Label-free and spike-in standard-free mass spectrometry in the proteomic analysis of plasma membrane proteins and membrane-associated protein networks

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften

der Technischen Universität Dresden

von

Diplom-Biochemiker, Diplom-Ingenieur (FH)

Christian Niehage

geboren am 14.2.1967 in Berlin

ACKNOWLEDGMENT

CONTENTS

SUMMARY

Mass spectrometry is the primary technology of proteomics. For the analysis of complex proteomes, protein identities and quantities are inferred from their peptides that are generated by cleaving all proteins with the endopeptidase trypsin. But there is one major disadvantage that is due to biophysical differences, different peptides cause different intensities. Miscellaneous approaches have been developed to circumvent this problem based on the chemical or metabolic introduction of heavy stable isotopes. This enables to monitor protein abundance differences of two or more samples on the same tryptic peptides that differ in mass only. Absolute quantification can be achieved similar by spiking-in synthetic isotopical labeled counterparts of a sample's tryptic peptides. However, labeling technics suffer from high prices, introduced biases, need for extensive manual control, laborious implementation and implementation restrictions. Therefore, a multiplicity of label-free approaches have been developed that profit from instrumental improvements targeting reliability of identifications and reproducibility of quantitative values. No extensive systematic comparison of label-free quantitative parameters has been published so far presumably because of the laborious implementation. An analysis of primary label-free parameters and associated normalization methods is presented here that compares dynamic and linear ranges and accuracies in the estimation of protein amounts. This facilitated the establishment of label-free procedures addressing three fundamental questions in proteomics: what is a sample's composition, are proteins that share a specific property enriched and what are the differences between two (or more) samples. A new mathematic model is presented that defines and elucidates enrichment.

The procedures were applied first to analyze and compare stem cell plasma membrane proteomes. This is an ambitious model for proteomics because of only small amounts of arduous to analyze, partial hydrophobic proteins in a complex proteomic and chemical background. It is of scientific relevance, as membrane proteins are the cell's communication interface that enable cell type specific processes and hence can be used to define, isolate and quantify those. The success of cell surface proteome enrichment, the quantitative composition of the proteome and the proteomic difference between stem cells isolated from the dental pulp and cultivated in different media is shown.

Secondly, the procedures were applied to the analysis of transient protein networks that assemble onto proteo-liposomes in a newly designed recruitment assay that fully recapitulates membrane sorting as seen in vivo. All transmembrane proteins need to be trafficked to other organelles' membranes by vesicular trafficking. Sorting signals within the cytosolic regions of the protein cargos trigger the formation of trafficking complexes around those. The transient membrane complexes additionally recognize organelle or organelle-domain specific membrane lipids, such as phosphatidylinositol phosphates. Different trafficking ways are characterized by different trafficking complexes. The elucidation of trafficking complexes that form around a transmembrane protein of interest discloses its trafficking routes and involved signaling processes. The synthetic proteo-liposomes were

prepared from chemically defined lipids and heterologous expressed cytosolic domains of type-I or type-II membrane receptors. The proteomic analyses of such samples are challenging because of huge proteomic backgrounds of proteins binding to the liposomes irrespective of the receptor and relatively small amounts and numbers of receptor-specific binders. Though the basic idea is to elucidate sorting machineries and study membrane trafficking processes, such experiments are untargeted and miscellaneous discoveries were achieved. We elucidated that the apical determinant crumbs 2 is a cargo of the retromer complex. This revealed a fameless level of control for the establishment of cell polarity. We found retromer along with the adapter complexes AP-4 and AP-5 trafficking the beta amyloid precursor protein APP. This confirmed recent publications and yielded new insights. Moreover, many more proteins and complexes appeared to associate with the cytosolic part of APP (AICD) in a membrane context-dependent or -independent manner. Among those, some were so far unknown to interact with AICD, like mTORC1 and the PIKFyve complex.

1. INTRODUCTION

1.1 Prologue

In the last twenty years, the biological sciences substantially changed with the introduction of -omics techniques. With the Human Genome Project and the publication of the first drafts of the human genome in 2001 (Lander, 2001; Venter, 2001), this process reached a milestone and a cornerstone for the system-wide analysis of the proteome. The basic technology of proteomics is mass spectrometry. A key step in the development of instruments for the analysis of proteins was the introduction of the soft-ionization techniques, matrix-assisted laser desorption ionization (MALDI) by Karas and Hillenkamp (Karas, 1988) as well as Tanaka (Tanaka, 1988), and electrospray ionization (ESI) by Fenn et al. (Fenn, 1989). The latter is the dominant ionization technique today in proteomics as it allows coupling a high-performance liquid chromatography system in front, for high content analysis (LC-MS). Additionally, the application of mass spectrometry for the study of the proteome reached several breakthroughs. One such was the use of trypsin to specifically cleave proteins in a subset of 'tryptic peptides' which masses were estimated to identify the protein they were generated from (Cottrell, 1994). This so-called peptide mass fingerprint relied on data bases with protein sequence data that were either experimentally yielded before or deduced from genome data. Nowadays, tryptic peptides are fragmented in the mass spectrometer to assess their partial or full sequence (LC-MS/MS). This increases the reliability of matching tryptic peptides to particular proteins and hence it allows studying a whole proteome in a single LC-MS run (Nagaraj, 2012). The analysis of such immense amounts of data benefits from the development of powerful and affordable computers. With the drastic increase in capacity and instrumental sensitivity, lists of identified proteins of samples and related controls became more and more interchangeable. This necessitated the development of quantitative approaches (Ong, 2005). Quantification is hindered by the fact that due to physicochemical distinctions different peptides cause different intensities. The introduction of heavy stable isotope labels circumvented this problem as it allows the quantitative study of proteins by the measurement of chemically identical peptides. Protein amounts can be estimated by peak comparison, either to a related sample or to a spike-in standard of synthetic tryptic peptide analogs. There was also a constant instrumental development that reduced the number of necessary protein fractionation steps and run to run deviations in LC-MS/MS, a prerequisite for a multitude of label-free methods that were developed in the last years. Label-free methods are attractive, because they omit long sample adaptation procedures, like the adaptation of a cell line to a specific medium to ensure high labeling efficiency and they omit implicated restrictions. Moreover they are cost-saving and hence reduce the entry barrier for work groups to make use of mass spectrometry services for their projects.

In modern proteomics, holistic approaches are used to study system perturbations and reductionistic approaches are performed for validation of specific findings. Mass spectrometry is also a key technology for the analysis of the lipidome (Blanksby, 2010) and

metabolome (Villas-Bôas, 2005). We expect this holism in near future to not only encompass the proteome, but also the lipidome and metabolome when studying a single sample – a concept called systems biology which seems now at hand with modern mass spectrometry instruments and approaches.

1.2 Mass spectrometry vocabulary

Only charged peptides are measured in mass spectrometry. If protonated peptides are measured, the instrument is operated in the positive mode which is the also the standard mode. The instrument can also be operated in the negative mode, for instance when multiphosphorylated peptides are analyzed in special approaches. The measurement doesn't directly yield masses, but mass to charge ratios instead. The unit is Thomson (Th) and is defined as one atomic unit per elementary charge: 1 Th = 1 u / e = $1.660540 \cdot 10^{-27}$ kg / $1.602177 \cdot 10^{-19}$ C = $1.036426 \cdot 10^{-8}$ kg C^{-1} . Peptides don't appear as a single mass to charge ratio in the spectrum, but as a number of peaks instead, due to the probabilistic distributed presence of isotopes. This is mainly due to the 1.1% of naturally occurring ^{13}C and 4.21% ^{34}S. Such patterns are called isotopic clusters. Because the distance of the isotopic peaks is always one atomic unit divided by the charge, the charge can be inferred from that. The mass value which is assigned to a peptide then, is the monoisotopic mass which is inferred from the most abundant isotopes of each element. For peptides, it is the mass of the first (most left) isotopic peak. This differs from stoichiometric calculations where a compound's mass (average mass) is calculated from the atomic weight, a weighted average of the atomic masses of the different isotopes of each element. For small peptides, below 1.5 kDa, the first isotopic peak is the most intensive one within the cluster. If peptides are bigger, it is more likely that at least one of the atoms within the peptide is a heavier isotope and the second isotopic peak is the most intensive one. For proteins, the monoisotopic peak cannot even be seen in the spectrum. The difference between the observed and a calculated monoisotopic mass is the mass accuracy which is expressed in parts per million (ppm). The ability of a mass analyzer to resolve peaks of ions with a small m/z difference is the resolution. It is commonly defined as m/Δm, with m=400 Th and Δm as the difference to the nearest (theoretical) peak at which the two peaks are separated by a valley with 10% of the smaller peaks height. As this is troublesome to estimate, the resolving power can also be approximated by half of the full width at half maximum (FWHM) of a single peak. Isotopic resolution is achieved when the peaks of an isotopic cluster are resolved. The mass range is the limit of m/z over which the mass analyzer can measure ions and the scan speed is the rate at which the analyzer measures over a particular mass range. Beside the estimation of masses, the mass spectrometer also delivers intensity values from the ion detection devices. Within this thesis this intensity values will be investigated how they can be used best to assess protein abundance. Another measurement parameter in LC-MS is the retention time at which a specific peptide is been detected at its maximum intensity in the mass spectrometer. As routine in chromatography, a peptide's intensity signal is integrated over

time and normalized by locally weighted polynomial regression (LOWESS) and the resulting paired retention time and intensity data that represents the respective peptide is called a peptide feature. The smallest quantity of compound yielding a definite signal-to noise ratio (10:1) is the sensitivity. The ratio of the number of ions reaching the detector and the number of ions entering the mass analyzer is the transmission of the mass analyzer and the part of ions of a particular m/z produced in the source that are effectively analyzed after passing the whole mass spectrometer is called the duty cycle of the instrument. The modern mass spectrometer is not only a balance and counter, but is also capable for fragmentation of the analytes and subsequent analysis of the fragments. The instrument which is used in this work is the LTQ Orbitrap XL (Thermo Fisher Scientific, Germany). It is capable for four different fragmentation technics: collision induced fragmentation (CID), pulsed Q-dissociation (PQD), high-energy collisional dissociation (HCD) and electron transfer dissociation (HCD). Here, we use solely CID which is described in the linear ion trap chapter of this introduction. The spectrum of the unfragmented peptide is called MS or MS^1 spectrum, the fragment ion spectrum is called MSMS or MS^2 spectrum and when fragmentation is applied again on a specific fragment, the spectrum of the new fragments is called MS^3 spectrum. When all MS^1 peaks are added and this signal is plotted over the time, this chromatogram is called total ion chromatogram (TIC). When noise is filtered in the MS^1 spectra prior to this and a baseline correction is carried out, then the resulting chromatogram is called base peak chromatogram (BP). It is also possible to plot only the signals of a specific mass. Doing so leads to the extracted ion chromatogram (XIC).

1.3 The mass spectrometer and its operation modes

A mass spectrometer as it is used in proteomics consists of the following components: A source, were peptides are transferred into the gas phase, ion lenses to filter and constrict the ion cloud, multipoles as ion transport devices, a mass analyzer, a collision cell and detectors. Some components are realized as a combination of those, like linear ion traps that consists of quadrupoles and lenses and are used as a fragmentation cell and as a mass analyzer, or an orbitrap that represents a combination of a mass analyzer and detector.

1.3.1 Electrospray ionization (ESI) is not the only source allowing online-coupling of a mass spectrometer to a chromatography system. Others are e.g. atmospheric pressure chemical ionization (APCI), thermo spray and gas chromatography-MS (GC-MS). These methods have specific preferences for the analytes' molecular weight and polarity with respect to their operating efficiency (Figure 1.1-A) making electrospray ionization coupled to HPLC the method of choice for the analysis of peptides. Peptides elute from the reversed phase HPLC-column, with an acidic water/acetonitrile gradient in a constant flow of around 200 nl/min and are transported to a metal coated needle (spray-tip). In between this needle coat and an appr. 5-10 mm distant orifice of the mass spectrometer, a voltage of about 2 kV is applied. Drops formed at the spray-tip are accelerated through this field and the low

pressure inside the mass spectrometer. On the way to the orifice, the solvent evaporates. When the surface- accumulated charges are too dense, the drop explodes leaving the charges remaining on the now gaseous peptides. The majority of peptides receive two or three charges (protons) in that way. However, the average number of charges increases with increasing size of the peptides. Proteins injected in that way receive 20, 30 or more charges (Figure 1.1-B).
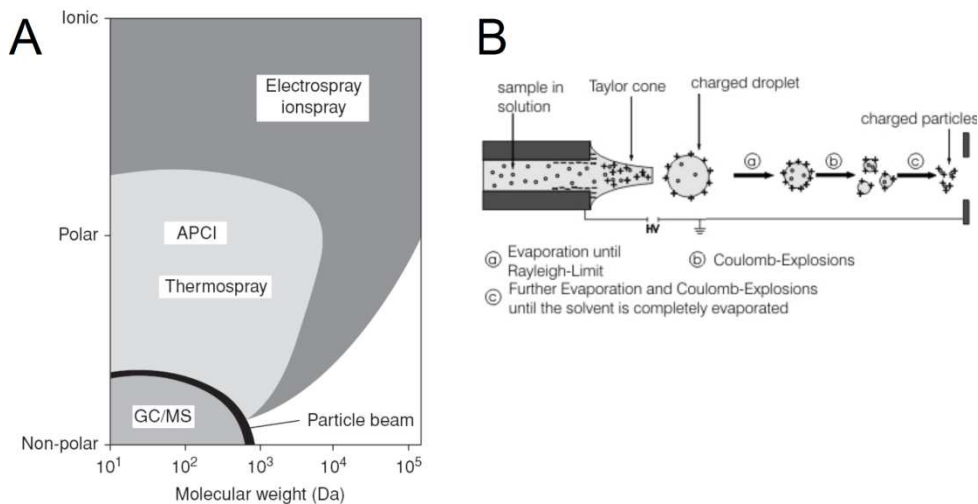
Figure 1.1: Electrospray ionization is the ion source in LC-MS based proteomics. A - Ionization capabilities of different source/mass spectrometer combinations with respect to the analytes' molecular weight and polarity. Tryptic peptides occur as ions with a suitable mass range of 500 to 10000 dalton. This elucidates electrospray ionization as the source of choice when analyzing tryptic peptides (taken from de Hoffmann, 2007). B - Electrospray ionization showing the needle tip on the left and the orifice of the mass spectrometer on the right side between which a constant voltage of about 2kV is applied (modified from www.physik.hu-berlin.de).

Only charged peptides can be analyzed in the mass spectrometer, which can be operated in two modes: negative, which means that peptides are negatively charged, or positive when peptides with one or more protons attached are analyzed. As protonated peptides are more stable than peptides with an overload on electrons, achievable intensities and resolutions are higher and the positive mode is the standard in proteomics. Cleaving proteins with trypsin ensures that, except the carboxy-terminal peptide, all tryptic peptides receive one or more basic residues, arginine or lysine, that stabilizes protons.

