U.; Tuschl, T.; Sander, C.; Marks, D. S. *Genome Biol.* **2003**, *5* (1), R1.

(182) Agarwal, V.; Bell, G. W.; Nam, J.-W.; Bartel, D. P. *Elife* **2015**, *4*, e05005.

(183) Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E. *Nat. Genet.* **2007**, *39* (10), 1278–1284.

(184) Kruger, J.; Rehmsmeier, M. *Nucleic Acids Res.* **2006**, *34*, 451–454.

(185) Wong, N.; Wang, X. *Nucleic Acids Res.* **2015**, *43* (D1), D146–D152.

(186) Tsang, J. S.; Ebert, M. S.; van Oudenaarden, A. *Mol. Cell* **2010**, *38* (1), 140–153.

(187) Hsu, S.-D.; Chu, C.-H.; Tsou, A.-P.; Chen, S.-J.; Chen, H.-C.; Hsu, P. W.-C.; Wong, Y.-H.; Chen, Y.-H.; Chen, G.-H.; Huang, H.-D. *Nucleic Acids Res.* **2007**, *36*, D165–D169.

(188) Miranda, K. C.; Huynh, T.; Tay, Y.; Ang, Y.-S.; Tam, W.-L.; Thomson, A. M.; Lim, B.; Rigoutsos, I. *Cell* **2006**, *126* (6), 1203–1217.

(189) Blin, K.; Dieterich, C.; Wurmus, R.; Rajewsky, N.; Landthaler, M.; Akalin, A.

Nucleic Acids Res. **2015**, *43*, D160–D167.

(190) Hsu, S.-D.; Tseng, Y.-T.; Shrestha, S.; Lin, Y.-L.; Khaleel, A.; Chou, C.-H.; Chu, C.-F.; Huang, H.-Y.; Lin, C.-M.; Ho, S.-Y.; Jian, T.-Y.; Lin, F.-M.; Chang, T.-H.; Weng, S.-L.; Liao, K.-W.; Liao, I.-E.; Liu, C.-C.; Huang, H.-D. *Nucleic Acids Res.* **2014**, *42* (D1), D78–D85.

(191) Reimand, J.; Arak, T.; Adler, P.; Kolberg, L.; Reisberg, S.; Peterson, H.; Vilo, J. *Nucleic Acids Res.* **2016**, *44* (W1), W83–W89.

(192) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nucleic Acids Res.* **2009**, *37* (1), 1–13.

(193) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nat. Protoc.* **2008**, *4* (1), 44–57.

(194) Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. *Nucleic Acids Res.* **2016**, *44*, D457–D462.

(195) Fabregat, A.; Sidiropoulos, K.; Garapati, P.; Gillespie, M.; Hausmann, K.; Haw, R.; Jassal, B.; Jupe, S.; Korninger, F.; McKay, S.; Matthews, L.; May, B.; Milacic, M.; Rothfels, K.; Shamovsky, V.; Webber, M.; Weiser, J.; Williams, M.; Wu, G.; Stein, L.; Hermjakob, H.; D'Eustachio, P. *Nucleic Acids Res.* **2016**, *44* (D1), D481–D487.

(196) Xiao, D.; Ohlendorf, J.; Chen, Y.; Taylor, D. D.; Rai, S. N.; Waigel, S.; Zacharias, W.; Hao, H.; McMasters, K. M. *PLoS One* **2012**, *7* (10), e46874.

(197) Ji, H.; Chen, M.; Greening, D. W.; He, W.; Rai, A.; Zhang, W.; Simpson, R. J. *PLoS One* **2014**, *9* (10), e110314.

(198) Crescitelli, R.; Lässer, C.; Szabó, T. G.; Kittel, A.; Eldh, M.; Dianzani, I.;

Buzás, E. I.; Lötvall, J. *J. Extracell. Vesicles* **2013**, *2* (0).

(199) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 4th Ed.; Garland Science: New York, 2002.

(200) Zhang, K.; Corsa, C. A.; Ponik, S. M.; Prior, J. L.; Piwnica-Worms, D.; Eliceiri, K. W.; Keely, P. J.; Longmore, G. D. *Nat. Cell Biol.* **2013**, *15* (6), 677–687.

(201) Toy, K. A.; Valiathan, R. R.; Núñez, F.; Kidwell, K. M.; Gonzalez, M. E.; Fridman, R.; Kleer, C. G. *Breast Cancer Res. Treat.* **2015**, *150* (1), 9–18.