1.3.2 Multipoles. Ions are transported inside the mass spectrometer by multipoles. Such multipoles commonly consist of either eight (octapole), six (hexapole) or four (quadrupole) metal rods with round, square or hyperbolic cross-sections. The transport characteristics of ion clouds depend on the design of the multipoles (Figure 1.2-A). To construct a linear ion trap a round quadrupole exhibits best ion filtering characteristics, whereas for the transport of ions from one component to another octapoles are suited best. Those transport ways are important, because they allow in combination with the separation of different instrumental compartments by electrical lenses to operate at different pressures, as it is the case for

instance with quadrupole filters and collision cells. Also, those components need to be separated to minimize overlapping of electric fields.
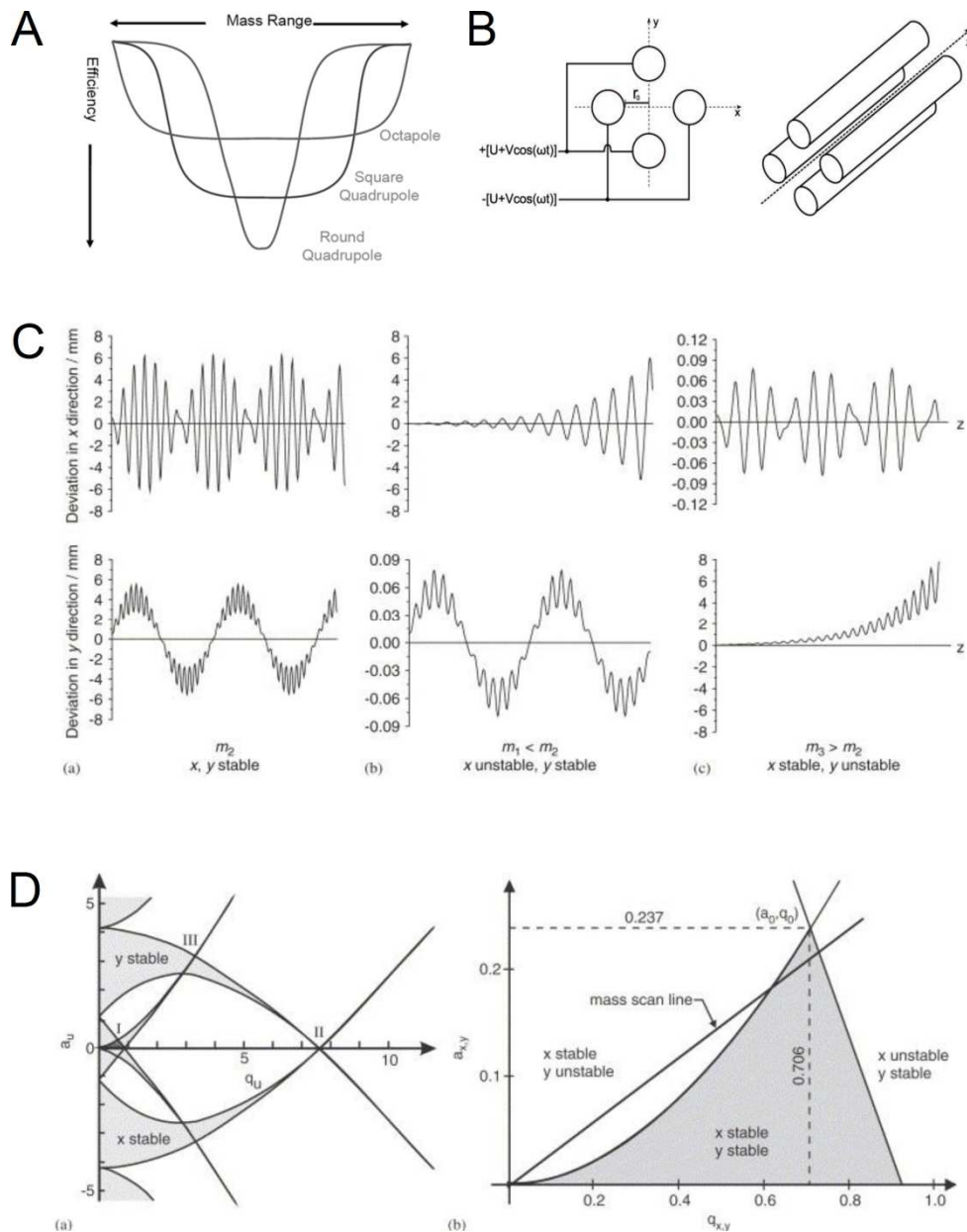
Figure 1.2: Multipoles as ion guides and filters. A – Transport characteristics are depending on the number and profile of the rods (source: www.thermofisher.com). B – A constant voltage superposed with an alternating is applied with different polarities on neighboring rods (taken from Benedikt, 2012). C - Examples for ion trajectories. D – Mathieu stability diagram for a quadrupole. Right pane: zoom-in of the first stability region which is generally the operating region (C,D taken from Blaum, 2006).

In addition to a constant voltage (U), an alternating voltage (V) is applied with opposite polarity on adjacent rods (Figure 1.2-B). Ions which are axially (z-direction) entering the multipoles, start to oscillate in the orthogonal x and y directions while continuing their axial movement resulting in a complex trajectory. The intensities and frequency of the applied

voltages determine the trajectories' stability of a specific mass to charge ratio range of the ions. In addition to the constant movement of the ions in the z-direction, the mathematical deduction of ion motion in the x,y-pane lead to the following differential expressions of ion movement (Paul equations):

$$\frac{d^2x}{dx^2} + \frac{2ze}{mr_0^2}(U - V\cos\omega t)x = 0$$

$$\frac{d^2y}{dy^2} - \frac{2ze}{mr_0^2}(U - V\cos\omega t)y = 0$$

With x, y being the coordinates of an ion of mass m and charge z, e the elementary charge, r_0 the inner space between the rods, t the time, ω the angular frequency and U and V the constant and alternating voltages applied to the rods as described above. These equations were seen in analogy to the former work of Mathieu (1866) on the propagation of waves in membranes and lead to two terms and parameters:

$$a = \frac{8zeU}{m\omega^2 r_0^2} \quad \text{and} \quad q = \frac{4zeV}{m\omega^2 r_0^2}$$

These terms describe a relationship between ion coordinates and time (de Hoffmann, 2007). Ions are instable when their x- or y-coordinates are reaching r_0 and they thus touch the rods and discharge. The a- and q-parameters are used to draw a Mathieu stability diagram (Figure 1.2-D), showing regions of pairs of a- and q-values that lead to stable ion trajectories (Figure 1.2-C). The stability diagram is an overlay of calculations for the x-and y-direction. Ions with different masses (mass to charge ratios) can be positioned in the stability diagram by altering V and U. The right pane in figure 1.2-D shows a zoom-in representing the common operating region. At a q-value of 0.706 and an a-value of 0.237, the stability region narrows. When the voltages set a specific m/z-ion to that point, other ions are instable and get lost enabling to operate the quadrupole as a filter in that way. When multipoles are operated as an ion guide, the a-parameter is set to zero by switching the constant voltage U off. All ions with a higher m/z than the one that for a given voltage V holds a q-value of 0.908 have stable trajectories and thus pass the quadrupole. However, ions with a very low q-value are poorly focused and hence also get lost. These are the reasons for the symmetric function shapes in figure 1.2-A. As a rule of thumb, for a quadrupole the lowest mass (m/z) observable (m_{min}) by adjusting V to yield a q-value of 0.908 for that m/z restricts the maximum observable mass to be $m_{max} = 20 \times m_{min}$.

1.3.3 Linear ion traps (LIT, LTQ) are constructed by superposing a potential well along the z-axis to a quadrupole field. This can be achieved by additional lenses or segmentation of the rods and the application of a constant voltage between the outer rods/lenses and the central rods (Figure 1.3-A and –B). This potential well causes ions to oscillate along the z-axis. Figure 1.3-A also depicts that the x-rods have slits for radial ejection of ions through the

x-rods to detectors. The Mathieu stability diagram of the linear ion trap favors the quadrupole's, but is asymmetric with respect to the q-axis (Figure 1.3-C).
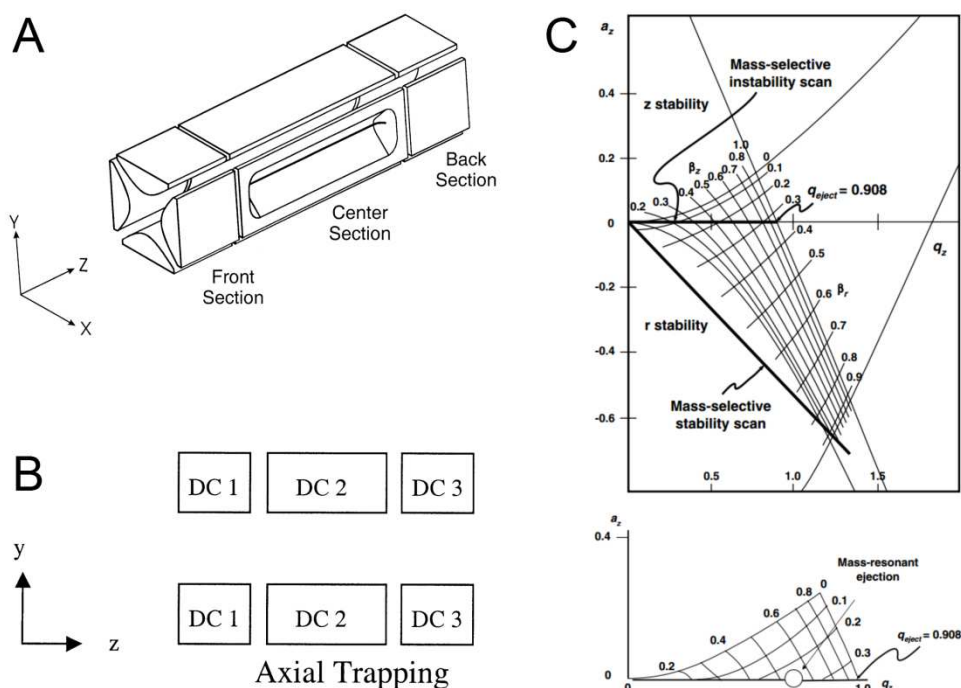
Figure 1.3: The linear ion trap is a versatile component and mass analyzer. A – Diagram of a linear ion trap as it is implemented within the LTQ orbitrap system. The two x-rods are provided with slits to enable radial resonance ejection. B - The hyperbolic rods are divided into inner and outer sections to add a constant voltage in z-direction that causes trapping. C –Matthieu-stability diagram for the two dimensional Paul-trap and operating region, lower panel: zoom-in of the operating region. (A,B from Schwartz, 2002; C from Fernández, 2007)

Ions enter the linear ion trap with a dispersed translational energy, but can be synchronized by collisional cooling with helium gas. This can be directly observed by an increase in resolution.

If the constant voltage U on the center rods is switched off, all ions are on the q-axis in the stability diagram. By ramping the alternating voltage V the ions are subsequently (smallest m/z values first) passing the q=0.908 instability barrier and are axially ejected. This is called the mass-selective instability mode. Ions are ejected as a function of the applied alternating voltage V (Fernández, 2007) and the mass spectrum can be obtained by rearranging the expression for q:

$$\frac{m}{z} = \frac{4eV}{\omega^2 r_0^2 q_z}$$

Another ejection mode is the resonance ion ejection. The basis of this method is the secular frequency with which all ions of a specific m/z oscillate along z. An alternating voltage applied to the end caps, or x-rods of the same frequency selectively increases the amplitude of this ions through resonance and causes those ions to be ejected axillary or radially in –x- or x-direction through the slits of the rods. Resonance ejection can be applied

to ions irrespective of their position within the stability diagram. An important application of axillary resonance ejection is the precursor isolation. It is conducted by ejection of all ions except the precursor. There is also a mass-selective stability mode, were the applied potentials U and V are selected that the a- und q-values are close to the apex of the stability diagram (Dass, 2007) and as described before for the multipoles, a single species is remaining stable und thus be isolated. Finally, the linear ion trap can be operated as an ion guide like ordinary multipoles.

After precursor isolation, the precursor can be positioned at a low q-value (0.25) and axillary resonance can be applied at an intensity omitting ejection. The excited ions will collide with the helium atoms and collision induced fragmentation occurs. The disadvantage of this is that fragment ions with m/z values of 1/3 of the precursor m/z are outside the q=0.908 instability barrier and will thus be not detected. This is the reason why iTRAQ is not applicable to standard linear ion traps, because of the small size of the reporter ions.

1.3.4 Orbitrap. The orbitrap is a combination of a high resolution mass analyzer and detector. It is composed of a central spindle electrode and an outer shell that consists of two electrical isolated halves. A constant voltage is applied between both shell halves on the one side and the spindle (electrostatic trap). Ions injected perpendicular to the spindle axis (z-axis) develop an intricate spiral trajectory composed of a circulation around and an oscillation along the spindle (Figure 1.4) due to the influence of the potential:

$$U(r,z) = \frac{k}{2} \cdot z^2 - \frac{k \cdot r^2}{4} + \frac{k}{2} \cdot R_m^2 \cdot \ln\left(\frac{r}{R_m}\right) + C$$

Where r and z are cylindrical coordinates with z as the axis of symmetry, k the field curvature, R_m the characteristic radius and C a constant. When integrating this equation with respect to z, this equation is reduced to k·z. The force that effects the z-oscillation is:

$$F_z = -q \cdot \frac{\delta U}{\delta z}$$

And equals the force that has to be applied to an ion of mass m and charge q (don't intermix with z, as z is the position along the spindle axis here) to move it:

$$F_z = m \cdot \frac{d^2z}{dt^2} = -q \cdot \frac{\delta U}{\delta z} = -qkz \quad \Leftrightarrow \quad \frac{d^2z}{dt^2} = -\frac{q}{m} \cdot k \cdot z$$

This is the equation of movement for the oscillation of an ion of mass m and charge q along the z-axis. Solving this equation leads to a combined sinus and cosinus function (de Hoffmann, 2007) with the angular frequency term:

$$\omega = \sqrt{\frac{q}{m \cdot k}}$$

This means that the frequency of the oscillation along the z-axis depends only on the mass to charge ratio of the ions (k is a constant). This oscillation results in an induced electric field which can be measured as a potential difference between the orbitrap's outer shell halves. A Fourier transformation of this image current yields all angular frequencies and hence mass to charge ratios inside the orbitrap, along with the respective intensity proportions.
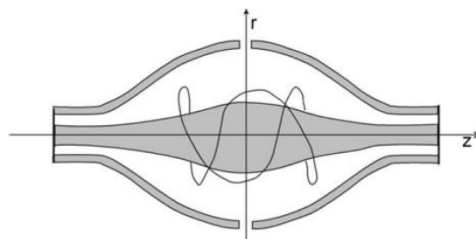
Figure 1.4: An orbitrap is an electrostatic trap. After radial injection, the ions start to oscillate around and along the inner spindle electrode (taken from de Hofmann, 2007; Makarov, 2000).

1.3.5 Secondary electron multipliers. As the orbitrap has intrinsic ion detection, the linear ion trap has not and is hence equipped with two secondary electron multipliers to detect ions that are radially ejected through the slits of the respective rods in x- and opposite direction. Ions are hitting a conversion dynode and release electrons that are accelerated by a static electric field, hitting themselves another dynode and release a multiple of other electrons and so on. An initial ion is triggering an amplifying cascade of secondary electrons leading to an electric current that represents the measurement intensity (Figure 1.5).
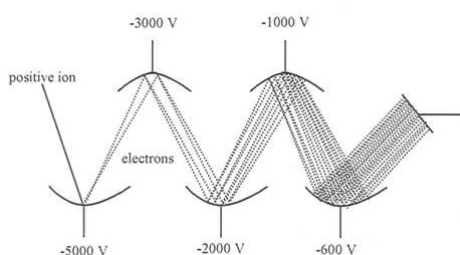
Figure 1.5: The secondary electron multiplier is a high sensitive ion detector. An ion causes an amplifying cascade of electrons which can be measures as an electric current. The disadvantage of multipliers is the decomposition of the dynodes. This has to be compensated by a frequent calibration, i.e. determination and application of an increased acceleration voltage. Nevertheless, they have to be replaced annually which makes them a major part in the running costs of the instrument.

Beside their respective operation modes, the mass analyzers described here differ in principle in scanning speed (Multipoles > LIT > orbitrap), upper mass limits (Quadrupole: 4000 Th, LIT: 6000 Th, Orbitrap: 50000 Th), resolutions (Quadrupole: 2000, LIT: 4000, Orbitrap: 100000) and accuracies (Quadrupole: 100 ppm, LIT: 100 ppm, Orbitrap: <5 ppm).

1.3.6 LTQ Orbitrap XL mass spectrometer. In the LTQ Orbitrap XL system (Thermo Fisher Scientific, Germany), a linear ion trap (LTQ) is coupled to an orbitrap. This creates a flexible system allowing for different fragmentation technics, namely CID and PQD and as an ETD (fluoranthene radical ion) source and a HCD cell is added as well, ETD and HCD. This is combined with the opportunity to acquire high resolution spectra in parallel. It is this, the high resolution, high flexibility and high scanning rate which accounts for the impact of this instrument to the –omics labs. We found CID superior in terms of protein and proteome coverage and used this fragmentation method throughout the experiments presented in this thesis. However, it is also possible to use a conditional selection procedure (decision tree) of fragmentation methods based on the masses and charges of the precursor ions (Figure 1.6).
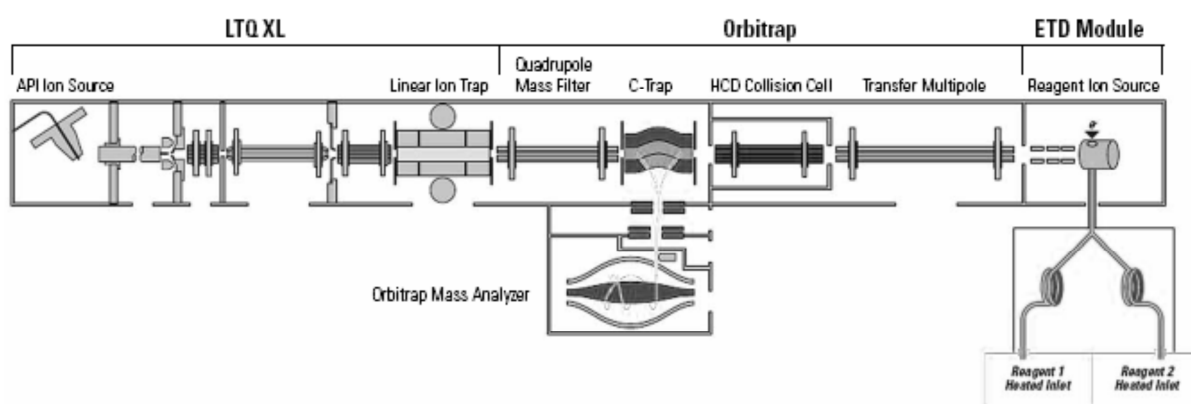
Figure 1.6: The LTQ Orbitrap XL ETD System as it is used in our lab. Other than the shown atmospheric pressure ionization source, for the measurement of peptides we use a nano-electrospray ionization source (nanoESI) (source: www.thermofisher.com).

1.3.7 TopN-approach. In shotgun proteomics, the standard operation procedure for the LTQ orbitrap system is the TopN-approach. A MS^1 spectrum is acquired with high resolution in the orbitrap. Right after beginning this process the system is using the preliminary data to automatically select precursors for parallel fragmentation and detection in the linear ion trap. While the MS^1 spectrum is still measured, a number of MS^2 spectra are acquired. The number of MS^2 spectra should be programmed in a way that when the acquisition of the MS^1 spectrum is finished, the acquisition of MS^2 spectra is also finished and the cycle can be repeated again. The duration of the MS^1 spectrum acquisition is depending on the resolution that should be achieved. For a resolution of 60000, the cycle time is appr. 2.8 s allowing to acquire 8 MS^2 spectra. The criterion for the precursor selection is intensity. When the 8 most intense peptides are selected for fragmentation in that way, the method is called Top8-method (Figure 1.7). When a complex proteome is measured, peptide separation is imperfect and a lot of peptides co-elute from the reversed phase column and enter the mass spectrometer. The TopN method could lead to miss a lot of peptides that are never among the top eight, while on the other hand many fragmentation spectra are acquired for the same peptides. To adjust for this, the operation program contains a

dynamic exclusion option. A peptide can be excluded thus from being fragmented again for a certain time (e.g. 15 s), if it has already been fragmented a couple of times (e.g. 2 times) within a given period (e.g. 5 s). The transient exclusion list can hold up to 500 peptides allowing a much deeper coverage of the peptidome and hence number of proteins identified. Resolution, cycle time, number of MS^2 spectra triggered and the settings for dynamic exclusion are parameters that can highly influence protein identification and label-free quantification such as spectral count. The influence of dynamic exclusion settings on spectral counting has been studied by Zhang et al. (Zhang, 2009). They found an ideal duration time of 100 s. However, as they also showed that this is a matter of the chromatographic parameters, like gradient time (132 min) and peak capacity, we found a duration time of 15 s with a repeat count of 2 within 5 s ideal for our assembly (gradient time 60 min, peak capacity around 1000). We didn't optimize the size of the exclusion list but set it to the maximum of 500 peptides instead. Kim et al. (Kim, 2010) analyzed the influence of different resolution settings, cycle times and if high resolution MS^2 spectra subsequent MS^1 in the orbitrap are superior to a shorter cycle time when MS^2 spectra are acquired in parallel in the LTQ. They found that parallel acquisition is clearly superior with slight advantages for shorter cycle times, i.e. lower MS^1 resolution. Their ideal resolution of 30000 is close to our experience (60000).
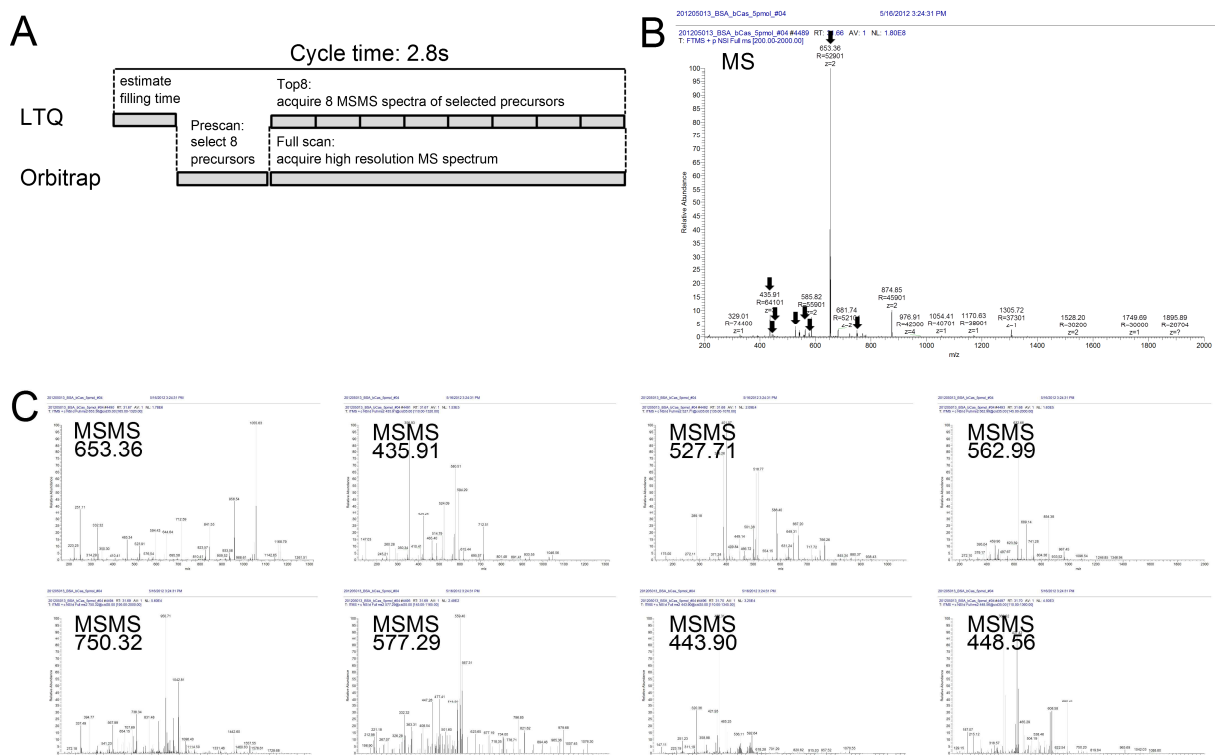
Figure 1.7: Top8 operation method for the LTQ Orbitrap XL. A - Low resolution MS^2 spectra are acquired in parallel to a high resolution MS^1 spectrum allowing higher duty cycles. B - MS^1 spectrum with labeled selected precursors. Not all of the 8 precursors are among the 8 most prominent MS^1 peaks. This is because dynamic exclusion was allowed that is aimed to prevent repetitive selection of the same precursors more than twice within 5 cycles and hence enhance the chance of co-eluting peptides to get fragmented as well. C - MS^2 spectra of the selected precursors.

1.4 Protein identification

1.4.1 Collision induced fragmentation of peptides leads to a rupture of the peptides at the amide bond of the peptide backbone. Other types of ruptures also occur but less likely, expressed in less intense MS^2 peaks. However, this is not true for phosphoserine or -threonine containing peptides as the collision energy is consumed here by the loss of phosphate. Fragmentation is a stochastic process, insofar as the amide bond that is ruptured is not the same one for every peptide molecule. This results in a ladder of peptide fragments with mass differences according to the amino acids between them. The situation is a bit complicated by the fact that the peptide charges can be transferred after cleavage to the carboxy- (y-ions) or amino-terminal (b-ions) part (only charged fragments can be observed in the mass spectrometer). Due to the basic residue at the carboxy-terminal part the acceptance of protons is more likely and thus the y-ions are more intense and dominate the MS^2 spectra. Figure 1.8 gives an example for the manual interpretation of such a spectrum. However, spectra of such quality are rare and as in LC-MS, within a two hour measurement ten thousands of spectra are generated, manual interpretation is not possible and algorithms have to be used.
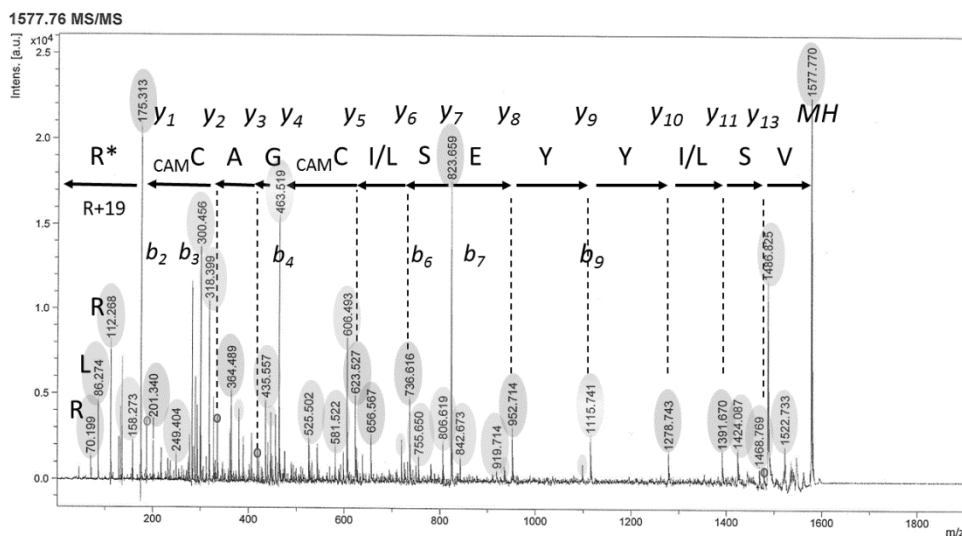
Figure 1.8: Example of a manual interpretation of a MS^2 spectrum performed on a MALDI-TOFTOF mass spectrometer (z=+1). The whole sequence had been deducible from this good quality spectrum. Unfortunately, this is not typical and only partial sequences can usually be elucidated. In addition, MS² spectra are commonly affected by imperfect precursor isolation in the LTQ and hence polluted by interfering fragment peaks.

1.4.2 Error control. Protein identification is a multi-stage process. This starts basically with the selection of precursors by the operating program for the mass spectrometer. The next step is the assignment of MS^2 spectra to peptides. This does not lead to a single peptide identification, but to a list of peptides of the same precursor mass that can represent the spectrum with a certain score. The score is a rating by the algorithm reflecting the grade of

matching, but it is not a probability. Characteristics that influence the score are the number of peaks matching to theoretical y- and b-ions and the number of remaining unmatched peaks. Every spectrum contains noise peaks, mainly from improper precursor isolation and side peaks, for instance from side chain fragmentations. This blurs the matching process and hence also a peptide of a lower score can in fact be the true origin of a certain fragment ion spectrum. The next step is the assignment of identified peptides to proteins. This is also a probabilistic step and the algorithms used assign scores to the matchings as well. Often a matching is not unique and leads to a group of proteins rather than to a single protein (Nesvizhskii, 2005). If probabilistic steps are consecutively applied and highly parallelized, errors will potentize and thus need to be controlled. This error control is performed by fixing the false discovery rate (Figure 1.9).
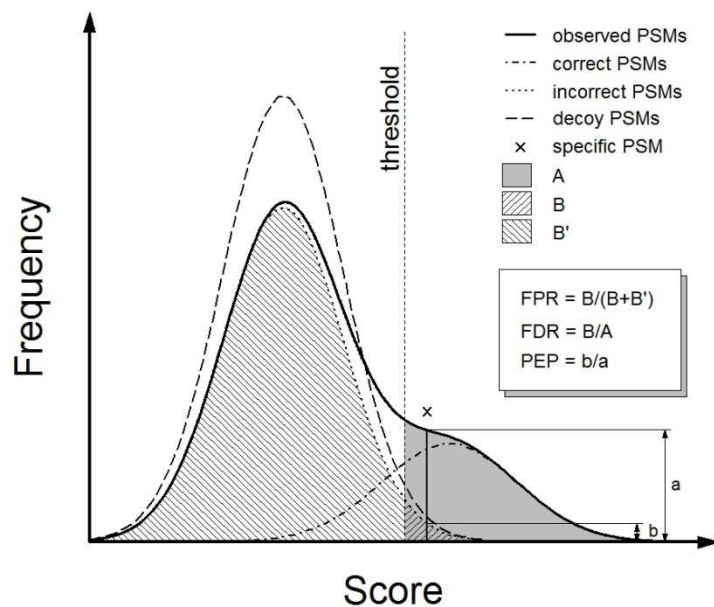
Figure 1.9: Error control demonstrated on peptide identification. The plot is a histogram reflecting the frequency at which peptides at specific scores (score ranges) occur. The unbroken line is what is commonly observed (observed peptide spectrum matches - PSM). This function is obviously a sum of two underlying Gaussian distributions. Nearly all assigned PSMs are falsely assigned (incorrect PSMs) while a small proportion with a higher mean score is correct. Unfortunately there is an overlap between both subdistributions. The distribution of incorrect PSMs can be experimentally assessed. All MS^2 spectra are probed hereby for matching to a nonsense (decoy) database which can be generated by reversing the real protein database before in-silico generation of theoretical tryptic peptides. This leads to a distribution simulating the incorrect PSM distribution, while omitting the fraction of correct PSMs. The question is now, how to set a score threshold for the acceptance of identified peptides. If the threshold is been set too high, while decreasing the false positives, the false negatives are increased finally resulting in a shortened list of identified proteins. If the threshold is too low, then the number of false positives is too high, leading to many false protein entries. From the distributions, one can calculate the false discovery rate which is defined as the fraction of false positives among all proteins passing the threshold. The score threshold is set allowing a certain false discovery rate (usually 1%) (adopted from Brosch, 2010).

1.5 Quantitative mass spectrometric approaches

Mass spectrometry has for long been considered as being inapplicable for quantitative proteomics, because of two major drawbacks. On the one hand, identifications and estimated quantities of proteins differ from LC-MS run to run. Slight differences in the sample loading by the autosampler as well as in the formation of the LC-gradient occur. One sample is usually constituted of ten thousands of peptides which are only partially separated in the chromatography, with a common peak capacity of around thousand. Hence, different peptides co-elute and compete for protons (charges) during ionization leading to suppression ('ionic suppression') of the inferior peptide species. Run to run differences in the chromatographic elution entail run to run differences in the detection of peptides and therefore protein identification. On the other hand, even in a single LC-MS run, the same amounts of different peptides, like those generated by tryptic digestion of a certain protein differ in the mass spectrometric intensities they cause. Two different proteins of the same amount will result in different sum intensities of their tryptic peptides in that way. This is, because the peptide's primary sequences and hence physicochemical properties differ largely. Also, the number of tryptic peptides differ largely among different proteins. Although bigger proteins tend to cause larger numbers of tryptic peptides, this is not a strong correlation because of large variations in the occurrence of lysine and arginine after which trypsin cleaves. Those two amino acids and histidine are the only amino acids that are positively charged in the chromatographic separation prior to mass spectrometry at pH=3. On the other hand, a higher content of acidic amino acid residues reduces the chance of a tryptic peptide to successfully compete with co-eluting peptides for accepting charges during ionization in positive mode. Hydrophilicity and hydrophobicity are further primary sequence-depending values restricting identification success. Highly hydrophobic peptides are difficult to keep soluble or to remove from the reversed phase columns, whereas highly hydrophilic peptides could be lost because of improper binding and retaining in the chromatography. During sample handling and processing, peptides are exposed to many kinds of surfaces, different plastics, glass, metals and polymeric bead materials to which they could be differentially bound and diminished (Goebel-Stengel, 2011). These factors hinder direct inference of protein amounts from their tryptic peptide signal intensities. In addition, proteins can differ in the type and grade of post translational modifications on specific sites in between to experimental samples. Often there is no evidence of these and hence lead to false inference from peptide to protein amount changes.

1.5.1 Labeling approaches. Several strategies have been developed to circumvent these problems. In the so-called labeling approaches, peptides from proteins of one sample are differentially labeled than proteins from another sample or control making it possible to compare pairs or even octaplets of the same peptides and deduce related protein changes in a single LC-MS run. One of the first such approaches introduced isotope-coded affinity tags (ICAT) (Gygi, 1999). Those tags are heterobifunctional linkers exhibiting a thiol-reactive group for coupling to cysteines and a biotin-group for affinity isolation on the other side.