(202) Ren, T.; Zhang, W.; Liu, X.; Zhao, H.; Zhang, J.; Zhang, J.; Li, X.; Zhang, Y.; Bu, X.; Shi, M.; Yao, L.; Su, J. *J. Pathol.* **2014**, *234* (4), 526–537.

(203) Hanahan, D.; Weinberg, R. A. *Cell* **2011**, *144*, 646–674.

(204) McMahon, G. *Oncologist* **2000**, *5 Suppl 1*, 3–10.

(205) Yarden, Y.; Sliwkowski, M. X. *Nat. Rev. Mol. Cell Biol.* **2001**, *2* (2), 127–137.

(206) Komander, D.; Rape, M. *Annu. Rev. Biochem* **2012**, *81*, 203–229.

(207) Sengupta, N.; Seto, E. *J. Cell. Biochem.* **2004**, *93* (1), 57–67.

(208) Marcelo, K. L.; Goldie, L. C.; Hirschi, K. K. *Circ. Res.* **2013**, *112* (9), 1272–1287.

(209) Ridder, K.; Keller, S.; Dams, M.; Rupp, A.-K.; Schlaudraff, J.; Del Turco, D.; Starmann, J.; Macas, J.; Karpova, D.; Devraj, K.; Depboylu, C.; Landfried, B.; Arnold, B.; Plate, K. H.; Höglinger, G.; Sültmann, H.; Altevogt, P.; Momma, S. *PLoS Biol.* **2014**, *12* (6), e1001874.

(210) Carroll, A. P.; Goodall, G. J.; Liu, B. *Wiley Interdiscip. Rev. RNA* **2014**, *5* (3), 361–379.

(211) Brennecke, J.; Stark, A.; Russell, R. B.; Cohen, S. M. *PLoS Biol.* **2005**, *3* (3), e85.

(212) Liu, B.; Li, J.; Cairns, M. J. *Brief. Bioinform.* **2014**, *15* (1), 1–19.

(213) Taganov, K. D.; Boldin, M. P.; Chang, K.-J.; Baltimore, D. *Proc. Natl. Acad. Sci.* **2006**, *103* (33), 12481–12486.

(214) Zhao, J. L.; Rao, D. S.; Boldin, M. P.; Taganov, K. D.; O'Connell, R. M.; Baltimore, D. *Proc. Natl. Acad. Sci.* **2011**, *108* (22), 9184–9189.

(215) Hegde, V. L.; Tomar, S.; Jackson, A.; Rao, R.; Yang, X.; Singh, U. P.; Singh, N. P.; Nagarkatti, P. S.; Nagarkatti, M. *J. Biol. Chem.* **2013**, *288* (52), 36810–36826.

(216) Wang, S.; Aurora, A. B.; Johnson, B. A.; Qi, X.; McAnally, J.; Hill, J. A.; Richardson, J. A.; Bassel-Duby, R.; Olson, E. N. *Dev. Cell* **2008**, *15* (2), 261–271.

(217) Boldin, M. P.; Taganov, K. D.; Rao, D. S.; Yang, L.; Zhao, J. L.; Kalwani, M.; Garcia-Flores, Y.; Luong, M.; Devrekanli, A.; Xu, J.; Sun, G.; Tay, J.; Linsley, P. S.; Baltimore, D. *J. Exp. Med.* **2011**, *208* (6), 1189–1201.

(218) Perry, M. M.; Moschos, S. A.; Williams, A. E.; Shepherd, N. J.; Larner-Svensson, H. M.; Lindsay, M. A. *J. Immunol.* **2008**, *180* (8), 5689–5698.

(219) Tang, B.; Xiao, B.; Liu, Z.; Li, N.; Zhu, E.-D.; Li, B.-S.; Xie, Q.-H.; Zhuang, Y.; Zou, Q.-M.; Mao, X.-H. *FEBS Lett.* **2010**, *584* (8), 1481–1486.

(220) Li, L.; Zhang, J.; Diao, W.; Wang, D.; Wei, Y.; Zhang, C.-Y.; Zen, K. *J. Immunol.* **2014**, *192* (3), 1034–1043.

(221) Liao, J.; Liu, R.; Yin, L.; Pu, Y. *Int. J. Mol. Sci.* **2014**, *15* (9), 15530–15551.

(222) Guo, L.; Luo, C.; Fan, J.; Hou, Z.; Ji, X.; Chen, F.; Zhu, B.; Ni, C. *Ind. Health* **2015**, *53* (1), 38–47.

(223) O'Neill, L. A.; Sheedy, F. J.; McCoy, C. E. *Nat. Rev. Immunol.* **2011**, *11* (3), 163–175.