The middle region of the linker either contains eight deuteriums or not. While only slightly changing physicochemical properties, peptides originating from different samples can be distinguished by mass spectrometry and its intensities compared in the MS[1]–spectra of a single run. Some drawbacks of the method are the loss of all cysteine-free peptides and the limitation to duplex labels, both of which have been addressed by the use of isobaric tags for relative and absolute quantitation (iTRAQ) (Ross, 2004) and Tandem Mass Tags (TMT) (Dayon, 2008). These tags are amine-reactive and do not differ in their overall mass and therefore do not enhance complexity of MS-spectra which reduces the number of different peptides selected for MS[2] and hence protein identification rates. The fragmentation process releases a set of chemical identical reporter groups comprising different combinations of light and heavy stable isotopes of carbon, nitrogen and oxygen. Up to eight reporter group peaks appear apart in the lysine containing peptide's MS[2] spectra enabling octaplex experiments. One disadvantage of chemical labels is the bias of chemical reactions to specific sites, imperfect labeling and side reactions. Amine-directed labeling depends on unmodified lysine residues and all arginine containing tryptic peptides are useless for quantification. Also, labels are introduced late in the sample processing workflow and hence samples can be combined tardily. Errors during sample processing can influence all samples unequally and lead to introduced variations. This chain of errors is smartly prevented by the popular method of stable isotope labeling of amino acids in cell culture (SILAC) (Ong, 2002). Cells in culture are fed herein with the isotope labeled essential amino acids arginine or lysine. This early labeling allows for combining of samples right after cell lysis. However, several constraints restrict the usage of SILAC. In the SILAC-approach, the complexity of MS-spectra is multiplied and so is the shotgun proteomics immanent problem of stochastically undetected proteins. This necessitates more protein and peptide fractionation steps, more replicates and finally resulting in more LC-MS runs. SILAC relies on multipassage cell cultures to achieve an indispensable maximum labeling efficiency of 98%. If the labeling efficiency is even minimal reduced the calculated protein ratios will drastically underestimate huge actual ratios. The labeling efficiency of arginine-containing peptides can also be reduced, if the cell possesses an arginine-proline interconversion pathway (Park, 2009), while proline could be itself partially labeled in that way. All in all, SILAC is restricted to specific long term cultivated cell lines of specific organisms and cell types. This impedes its use in stem cell proteomics where cultivation time is known to affect the cell surface composition (Karbanová, 2010).

1.5.2 Label-free approaches. In recent years, label-free methods became very popular, as they can be applied to any biological sample. Additional benefits are low costs, ease of implementation and a non increase of MS[1] complexity. In former years, the higher number of LC-MS runs needed, as label-free mass spectrometry relies on parallel sampling, cleared some of the gains in costs. With the high sensitivity and high duty cycles of modern high resolution mass spectrometers, less prefractionation is needed which drastically reduces the numbers of LC-MS runs and enable the researcher to investigate more biological replicates.

Less steps in prefractionation diminishes also errors in the parallel treatment of samples. In label-free MS approaches, protein abundances and abundance changes could be elucidated from mass spectrometric data using either empirical indices based on sequence coverage, number of observed peptides or number of triggered MS^2 , or inferred from measurement values like MS^1-intensity or MS^2-intensity and retention time.

While the quantitative bases determine the dynamic ranges of label-free methods, their linear ranges and accuracies are enhanced by engineered normalization formulas (Figure 1.10). In the cell surface proteome, we expect a distribution of protein abundances over several orders of magnitude. This necessitates a huge linear range of a quantitation method, particularly when a sample's protein composition should be assessed. Sample composition and the difference between two samples are two types of information desired to gain in a quantitative approach. The sample composition, an estimation of relative amounts of different proteins within a sample, enables definition of frequency classes and group identified proteins for their suitability to serve as cell surface markers. Furthermore, the sample composition provides information on the accumulation of specific gene ontologies or protein domains. This also enables to monitor the effectivity of dedicated enrichment procedures. Analysis of the proteomic difference between two or more samples elucidates changes upon perturbance, different donors, different time points during proliferation, different cultivation parameters, etc. Such changes are quantitatively expressed in ratios of estimated values and hence are sensitive for the accuracy of their estimation.
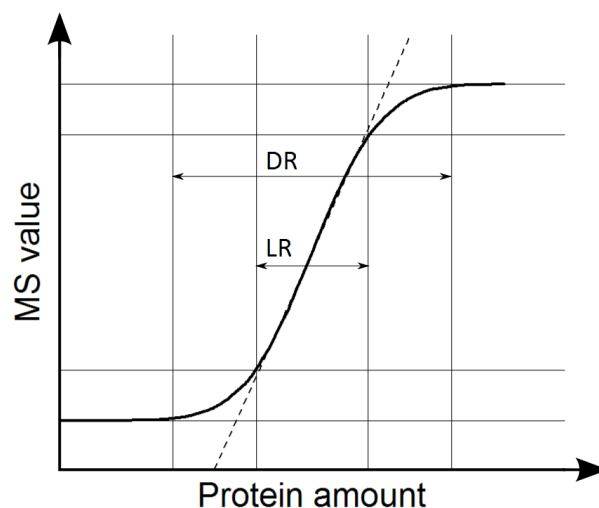
Figure 1.10: Dynamic range (DR) and linear range (LR) of a MS value that is used to assess protein amount (or concentration, respectively). The dynamic range is the range of protein amounts over which the MS value changes. The linear range is the range of protein amounts were the MS value has a certain degree of linear correlation with the protein amount.

In shotgun mass spectrometry early label-free indices used were sequence coverage and the number of different non overlapping peptides (PN) identified for a particular protein.

Their usage was mainly restricted to reflect the reliability of identification. Both values depend on the protein size, as high sequence coverage of smaller proteins is easier to achieve whereas a tryptic digestion of larger proteins will more likely result in a higher number of observable peptides (PO). An attempt to normalize PN was to simply divide by PO to yield the protein abundance index (PAI) (Rappsilber, 2002). Observation of a peptide can be restricted through mass ranges defined in the workflow (e.g. linear ion trap) and rated by its predicted chromatographic behavior. Further studies modified the PAI and defined an exponentially modified protein abundance index (emPAI) as 10^{PAI}-1 (Ishihama, 2005) and on the base of 6.5 (emPAI65) (Kudlicki, 2012). The concept of spectral counting (SC) was introduced by Liu et al. (Liu, 2004). In shotgun proteomics the mass spectrometer is programmed to run in acquisition cycles of MS^1 and several intermediate MS^2 spectra. From MS^1 spectra a number of most intensive signals trigger the selection of corresponding peptide ions as precursors for fragmentation. An enrichment of a certain protein in a sample will result in more intensive MS^1 signals of its tryptic peptides compared to a control and hence being more often selected for MS^2. Therefore, the MS^2 sampling rate, or the sum of peptide spectrum matches (PSM) assigned to a protein is an indirect quantitative value for a protein's abundance in one sample compared to another. In their work, Liu et al. demonstrated for six standard proteins mixed into a yeast proteome that spectral counts correlated well over the tested two orders of magnitude with the input amount. The value was clearly an improvement to the sequence coverage or number of identified tryptic peptides (PN). This work raised the task to test this value for its operating capacity over a broader dynamic range. A number of normalization procedures and derived values were subsequently developed. One of those is the normalized spectral abundance factor (NSAF) which normalizes spectral counts twofold. Firstly, all spectral counts of a specific protein are normalized by the protein's length (L). Secondly, the resulting SC/L ratio is divided by the sum of all such ratios for all proteins in a sample (Zybailov, 2006). Griffin et al. developed the normalized spectral index of an individual protein (SIN) that combines spectral counts with unique peptide numbers and MS^2 intensities (Griffin, 2010). They demonstrated that the index correlates well over three orders of magnitude with input amounts of standard proteins. It was further shown that SIN suits better on replicate measurements than PN, SC, NSAF and ratios of spectral counts (RSC) computed as proposed by Old et al. (Old, 2005). However, as all of the normalized indices rely also on the estimation of spectral counts, the availability of spectral count values determines also the dynamic range of all the indices.

Intensity based methods have been widely used to assess analyte amounts as in the analysis of small molecules. Chelius et al. (Chelius, 2002) prepared a concentration series of a tryptic digest of myoglobin from 10 to 105 fmol. They demonstrated that the sum area under the curve (AUC) of five peptide's MS^1-intensities in the chromatogram correlated well (r=0.995) with the amount of the protein over the five orders of magnitude tested. In addition, they still could quantify two different amounts of horse myoglobin (250 and 500 fmol) spiked into a digest of human serum proteins. We wanted to extend this analysis for a

broader range in a complex proteomic background. We did so using digests of bovine serum albumin and beta casein as standard proteins and further tested the accuracy of normalization methods that should obliterate physicochemical differences of the proteins on a mixture of 49 proteins spreading over a range of 0.5 to 50000 fmol. An intrinsic problem in the endeavor to create an intensity-protein amount correlation independently on the proteins' primary sequences are the different numbers of observable tryptic peptides that can be generated from them. Those numbers correlate not simply with the protein size, but moreover with the distribution of lysine and arginine residues. Too small and too large peptides are not observed in the mass spectrometer because of a maximum operation m/z-range of the mass analyzers (e.g. quadrupoles, linear ion traps). Also different physicochemical properties of the peptides could hinder them on binding to or eluting of the chromatographic column, or discriminate them in the ionization process. One approach is to reckon only the three most intensive peptides of a protein and constructing the protein intensity either by summing up (Ning, 2012), or averaging (Ishihama, 2008) those peptide intensities. The latter method is named extracted ion intensity-based protein abundance index (xPAI). This necessitates on the other hand the availability of at least three tryptic peptides and hence restricting the dynamic range on the bottom. On the other hand, a restriction to three tryptic peptides increases the susceptibility to appearing post translational modifications, if two experimental states are compared. Another approach is intensity normalization by division through the number of calculated observable tryptic peptides. This approach was introduced using a spiked-in universal proteomic standard and hence allowing even an intensity-based absolute quantification (iBAQ) (Schwanhäusser, 2011). The observability defined in that method was a sequence length restriction from 6 to 35 amino acids. When proteins across two or more samples ought to be compared, normalization should adjust for loading discrepancies. An approach realizing this is implemented in MaxQuant resulting in the label free quantification intensity (LFQ) values (Hubner, 2010). We compared ordinary MaxQuant intensities (MQ) (Cox, 2008), which represent the sum of intensity peak areas of extracted ion chromatograms (XIC) for all a protein's tryptic peptides, LFQ and xPAI for their dynamic and linear ranges and accuracies in the estimation of protein abundance.

Sample comparison approaches also generate the issue of an appropriate test statistic in which false discovery rates can be controlled. Roxas et al. (Roxas, 2008) adopted a method which has been developed for analyses of microarray experiments, the significance analysis of microarrays (SAM) (Tusher, 2001). This method is available through MaxQuant's Perseus package (www.maxquant.org) (Hubner, 2010). We prepared yeast 'background' proteomes with different spiked-in amounts of universal proteomic standard 1 (UPS1) proteins and used MaxQuant LFQ values to assess this method's capability to detect defined protein ratios.

1.6 Application I – Analysis of plasma membrane proteins

38% of all proteins encoded by the mammalian genome are classified as membrane proteins and they represent more than one-third of the current list of biomarker candidates and two-thirds of the existing drug's targets (Rusevic, 2011). Cell surface proteins arrange cell adhesion and thus determine cell localization and constitute an interface for communication to the cell's environment. Both adds to the cell's function, state and hence identity. The analysis of membrane proteins is still a challenge, as they are usually under-represented in proteomic studies (Mindaye, 2012). Low detection rates are thought to be reasoned by low abundance and experimentally by their poor solubility (Vuckovic, 2012). Many of the technical advances in plasma membrane proteomics have been made so far using cancer cell lines and a further development to the direct analysis of clinically relevant patient samples was demanded to finally identify biomarkers that improve early disease detection and allow prognosis and prediction of treatment response (Leth-Larsen, 2010). A major focus with this respect is the identification and treatment of cancer though, but beside the analysis of pathogenic cells knowing the cell surface proteome of stem cells will aid in their isolation. In fact, there are crucial experimental differences to cancer biopsies, like a small number of isolatable cells, small rate of cell division and alterations in the cell surface proteome during cultivation. We did a detailed identification of the cell surface proteome of mesenchymal stem cells at first (Niehage, 2010), which is still the most thorough study on this type of cells promising for their potentially therapeutic relevance. Upon implementation of more sophisticated and quantitative mass spectrometric workflows, we wanted to apply these to a related type of stem cells, the dental pulp stem cells (DPSC). The dental pulp has been demonstrated as an easy accessible source of somatic stem cells (Karbanová, 2011). Karbanová et al. also showed that cell surface proteomes develop differences in dependency on the culture media used, but did not identify those differences in an extensive untargeted approach. We wanted to fill the gap, as the mentioned experimental obstacles upon stem cells define this as a challenging approach to benchmark our procedures.

A common assumption of membrane proteomics is that membrane proteins need to be enriched prior analysis (Sprenger, 2010). A multiplicity of enrichment procedures for plasma membrane proteins have been developed among which the most important are density centrifugation, lysine-directed biotinylation (Zhang, 2003a) and hydrazide capturing of reduced N-glycosidic carbohydrate side chains (Zhang, 2003b). Density centrifugation is highly affected by contaminant proteins, since the densities of plasma membranes, endoplasmic reticulum and mitochondria overlap (Zhang, 2003a). The hydrazide capturing technic can be considered to deliver the least contaminants as captured proteins are temporarily covalently coupled to a solid matrix allowing intense washing. The hydrazide capturing technic on the other hand, select only proteins that are glycosylated and are expected to introduce the hugest bias though. Because of this, we chose lysine-directed biotinylation to enrich for cell surface proteins and compare this with two types of controls.

This is at first, a similar experiment with non biotinylated samples to elucidate proteins that bind to the bead material and constitute the proteomic background and at second, a total proteome sample to elucidate possible biases by comparing intensity distributions of all plasma membrane proteins that occur in both samples.

1.7 Application II – Analysis of membrane-associated protein networks

1.7.1 Vesicular trafficking. The plasma membrane proteome is very dynamic. Transmembrane proteins are transported in the secretory and endocytic pathways by vesicular trafficking. After synthesis, transmembrane proteins are transported from the endoplasmic reticulum (ER) to the cis-Golgi and further trough the Golgi to the trans-Golgi by COPI and COPII coated vesicles (see chapter 1.7.4). The trans-Golgi network (TGN) represents a sorting hub were different proteins are directed to different trafficking ways (Figure 1.11). The following chapters on trafficking complexes provide a more detailed view on the respective trafficking mechanisms.
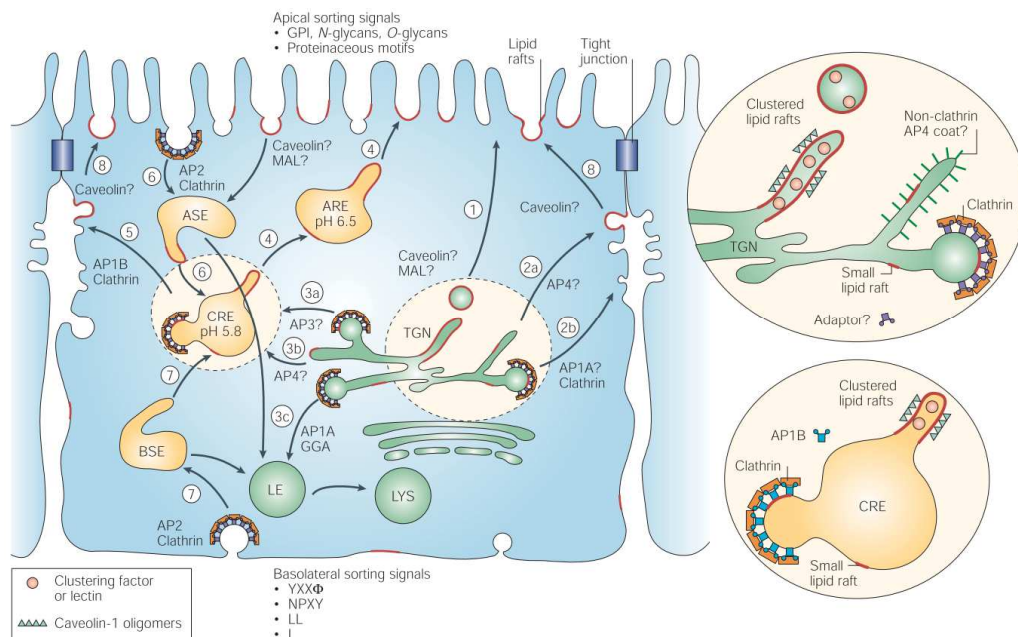
Figure 1.11: Some post-Golgi trafficking routes of transmembrane proteins in epithelial cells. Proteins exit the trans-Golgi network on unknown mechanisms to the apical plasma membrane (1) or basolateral, supposedly via AP-1A (2b) or AP-4 (2a) adapter complexes (AP). Other routes lead either to late endosomes, via AP-1A or GGAs (Golgi-localized, γ -ear-containing, Arf-binding protein) (3c), or to common recycling endosomes (CRE), supposedly via AP-3 (3a) or AP-4 (3b) adapter complexes. Proteins could leave the CRE either basolateral via the adapter complex AP-1B (5) or via apical recycling endosomes (ARE) (4) to the apical plasma membrane. Endocytosis directs transmembrane proteins from the plasma membrane without or with involvement of AP-2 (6 and 7) back to the endosomal system. From the endosomes, proteins can also traffic back to the Golgi via the retromer complex (not shown), directed to the lysosomes, or incorporated to multivesicular bodies via ESCRT-complexes (not shown) and even integrated into exosomes hereafter. (Figure: Rodriguez-Boulan, 2005)

The vesicles generate from a donor membrane where transmembrane proteins need to be transported. At the beginning, trafficking complexes recognize and gather protein cargos. Vesicles are formed hereupon that are cutted off the donor membrane (Vesicle scission or budding), moved to the acceptor membrane (Vesicle motility or movement) to which they attach (Vesicle tethering) and fuse (Vesicle fusion) (Figure 1.12). These steps are regulated by peripheral membrane proteins, as for instance distinct families of GTPases (Arfs regulate formation, dynamins regulate scission, and Rabs and Rhos tethering and fusion) (Segev, 2011; Behnia, 2005).
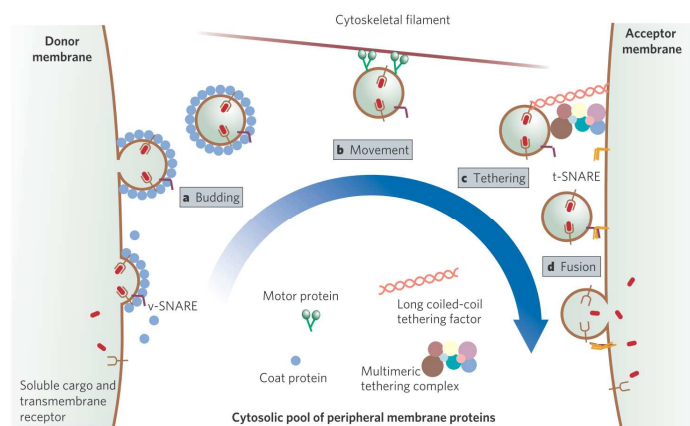
Figure 1.12: Steps of vesicular trafficking (Behnia, 2005).

The attachment of peripheral membrane proteins during vesicle formation premises the recognition of the type of membrane that represents the respective organelle. Also, the tethering of vesicles necessitates the recognition of the target membrane by the trafficking components involved. This implicates that organelle membranes differ in lipid composition, as shown by van Meer et al. (Figure 1.13) (van Meer, 2008). Although quantitative lipid analysis is hampered by inadequate isolation procedures that allow only insufficient separation of the different types of cellular membranes, it is known that the diverse types of lipids distribute heterogeneously across them. Despite of this lipid-content heterogeneities, phosphoinositides are thought to basically represent the identity of a membrane though.
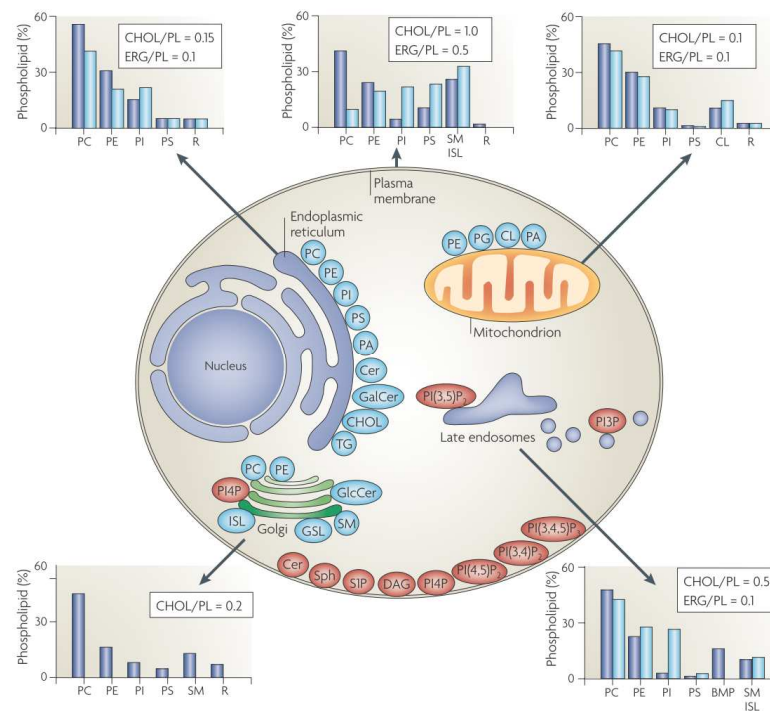
Figure 1.13: Lipid composition of cell membranes. Blue bars - mammals; Light blue bars - yeast (van Meer, 2008)

1.7.2 Cargo sorting motifs. The transmembrane protein cargo, with its embedded sorting motifs, is the nucleation core of trafficking complexes. The individual sorting of intrinsic membrane proteins requires sequence information in the cytosolic domains of the cargo that can be accessed by cytosolic components of the trafficking complexes. One of this sequence classes causes its ubiquitination, while others are directly recognized by the trafficking complexes. Table 1.1 lists some known sorting motifs. Embedded thyrosines, serines and threonines are often phosphorylated and thus allow controlling trafficking by intervening of signaling processes. Just a minor part of the actual cargos of known trafficking complexes can be assumed to be known. Thus, table 1.1 can be assumed to be incomplete as well. Trafficking complexes are also believed to bind cargo motifs with a moderate affinity only. In the coincidence detection model, the affinity is dramatically increased by the presence of specific phosphoinositides and small GTPases.

A former colleague in our group, Thorsten Baust, used a part of the cytosolic domain of varicella zoster virus glycoprotein E (gE, formerly gpI), covalently coupled to artificial liposomes to mimic trafficking complex formation in vitro (Baust, 2006). Figure 1.14B shows that the wildtype gE heavily recruits γ-adaptin (AP1G) along with the clathrin coat and the small GTPase Arf1, but as it seems, low amounts of members of the adapter complexes 2 (α-adaptin), 3 (σ3-subunit) or COP complexes (β-COP). Wild type and truncated or mutated gE variants (Figure 1.14A) were used to demonstrate the importance of the sorting motif in the recruitment of γ-adaptin and Arf1 (Figure 1.14C).

Table 1.1: Endosomal/lysosomal sorting motifs (adopted: Bonifacino, 2003; Burgos, 2010).

Motif	Proposed recognition protein or domain	Functions
NPXY	Clathrin terminal domain,μ2 subunit of AP-2, PTB domain of Dab2	Internalization
YXXØ	μ subunits of AP complexes	Internalization, lysosomal targeting, Basolateral targeting
(Y)KFFE	AP-4	TGN-to-endosome sorting
[DE]XXXL[LI]	μand/or β subunits of AP complexes	Internalization, lysosomal targeting, Basolateral targeting
DXXLL	VHS domain of the GGAs	TGN-to-endosomes sorting
Acidic cluster	PACS-1	Endosomes-to-TGN sorting
FW- or P-rich	TIP47	Endosomes-to-TGN sorting
NPFX(1,2)D	SHD1 domain of Saccharomyces cerevisiae S1a1p	Internalization
Ubiquitin	UIM, UBA, and UBC domains, ESCRT complexes	Internalization, lysosomal/vacuolar targeting

X - any amino acid; Ø - amino acid residue with a bulky hydrophobic side chain
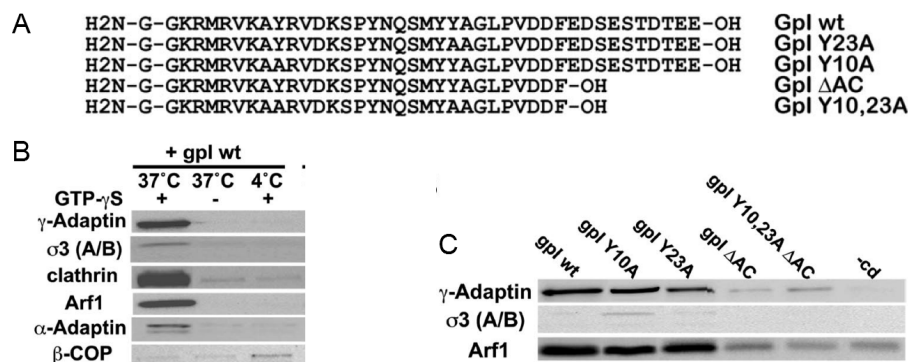
Figure 1.14: Recruitment experiments with gE coupled to artificial liposomes. A – sequences of wildtype and truncated or mutated variants of gE used. B – Western blots showing temperature and GTP-γS dependent recruitment of Arf1, γ-adaptin (AP1G) and clathrin. There is also indication of recruitment of some AP-2 and AP-3 (α-adaptin, σ3-subunit), whereas β-COP doesn't seem to be recruited. C – Sorting motif dependent recruitment of γ-adaptin and Arf1. –cd – control experiment without cytosolic domain of gE. (Baust, 2006).

1.7.3 Phosphoinositides. Along with the sorting motif, membranes are also recognized by the trafficking complexes. Often this is achieved through protein domains recognizing phosphoinositides. Phoshoinositides, or phosphatidylinositol phosphates (PIP) are derivatives of phosphatidylinositol (PI), were the inositol ring is decorated with organic phosphate in a combinatory manner on the 3, 4 and 5-positions (Figure 1.14). The different phosphoinositides are converted into another by specific kinases and phosphatases that act on specific membranes. Hence phosphoinositides are heterogeneously distributed among the membrane types. The exocytotic pathway is governed by PI4P that is predominantly occurring in the Golgi membrane but also found in the plasma membrane. The endocytic pathway on the other hand, is governed by PI3P in the early endosomes that is converted into PI35P2 which represents the late endosome or lysosomal membrane. PI4P in the plasma membrane is converted to PI45P2, the predominant phosphoinositide of the plasma membrane. Two phosphoinositides are generated in response to external signals, PI34P2

and PI345P3. PI345P3 is described to occur in specific regions of the plasma membrane (van Meer, 2008) or phagosomes (Byekova, 2010), PI34P2 at the plasma membrane (van Meer, 2008) (Figures 1.13 & 1.17). The subcellular localization and cellular function of PI5P remains elusive. As PI5P is produced via PI3P and PI35P2 by phosphoinositide 5-kinase PIKfyve and the phosphoinositide 3-phosphatase MTMR3 (Oppelt, 2013), it might be embedded into a PI35P2-containing organelle, its maturation products or derivated vesicles. Phosphoinositides work together with small GTPases as sign posts of membrane recognition by trafficking complexes (Behnia, 2005). Phosphoinositides aid to acquire the specific GTPases when they are activated by guanine nucleotide exchange factors (GEFs). Trafficking complex constituents concertedly recognize both then, the phosphoinositide and the GTPase (Figure 1.16) (Di Paolo, 2006).
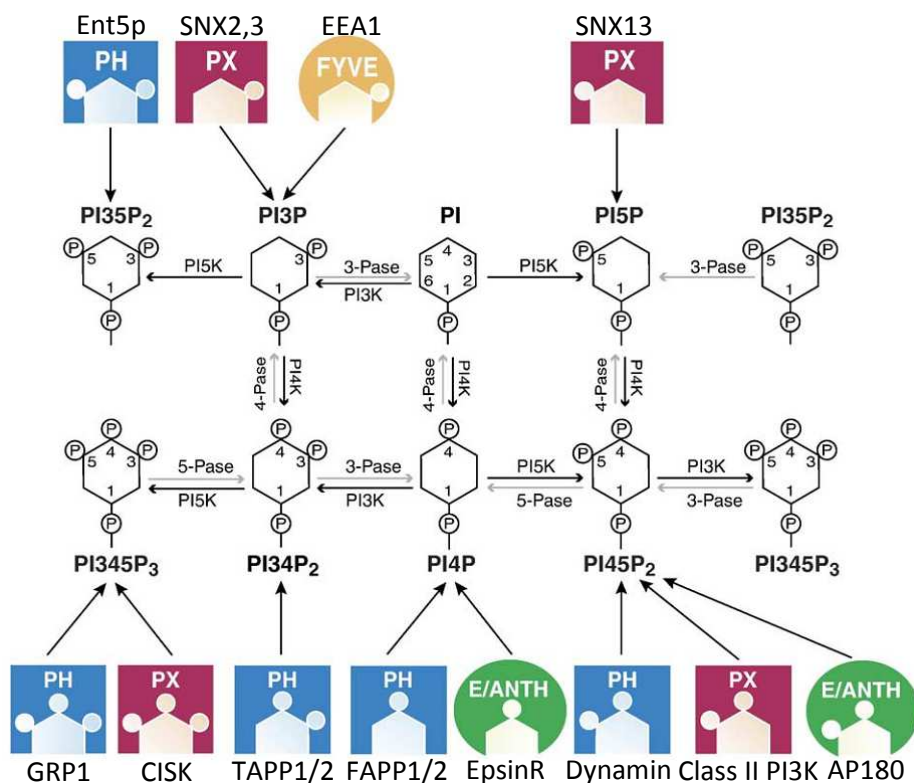
Figure 1.15: Phosphoinositides (PIPs), their kinases and phosphatases and protein domains that bind PIPs (altered from: DeMatteis, 2005), along with exemplary proteins (taken from DiPaolo, 2006).

1.7.4 Small GTPases. The identity of organelle membranes is also defined by small GTPases. The main classes of small GTPases contributing to this are Rab and Arf families (Behnia, 2005). GTPases switch between the inactive state when GDP is bound and the active state with a bound GTP (Figure 1.16-A). In the inactive state, those GTPases are mostly cytosolic but can also bind to membranes transiently. Membrane binding is enabled by a myristoyl group (Arfs) or palmitoyl groups (Rabs). The Rab GTPases are associated in the inactive state with a GDP-dissociation inhibitor (GDI) that needs to be released by an intrinsic membrane protein, the GDI-displacement factor (GDF). The interconversion from active to inactive

states is regulated by guanine nucleotide exchange factors (GEF) and GTPase-activating proteins (GAP). Specific GEFs recognize and convert inactive GTPases by recognizing both, the specific GTPase and the phosphoinositide that represents membrane identity. GAPs are likewise detecting the phosphoinositide and GTPase coincidently. This coincident detection is a general principle that allows triggering effector affinity due to cooperative binding (Figure 1.16-C). Coincidence detection also occurs when trafficking complexes bind to phosphoinositides and cargo (membrane receptor to be trafficked) sequences. Interestingly, the phosphoinositide converting enzymes like kinases (Figure 1.16-B) also detect and convert its substrate when they are bound to activated small GTPases in parallel.
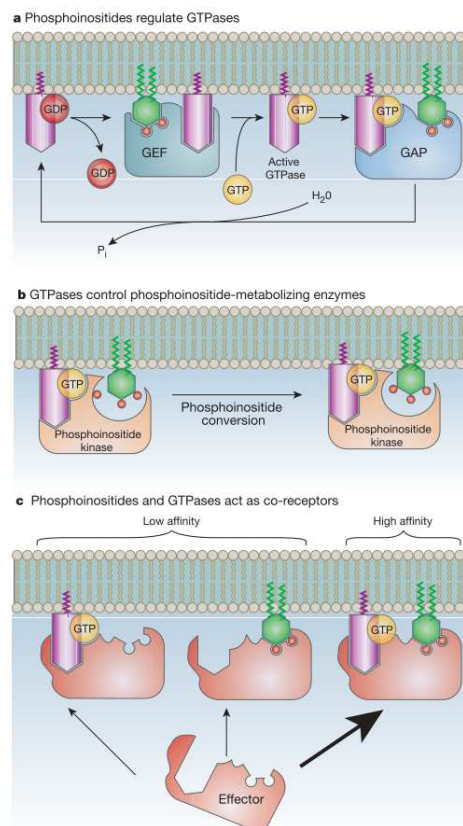
Figure 1.16 Interplay of phosphoinositides and small GTPases (Di Paolo, 2006). Purple - GTPases; Red – GDP; Yellow – GTP.