(224) O'Connell, R. M.; Rao, D. S.; Chaudhuri, A. A.; Baltimore, D. *Nat. Rev. Immunol.* **2010**, *10* (2), 111–122.

(225) Zhang, M.; Liu, Q.; Mi, S.; Liang, X.; Zhang, Z.; Su, X.; Liu, J.; Chen, Y.; Wang, M.; Zhang, Y.; Guo, F.; Zhang, Z.; Yang, R. *J. Immunol.* **2011**, *186* (8), 4716–4724.

(226) Chen, R.; Alvero, A. B.; Silasi, D. A.; Kelly, M. G.; Fest, S.; Visintin, I.; Leiser, A.; Schwartz, P. E.; Rutherford, T.; Mor, G. *Oncogene* **2008**, *27* (34), 4712–4723.

(227) Melnik, B. C. *J. Transl. Med.* **2015**, *13* (1), 202.

(228) McClure, C.; Brudecki, L.; Ferguson, D. A.; Yao, Z. Q.; Moorman, J. P.; McCall, C. E.; Gazzar, M. El. *Infect. Immun.* **2014**, *82* (9), 3816–3825.

(229) Meng, F.; Henson, R.; Wehbe–Janek, H.; Ghoshal, K.; Jacob, S. T.; Patel, T. *Gastroenterology* **2007**, *133* (2), 647–658.

(230) Sheedy, F. J.; Palsson-McDermott, E.; Hennessy, E. J.; Martin, C.; O'Leary, J. J.; Ruan, Q.; Johnson, D. S.; Chen, Y.; O'Neill, L. A. J. *Nat. Immunol.* **2010**, *11* (2), 141–147.

(231) Liu, Q.; Zhang, M.; Jiang, X.; Zhang, Z.; Dai, L.; Min, S.; Wu, X.; He, Q.; Liu, J.; Zhang, Y.; Zhang, Z.; Yang, R. *Int. J. Cancer* **2011**, *129* (11), 2662–2673.

(232) Fazi, F.; Rosa, A.; Fatica, A.; Gelmetti, V.; De Marchis, M. L.; Nervi, C.;

Bozzoni, I. *Cell* **2005**, *123* (5), 819–831.

(233) Liu, Y.; Lai, L.; Chen, Q.; Song, Y.; Xu, S.; Ma, F.; Wang, X.; Wang, J.; Yu, H.; Cao, X.; Wang, Q. *J. Immunol.* **2012**, *188* (11), 5500–5510.

(234) Tian, J.; Rui, K.; Tang, X.; Ma, J.; Wang, Y.; Tian, X.; Zhang, Y.; Xu, H.; Lu, L.; Wang, S. *J. Immunol.* **2015**, *195* (3), 1301–1311.

(235) Wang, S.; Tang, Y.; Cui, H.; Zhao, X.; Luo, X.; Pan, W.; Huang, X.; Shen, N. *Genes Immun.* **2011**, *12* (2), 149–154.

(236) Li, H.-W.; Meng, Y.; Xie, Q.; Yi, W.-J.; Lai, X.-L.; Bian, Q.; Wang, J.; Wang, J.-F.; Yu, G. *Biochem. Biophys. Res. Commun.* **2015**, *467* (3), 595–601.

(237) Beury, D. W.; Parker, K. H.; Nyandjo, M.; Sinha, P.; Carter, K. a.; Ostrand-Rosenberg, S. *J. Leukoc. Biol.* **2014**, *96* (6), 1109–1118.

(238) Kroesen, B.-J.; Teteloshvili, N.; Smigielska-Czepiel, K.; Brouwer, E.; Boots, A. M. H.; van den Berg, A.; Kluiver, J. *Immunology* **2015**, *144* (1), 1–10.

(239) Lu, L.-F.; Thai, T.-H.; Calado, D. P.; Chaudhry, A.; Kubo, M.; Tanaka, K.; Loeb, G. B.; Lee, H.; Yoshimura, A.; Rajewsky, K.; Rudensky, A. Y. *Immunity* **2009**, *30* (1), 80–91.

(240) Chauhan, S.; Danielson, S.; Clements, V.; Edwards, N.; Ostrand-Rosenberg, S.; Fenselau, C. *J. Proteome Res.* **2016**, *In Press*, DOI: 10.1021/acs.jproteome.6b00811.

(241) Villarroya-Beltri, C.; Gutiérrez-Vázquez, C.; Sánchez-Cabo, F.; Pérez-Hernández, D.; Vázquez, J.; Martin-Cofreces, N.; Martinez-Herrera, D. J.; Pascual-Montano, A.; Mittelbrunn, M.; Sánchez-Madrid, F. *Nat. Commun.* **2013**, *4*, 581–593.