These processes can occur simultaneously forming a negative feedback loop. The conversion of phosphoinositides changes the membrane identity and can prepare suborganelle membrane domains or trafficking vesicles for fusion with the target membrane. Figure 1.16 shows the subcellular distribution of phosphoinositides and their metabolizing enzymes.

Due to their phosphoinositide affiliation, Rab and Arf GTPases are heterogeneously distributed in the cell. Just a subset of them is constitutively expressed, while the expression of others depend on the cell type or cell state (reviewed in Stenmark, 2009; Donaldson, 2011) (Figure 1.17). Two of the Arf GTPases, Arf1 and Arf6 have been studied particularly in

the context of coat formation. Arf1 localize at the Golgi, whereas Arf6 is predominantly found at the plasma membrane, but also occur in the endosomal system. Arf1 is required for the recruitment of the trafficking complex AP-1 (Stamness, 1993; Crottet, 2002) and Arf6 was shown to recruit AP-2 (Paleotti, 2005).
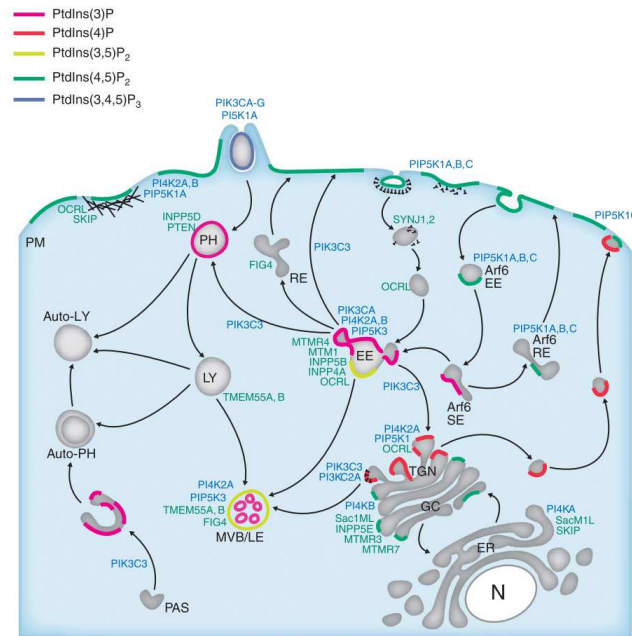
Figure 1.17: Subcellular distribution of phosphoinositides and their metabolizing enzymes (taken from Vivinanza, 2008). Blue names - phosphoinositide kinases; Green names - phosphoinositide phosphatases; PM - plasma membrane; EE - early endosome; SE - sorting endosomes; RE - recycling endosome; LY - lysosome; MVB/LE - multivesicular body/late endosome; PAS - pre-autophagosomal structure; PH - phagosome; TGN - trans-Golgi network; GC - Golgi complex; ER - endoplasmic reticulum; N - nucleus.

1.7.5 COPI, COPII complexes. Basic trafficking ways have been discovered that are hallmarked by the trafficking coat complexes involved. After synthesis and membrane-insertion in the rough endoplasmic reticulum, transmembrane proteins are trafficked to the cis-Golgi by vesicles coated with COPII. COPI coated vesicles are trafficking cargo either inside the Golgi or back to the endoplasmic reticulum. Post-Golgi trafficking vesicles are not coated with COPI/II but can be coated with clathrin instead.

1.7.6 Adapter complexes. At least six adapter complexes, AP-1A, AP-1B, AP-2, AP-3, AP-4 and AP-5, have been identified that serve on different post-Golgi trafficking routes. Adapter complex 1 or 2 link the clathrin coat to the vesicle. Several other trafficking complexes like AP-3 to 5, the retromer complex and a sequence of ESCRT-complexes ESCRT-0 to 3 do not seem to be coated. Clathrin has been implicated in retromer-mediated trafficking, but is shown to be not required for SNX-BAR-retromer-mediated carrier formation (McGough, 2013). Adapter complexes are heterotetramers, comprising of two large subunits (β 1–5, and either α , γ , δ , ε or ζ ; ~ 100 kDa), one medium (μ 1–5; ~ 50 kDa) and one small (σ 1–

5; ~ 20 kDa) subunit (Figure 1.17). Isoforms of subunits of the adapter complexes 1 to 3 have also been identified. The adapter complex 1 has two possible µ subunits, µ1 and µ2 and the resulting complexes are indexed as AP-1A and AP-1B. In epithelial cells, both complexes traffic cargo on different routes. AP-1A mediates bi-directional transport between the trans-Golgi network and the endosomes, while AP-1B conducts basolateral trafficking. AP-2 provokes clathrin-mediated endocytosis from the plasma membrane and AP-3 is thought to function in the endosome to lysosome or Golgi to lysosome transport. AP-4 may play a role in trafficking of proteins from the TGN to endosomes either directly or via the plasma membrane (Hirst, 2013). The trafficking routes and cargo of AP-5 that is believed to localize on late endosomal or lysosomal membranes remains elusive (Figure 1.18).
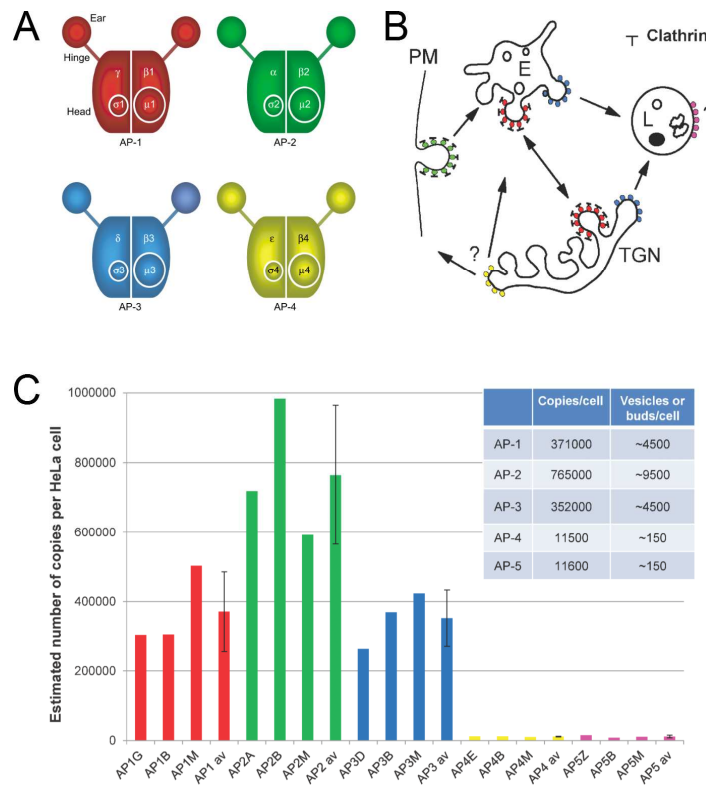
Figure 1.18: Trafficking by adapter complexes. A – Subunit structure of the adapter complexes 1-4 (taken from Robinson, 2001). B - Trafficking ways of the five adapter complexes. Red circles - AP-1; Green circles - AP-2; Blue circles - AP-3; Yellow circles - AP-4; Pink circles - AP-5; PM - Plasma membrane; E - Endosomes; L - Lysosomes; TGN - Trans Golgi network. C – Expression of the five adapter complexes in HeLa (B,C taken from Hirst, 2013).

1.7.7 Retromer complex. Retromer directs cargo from the endosomes to the Golgi, enabling recycling of receptors back to the plasma membrane. The retromer complex consists of two subcomplexes. It is at first, the cargo-selective core complex with VPS35, VPS29 and either VPS26A or B. Secondly, this core complex binds to sorting nexin subcomplexes that

recognize membranes by phosphoinositide-binding domains, like SNX1,2,5,6 (Figure 1.19). The involvement of clathrin in the retromer complex has been supposed (Popoff, 2007; McGough, 2013).
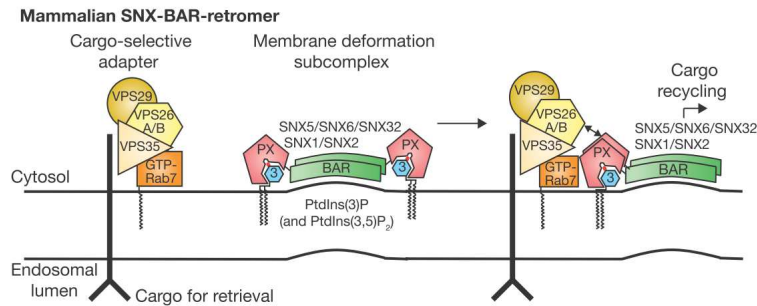
Figure 1.19: The retromer complex. The cargo-recognizing core of the retromer complex associates with sorting nexin subcomplexes that provide membrane-specificity (Cullen, 2011).

1.7.8 GGAs. Golgi-localized, γ-ear-containing, Arf-binding family proteins (GGA) are believed to be involved in the trafficking of mannose 6-phosphate receptors including their cargos, from the trans-Golgi network to early or late endosome. However, the fact that substantial amounts of GGAs localize with peripherally distributed structures can be explained by vesicular trafficking of GGAs to that sites (endosomes?) or GGA-mediated trafficking from
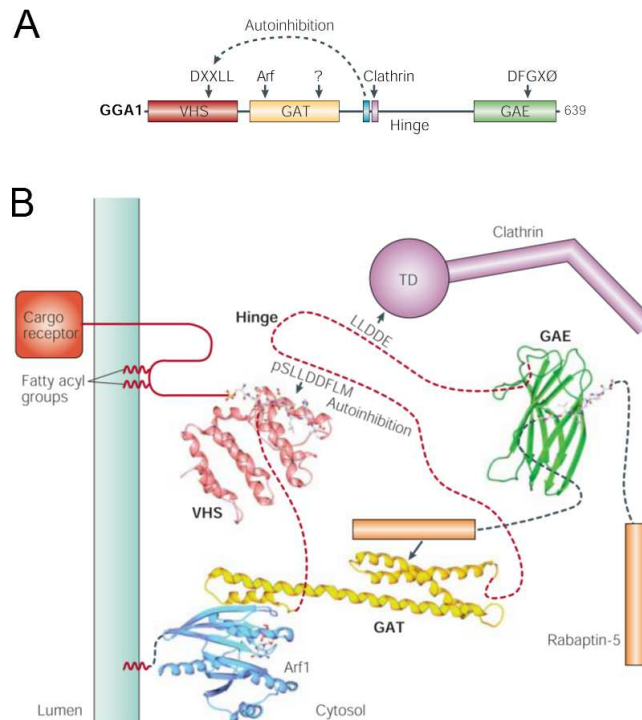
Figure 1.20: GGA coats. A - Domain organization of GGA1. B - Assembly of GGA-containing coats (Bonifacino, 2004).

that sites (Bonifacino, 2004). GGAs bind clathrin and they bind to the DXXLL sorting motifs of CI-MPR, CD-MPR, Sortilin, SorLA, LRP3, β-secretase and to other GGAs 1 or 3. They do also directly bind to Arf1, clathrin and accessory proteins, like rafaptin-5 (Figure 1.20).

Many more proteins that are capable of binding clathrin are recognizing membranes by phosphoinositides or Arfs, cargo sequences or ubiquitin were identified. Those proteins can have a similar central or an accessory role in a diversity of trafficking coats. Such proteins that localize at the plasma membrane are **epsin-1**, autosomal recessive form of hypercholesterolemia (**ARH**), huntingtin-interacting protein 1 (**HIP1**), disabled-2 (**Dab2**) and **numb** as well as hepatocyte-growth-factor-receptor substrate (**Hrs**) and signal-transducing adaptor molecule (**Stam**) on endosomes (Bonifacino, 2004).

1.7.9 ESCRT complexes. Parts of the plasma membrane along with membrane-associated cargo is internalized via endosomes and further internalized into the endosomes forming multivesicular bodies (MVB). These enclosed vesicles can subsequently be directed to the lysosomes or when MVBs fuse with the plasma membrane, be released from the cell as exosomes. Exosomal membrane proteomes differ in their composition from plasma membrane proteomes meaning that mechanisms exist that recognize cytosolic parts of membrane proteins and sort proteins for this route. The first step of these processes is the labeling of those membrane proteins with ubiquitin. Components of four ESCRT complexes (endosomal sorting complexes required for transport) (Table 1.2) recognize and cluster this cargo at the plasma membrane, provoke deubiquitination, sequestration and sorting (Figure 1.21).

Table 1.2: ESCRT complex components (Phyllis, 2009).

Complex	Yeast	Metazoan
ESCRT-0	Vps27	VPS27 (Hrs)
	Hse1	STAM1, 2 (HSE1, 2)
ESCRT-I	Vps23	VPS23 (Tsg101)
	Vps28	VPS28
	Vps37	VPS37A, B, C, D
	Mvb12	MVB12A, B
ESCRT-II	Vps22	VPS22 (EAP30)
	Vps25	VPS25 (EAP20)
	Vps36	VPS36 (EAP45)
ESCRT-III	Vps2	VPS2-1, 2 (CHMP2A, B)
	Vps24	VPS24 (CHMP3)
	Vps20	VPS20 (CHMP6)
	Vps32 (Snf7)	SNF7-1, 2, 3 (CHMP4A,B, C)
	Vps60 (Mos1)	VPS60 (CHMP5)
	Vps46 (Did2)	DID2-1, 2 (CHMP1A, B)
	-	CHMP7
Others	Vps4	VPS4A, B (SKD1)
	Vta1	LIP5 (VTA1)
	Ist1	IST1
	Vps31 (Bro1)	ALIX (AIP1)
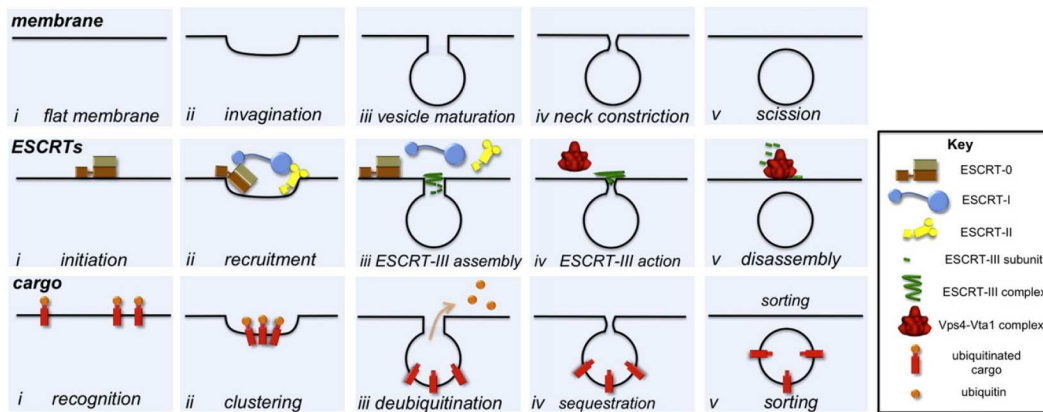
Figure 1.21: Sequential operation of ESCRT complexes (Henne, 2011).

1.8 Aim of the thesis

We want to establish label-free procedures that should enable us to answer the fundamental questions in proteomics: what is a sample's composition, is a group of proteins enriched that is defined by the sharing of specific properties and what are the differences between two (or more) samples. To do so, we will systematically compare primary label-free parameters and associated normalization methods for their dynamic and linear ranges and their accuracies to reflect protein amounts. For the enrichment analysis we will develop a mathematical model and for all the elucidated procedures we will implement bioinformatic workflows. To demonstrate the convenience of the workflows, we will apply this to two challenging types of biological models. It is at first, the comparative analysis of stem cell plasma membrane proteomes where we expect low sample amounts, small proteomic differences and arduous to analyze proteins in a complex chemical and proteomic background. At second, we want to study early trafficking events by the realization of proteo-liposome recruitment experiments. Hereby, membrane-associated protein networks of strong interaction partners as well as fragile protein complexes will be generated. For this purpose, we will develop a new recruitment assay approach enabling to couple bacterial expressed proteins to synthetic liposomes. These constructs will represent cytosolic parts of type-I and –II membrane proteins containing partially known trafficking motives. Those motives, presented in a lipid environment and exposed to a complex cytosolic proteome, will mimic early trafficking events by recruitment of proteins and protein networks. We expect a huge proteomic background of proteins binding to the liposomes irrespective of the receptor or simply co-isolated proteins and a relatively small amount and number of receptor-specific binders.

2. RESULTS

2.1 Label-free mass spectrometric parameters

2.1.1 Dynamic ranges. We were interested in establishing label-free mass spectrometric workflows, to estimate sample composition as well as the proteomic difference between two samples. At first, the dynamic and linear ranges of the basic mass spectrometric parameters spectrum count (SC) and intensity (MQ) were analyzed. As for the extracted ion intensity-based protein abundance index (xPAI) three different peptides are needed, this parameter was also included into the analysis. A dilution series of a mixture of albumin and beta casein was therefore analyzed (Figure 2.1) and the MaxQuant-Intensity (MQ) was found to provide superior dynamic and linear range over six orders of magnitude when compared to the mass input (Figure 2.2). The dynamic range of the spectrum count method was clearly inferior, with 5 orders of magnitude for albumin and 4 orders of magnitude for beta casein. For all three parameters, the curves converge and even superpose (for MQ) if they are compared to mass inputs instead of molar inputs. This suggests a general requirement of normalization for protein size. The measurements were repeated in presence of a yeast background proteome which affected the three parameters unequally. While in the spectral count method the detection of albumin seems unaffected, the detection of the smaller protein beta casein was reduced to the higher amounts where a dynamic dependency can even barely be visualized. This might be, because the upper limit of the SC method seems to be already achieved when first detecting casein. In the xPAI method, the detection of albumin still span over 6 orders of magnitude but having the detection of casein restricted to higher amounts. Clearly the outperforming parameter is MQ, with a linear range over five orders of magnitude for both proteins, while having both curves nearly superposed (Figure 2.3).

2.1.2 Accuracies. Secondly, the accuracies of normalized parameters in the estimation of protein abundances were observed on a broader spectrum of standard proteins. A tryptic digest of 49 proteins, presented in a range 0.5 to 50000 fmol (Universal proteomic standard UPS2) (Figure 2.5) was used therefore and the peptide numbers (PN), exponentially modified protein abundance indices (emPAI), spectral counts (SC), normalized spectral abundance factors (NSAF), MaxQuant-intensities (MQ) , label-free quantification intensities (LFQ), extracted ion intensity-based protein abundance indices (xPAI) and intensity-based absolute quantification values (iBAQ) were calculated. Both, emPAI and iBAQ require the number of observable tryptic peptides for normalization. Observability is defined by the size of the peptide, because the linear ion trap has to be operated in a certain mass range, it is defined by the capability to accept protons and to hold an overall charge of +2 or more, because of similar operative restrictions, and it is defined by a grade of hydrophilicity that allows binding to the reversed phase chromatography as well as its elution within the applied gradient. A total yeast proteome was analyzed and a python program was used, to compare experimental with theoretical retention times, and to spot retention time limits. Another python program was used then, to estimate numbers of observable tryptic peptides

from those limits, and a fasta-format data base of the respective organism, to carry out an in-silico digestion with trypsin. Other parameters were the charge (+2 or +3), number of amino acids (6 to 60) and the type of gradient (reversed phase, pH=3, with peptides eluting in between 5 and 35% acetonitrile) (Figure 2.4).
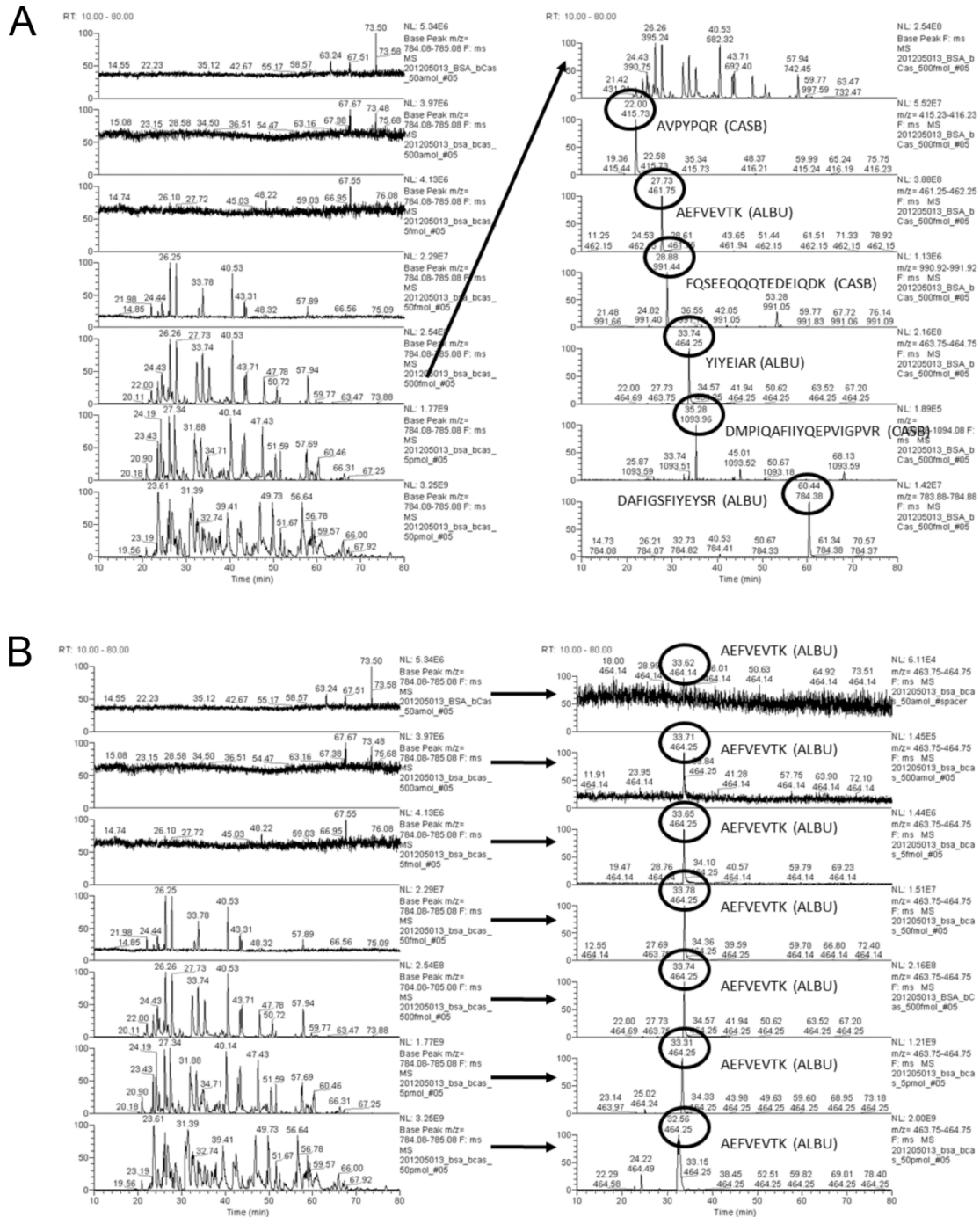
Figure 2.1: Assessment of the dynamic range in protein quantification. A & B – Left panels: base peak chromatograms for different amounts of a mixture of tryptic peptides of albumin and beta casein; A - Right panel: base peak chromatogram and extracted ion chromatograms of tryptic peptides from a mixture of 500 fmol tryptic peptides of albumin and beta casein; B – Right panel: extraced ion chromatograms for an albumin peptide across all dilution steps. This peak can be identified in the extracted ion chormatograms, even in experiments were there is no clear peak pattern in the base peak chromatograms.
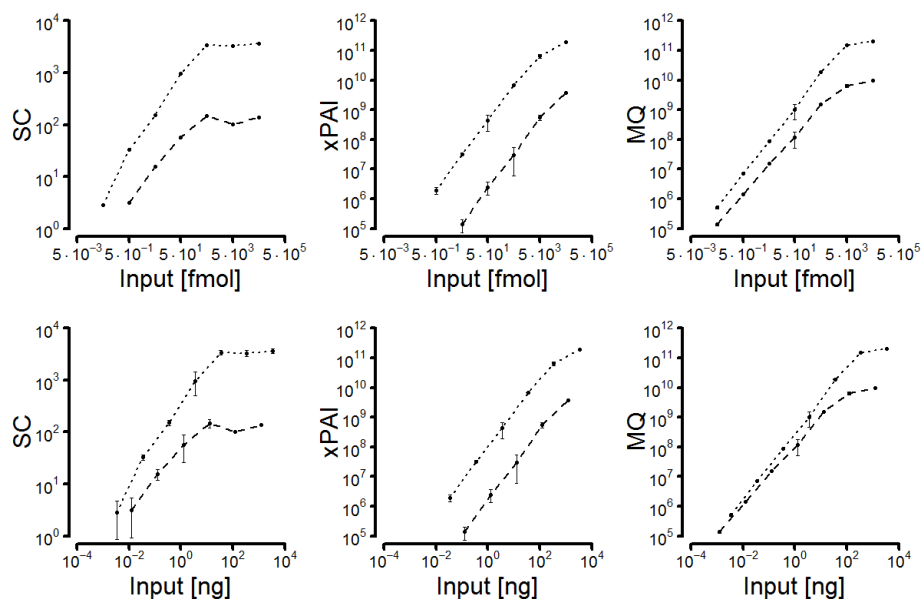
Figure 2.2: Dynamic range of quantitative values. Two-component mixtures of different amounts of albumin (dotted) and beta casein (dashed) were trypsin-digested and analyzed via LC-MS. Highest dynamic range over five orders of magnitude, along with a strong linearity, was achieved by the use of MaxQuant intensity values (MQ), and similar good by using the extracted ion chromatogram-based protein abundance index (xPAI). The spectral count index (SC) was restricted to a dynamic range over three orders of magnitude, in case of the smaller beta casein. Using mass amounts on the abscissa, normalizes the data for protein length and converges both proteins' curves.
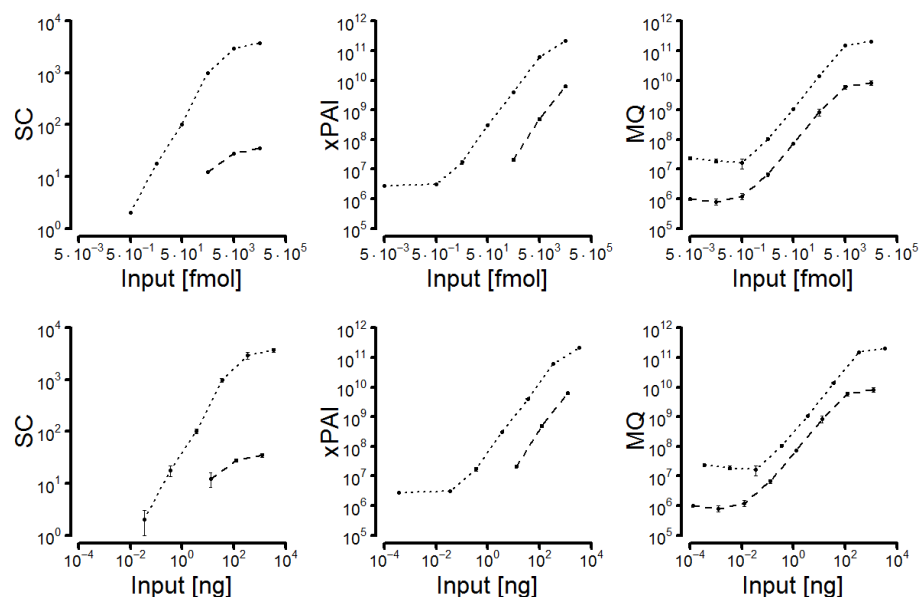
Figure 2.3: Dynamic range in a complex proteomic background. Albumin (dotted) and beta casein (dashed) were spiked in different amounts into a total yeast proteome, trypsin-digested and analyzed via LC-MS. Sample complexity slightly reduced the dynamic range of all values. The MaxQuant intensity performed best, in terms of dynamic range and linearity over four orders of magnitude.
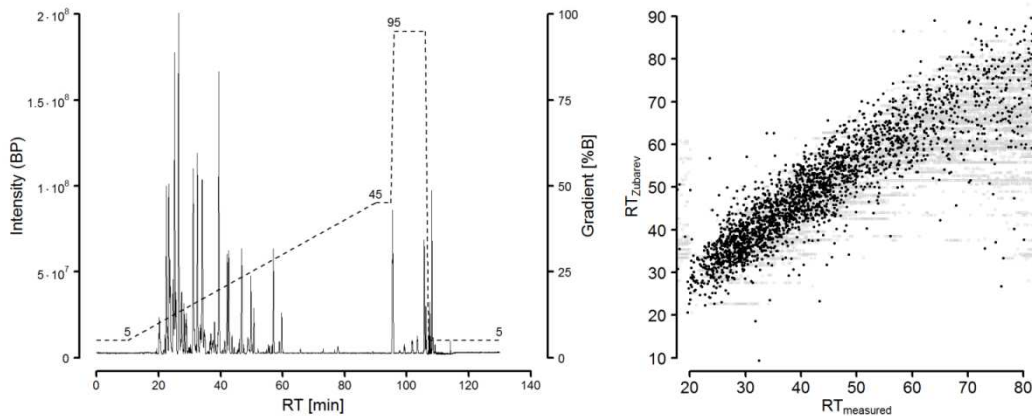
Figure 2.4: A - LC-gradient and base peak chromatogram (BP) as used in all experiments. The majority of peptides elute within 40 minutes of the gradient (RT = 20...60 min). B - Estimation of RT-limit presets. The calculation of theoretical observable peptides is a prerequisite for the calculation of emPAI and iBAQ. All peptides of a tryptic in-silico digest are filtered first for those observable in the operation mass over charge range of 200 to 2000 amu with allowed charges of 2+ and 3+. Extremely hydrophilic (at pH=3) or hydrophobic peptides can also not be observed because they either not bind to the LC-columns or could not be recovered from the bead or other surface materials. To elucidate those the method of Zubarev (Gorshkov, 2006) is used to calculate theoretical retention times and a filtering window is constructed by comparison of those theoretical with practical retention times within the gradient of a pre-experiment. Grey circles – all redundant peptides measured. Black dots - highest scoring peptide identification. Whereas most of the peptides elute in between 20 and 60 minutes, above 60 min grey horizontal lines indicate an increasing repetitive measurement. This corresponds to the lack of peaks in the base peak diagram. Hence, there is some optimization potential in altering the classical linear 5 to 45 percent acetonitrile gradient to a sophisticated non-linear concave gradient.