(242) Livneh, I.; Cohen-Kaplan, V.; Cohen-Rosenzweig, C.; Avni, N.; Ciechanover, A. *Cell Res.* **2016**, *26* (8), 869–885.

(243) De La Mota-Peynado, A.; Lee, S. Y.-C.; Pierce, B. M.; Wani, P.; Singh, C. R.; Roelofs, J. *J. Biol. Chem.* **2013**, *288* (41), 29467–29481.

(244) Kloetzel, P. M. *Nat. Immunol.* **2004**, *5* (7), 661–669.

(245) McCarthy, M. K.; Weinberg, J. B. *Front. Microbiol.* **2015**, *6*, 1–16.

(246) Adams, K. Investigation by mass spectrometry of the ubiquitome and protein cargo of exosomes derived from myeloid-derived suppressor cells. Thesis submitted for the M.S. degree, University of Maryland, 2016.

(247) Lai, R. C.; Tan, S. S.; Teh, B. J.; Sze, S. K.; Arslan, F.; de Kleijn, D. P.; Choo, A.; Lim, S. K. *Int. J. Proteomics* **2012**, *2012*, 1–14.

(248) Bartel, D. P. *Cell* **2004**, *116* (2), 281–297.

(249) Mihelič, M.; Teuscher, C.; Turk, V.; Turk, D. *FEBS Lett.* **2006**, *580* (17), 4195–4199.

(250) Wang, X.; Wang, E.; Kavanagh, J. J.; Freedman, R. S. *J. Transl. Med.* **2005**, *3* (1), 25.

(251) Lima, L. G.; Monteiro, R. Q. *Biosci. Rep.* **2013**, *33* (5), 701–710.

(252) Gorp, H. Van; Delputte, P. L.; Nauwynck, H. J. *Mol. Immunol.* **2010**, *47*, 1650–1660.

(253) Wang, T.; Chu, Z.; Lin, H.; Jiang, J.; Zhou, X.; Liang, X. *Mol. Biol. Rep.* **2014**, *41* (6), 4069–4076.

(254) Sano, H.; Hsu, D. K.; Yu, L.; Apgar, J. R.; Kuwabara, I.; Yamanaka, T.; Hirashima, M.; Liu, F.-T. *J. Immunol.* **2000**, *165* (4), 2156–2164.

(255) Kanno, H.; Nishihara, H.; Wang, L.; Yuzawa, S.; Kobayashi, H.; Tsuda, M.; Kimura, T.; Tanino, M.; Terasaka, S.; Tanaka, S. *Neuro. Oncol.* **2013**, *15* (7), 853–864.

(256) Gao, J.; Wu, Y.; Su, Z.; Amoah Barnie, P.; Jiao, Z.; Bie, Q.; Lu, L.; Wang, S.; Xu, H. *PLoS One* **2014**, *9* (8), e104453.

(257) Simpson, K. D.; Cross, J. V. *Oncoimmunology* **2013**, *2* (3), e23337.

(258) Simpson, K. D.; Templeton, D. J.; Cross, J. V. *J. Immunol.* **2012**, *189* (12), 5533–5540.

(259) Deuel, T. F.; Senior, R. M.; Chang, D.; Griffin, G. L.; Heinrikson, R. L.; Kaiser, E. T. *Med. Sci.* **1981**, *78* (7), 4584–4587.

(260) Nagaraj, S.; Gabrilovich, D. I. *Semin. Cancer Biol.* **2012**, *22* (4), 282–288.

(261) Parker, K. H.; Sinha, P.; Horn, L. A.; Clements, V. K.; Yang, H.; Li, J.; Tracey, K. J.; Ostrand-Rosenberg, S. *Cancer Res.* **2014**, *74* (20), 5723–5733.

(262) Ostrand-Rosenberg, S.; Sinha, P. *J. Immunol.* **2009**, *182* (8), 4499–4506.

(263) Melani, C.; Sangaletti, S.; Barazzetta, F. M.; Werb, Z.; Colombo, M. P. *Cancer Res.* **2007**, *67* (23), 11438–11446.

(264) Xiang, X.; Poliakov, A.; Liu, C.; Liu, Y.; Deng, Z.; Wang, J.; Cheng, Z.; Shah, S. V.; Wang, G.-J.; Zhang, L.; Grizzle, W. E.; Mobley, J.; Zhang, H.-G. *Int. J. Cancer* **2009**, *124* (11), 2621–2633.