The quantitative indices in Figure 2.5 differ not only in the overall correlation (correlation coefficients), but also in the magnitude of single outliers which would lead to extremely false estimations of single protein abundances. This is particularly the case for emPAI and PN, but can also be seen for LFQ were a single protein is underestimated by two orders of magnitude. PN and emPAI proved to correlate absolutely poor when no manual data deselection is allowed. All other parameters seem to be generally suitable for assessing proteome compositions. Interestingly, the magnitude of single outliers is diminished when the parameters are compared to the mass inputs, rather than to the common molecular inputs, enlighten another level of normalization. Also, the overall correlation is enhanced in that way. The two predominating indices are LFQ and iBAQ. The latter appear as an almost ideal index for protein abundance in a sample, with low standard deviations of repeated measurements, a high correlation coefficient over a huge dynamic range and outliers below one order of magnitude. LFQ on the other hand, demonstrated to be reliable as well and as it is constructed as normalization between samples, these LFQ values were subsequently assessed for its use in the comparison of different samples.
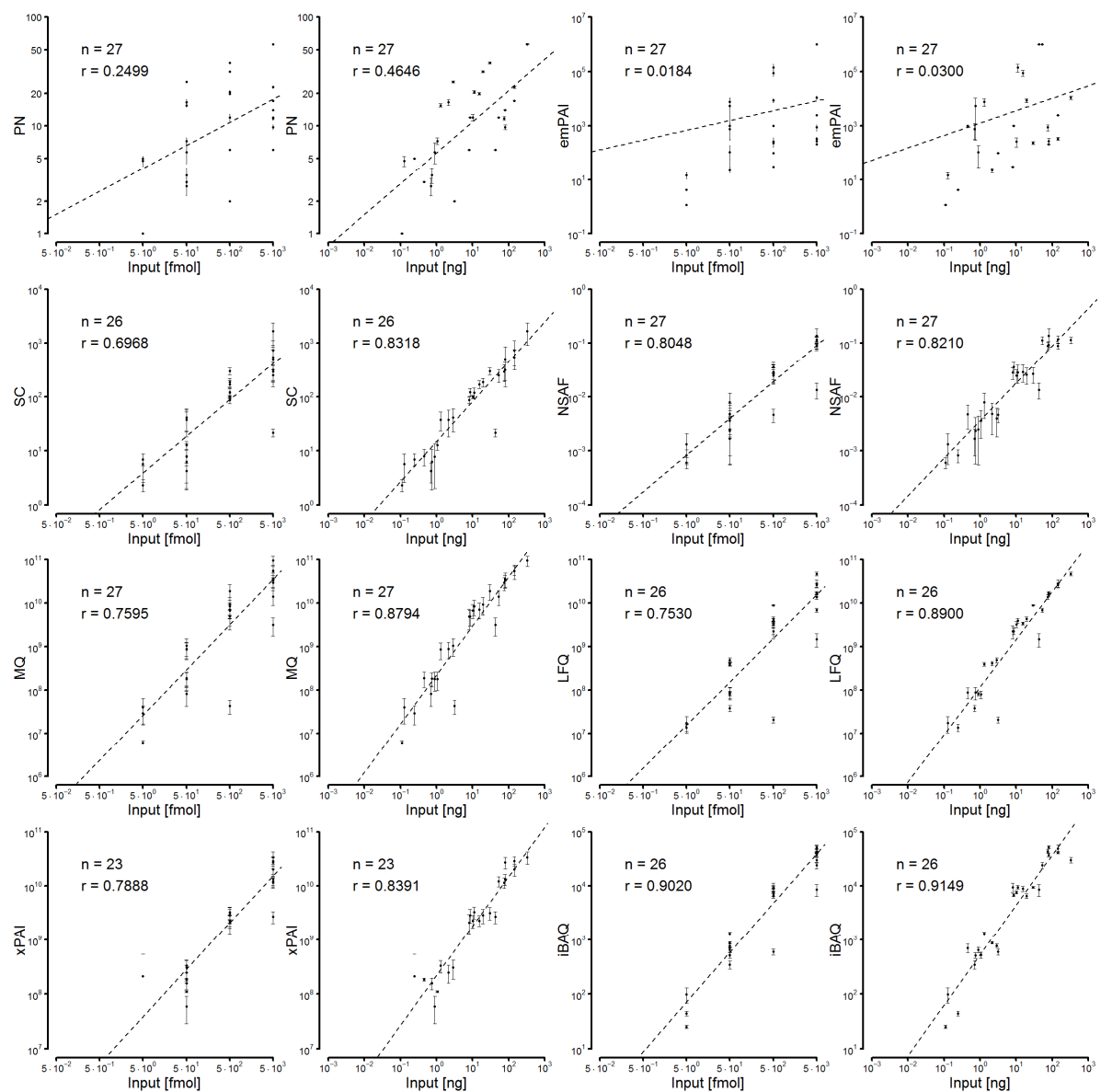
Figure 2.5: Accuracy of protein amount estimations. Different normalized parameters were compared for their linearity to a broad range of spike-in amounts of Universal proteomic standard 2 (UPS2) proteins. PN and its normalized derivative emPAI are highly affected by large outliers. SC, its normalized derivative NSAF and all intensity based parameters are generally suited for quantitative analyses. All parameters show slightly better Pearson correlation coefficients when been opposed to the mass inputs suggesting another level of normalization. Some proteins deviate largely from the regression lines, hampering confident estimation of single protein amounts whereas assumptions over group contingents, like proteins sharing a certain gene ontology, seem to be justified. This was not the case for iBAQ values, which suited best also in overall correlation. Deviations of single values were here less than one order of magnitude from the regression line. LFQ intensities were only slightly less accurate and as they are constructed for normalization between samples, were thus further used for differential proteome analyses.

2.2 Differential proteome analysis

2.2.1 Imputation. We assume that for proteins, which were not detected in all the triplicates of at least one sample group (i.e. experimental state), there is no significant proof of their principal presence and hence they were removed from the data. In all other cases, we imputed missing values under the assumption that they were present below the detection limit. However, this assumption is not obvious for proteins with a low number of observable tryptic peptides where ion suppression could have led to missing identifications. To keep the overall standard deviation for protein intensities from the low intensity region of the data, we refrained from imputation using a constant. Instead, intensity values for imputation were randomly drawn from a Gaussian distribution around the detection limit, with a standard deviation estimated from the low intensity region of the data (Figure 2.6).
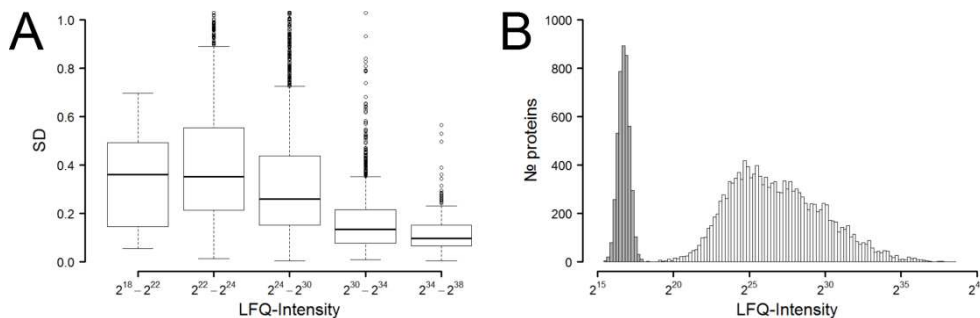
Figure 2.6: A - Standard deviation (SD) in units of $\log_2$(LFQ-Intensity), versus LFQ-Intensity classes in a boxplot analysis. Low intensities are generally poorer estimated than higher intensities, as demonstrated by their higher standard deviations. One major reason for this might be an enhanced risk for ionic suppression. (Data were taken from an APP-recruitment experiment in a PI345P3 membrane context). B – Original (light bars) and imputed (grey bars) LFQ-values in a histogram plot. Imputation values were randomly drawn from a Gaussian distribution around the detection limit, with a standard deviation estimated from the low intensity region of the data (same data set as in Figure 2.5).

2.2.2 Accuracy of protein change detection. Next, the universal proteomic standard 1 (UPS1) consisting of of 49 equimolar proteins was spiked into a yeast background proteome in a series of different amounts (Figure 2.7 and 2.8). Four technical replicates of all such samples were analyzed via LC-MS and protein identification was accomplished using MaxQuant. LFQ values were $\log_2$-transformed, and proteins were eliminated if not at least 3 values out of four replicates were achieved. Missing values were imputed with a normal distribution around the detection limit. A permutation based modified t-test, similar to a significance analysis of microarray experiments (SAM) (Tusher, 2001), was conducted using the MaxQuant Perseus software, with a threshold value of 0.01 and a slope value of 1.0. No protein was found to be significantly different when 0.74 versus 0.24 fmol UPS1 proteins were compared. This is, because with an input of 0.24 fmol, it was impossible to yield three LFQ values for anyone protein. As expected, the number of quantifiable proteins increases with the amounts spiked in, but also the accuracy of ratio estimation increases with the

actual ratio, but not with increasing input amounts. No saturation of this increase was reached within the experiments. The method elucidates even small changes in a complex background proteome. High accuracy is reached for high input ratios, in contrast to SILAC where the impossibility of a 100% labeling efficiency leads to great underestimation of high ratios.
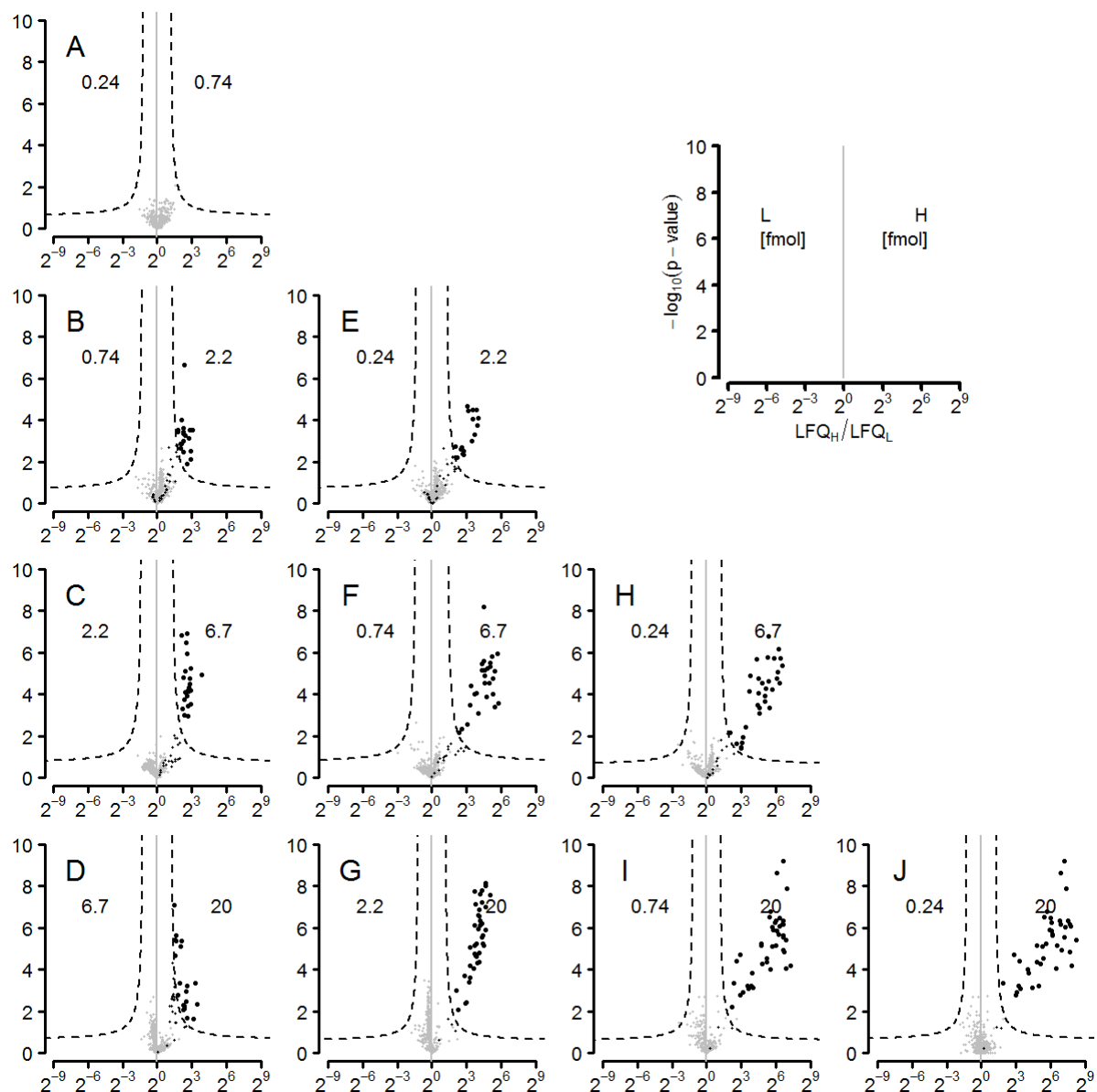
Figure 2.7: Volcano plots for differential proteome analysis. Different amounts of UPS1 standard proteins (black dots) were spiked into a total yeast proteome (grey dots). To assess the models response to protein abundance changes, four replicates of two of these samples were compared in each volcano plot. The dashed lines represent significance thresholds (TV=0.01 and s_0 = 1.0). Numbers inside the plot indicate spike-in amounts of UPS1 that are compared. H, L – High and low amounts of spiked-in UPS1 standard proteins.

Normal distributed data are preferred, and sometimes are even a prerequisite for statistical models of distributions. The general notion is that log-transformed abundances are approximately normal (Podwojski, 2010). To check if this is also the case for the normalized intensities that due to the preceding experiments reflect protein abundances, LFQ and iBAQ intensities from the yeast background proteome were compared to a normal distribution in a quantile-quantile (Q-Q) plot (Figure 2.9). iBAQ intensities prove to be perfectly normal distributed and the LFQ value can be represented by a normal distribution as well, although there is a certain deviation from the intersecting line for extreme values.
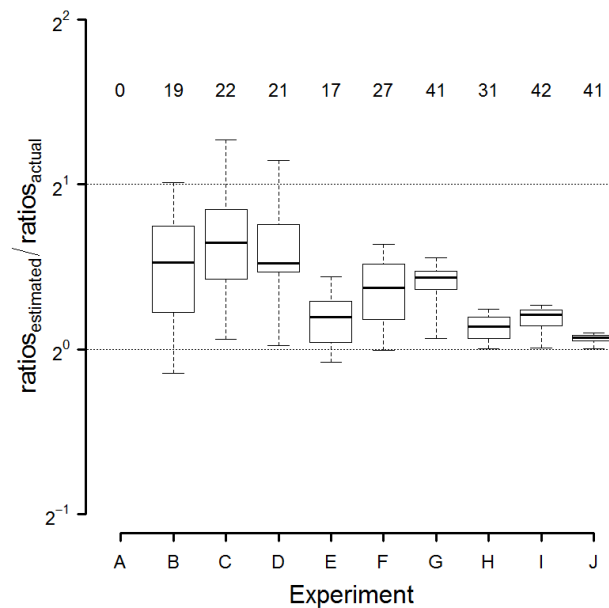
Figure 2.8: Boxplots of ratios of UPS1 proteins, normalized by the actual input ratios. The experiment labels refer to Figure 2.7. Numbers inside the plot assign the number of proteins passing the significance thresholds. The accuracy increases with higher input ratios, but not with increasing input amounts. Even small changes in a complex background proteome are elucidated.
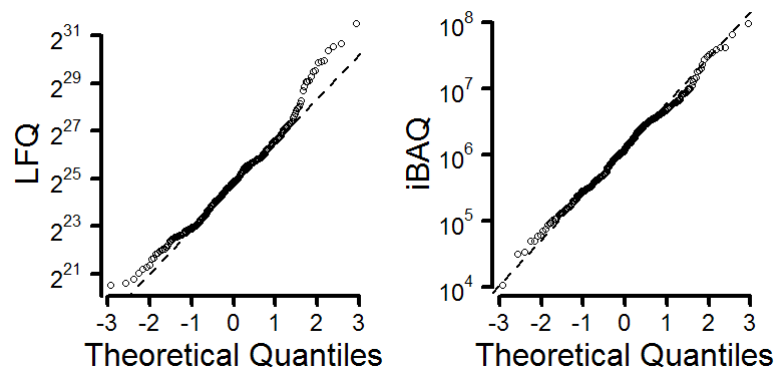
Figure 2.9: A trypsin-digestion of a total yeast proteome was used to mine MaxQuant label-free quantification intensity values (LFQ) for a comparison to theoretical quantiles in a Q-Q plot. Log-transformed LFQ-intensities proved to be normal distributed in a complex sample, and so were the protein amounts due to their demonstrated strict linear correlation to MQ/LFQ values.

2.2.3 Receiver Operating Characteristic. The choice of a threshold value and the slope parameter which has a minor influence on the curve is not obvious. The SAM approach was initially developed to resolve the issue of multiple testing (Roxas, 2008) and the threshold was designed to reflect the false discovery rate (FDR) (Hubner, 2010). However, we recognized that this doesn't hold true for proteomics data. To test the robustness of

Table 2.1: ROC-Analyses of the modified t-test (SAM), as implemented in PERSEUS. A - Experiment D (20 fmol versus 6.7 fmol UPS1 in a yeast background proteome); B - Experiment B (2.2 vs. 0.74 fmol UPS1)

A	TV	T	P	D^+	D^-	TP	FP	TPR	FPR
	0.001	477	45	11	466	11	0	0.244	0.000
	0.01	477	45	25	452	25	0	0.556	0.000
	0.02	477	45	34	443	34	0	0.756	0.000
	0.03	477	45	38	439	38	0	0.844	0.000
	0.04	477	45	42	435	41	1	0.911	0.002
	0.05	477	45	42	435	41	1	0.911	0.002
	0.06	477	45	42	435	41	1	0.911	0.002
	0.07	477	45	42	435	41	1	0.911	0.002
	0.08	477	45	42	435	41	1	0.911	0.002
	0.09	477	45	43	434	42	1	0.933	0.002
	0.10	477	45	43	434	42	1	0.933	0.002
	0.15	477	45	44	433	42	2	0.933	0.005
	0.20	477	45	44	433	42	2	0.933	0.005
	0.25	477	45	44	433	42	2	0.933	0.005
	0.30	477	45	45	432	42	3	0.933	0.007
	0.40	477	45	52	425	43	9	0.956	0.020
	0.50	477	45	53	424	43	10	0.956	0.023
	0.60	477	45	55	422	44	11	0.978	0.025
	0.70	477	45	137	340	44	93	0.978	0.177

B	TV	T	P	D^+	D^-	TP	FP	TPR	FPR
	0.001	477	45	15	462	15	0	0.333	0.000
	0.01	477	45	23	454	23	0	0.511	0.000
	0.02	477	45	23	454	23	0	0.511	0.000
	0.03	477	45	23	454	23	0	0.511	0.000
	0.04	477	45	24	453	24	0	0.533	0.000
	0.05	477	45	24	453	24	0	0.533	0.000
	0.06	477	45	24	453	24	0	0.533	0.000
	0.07	477	45	24	453	24	0	0.533	0.000
	0.08	477	45	24	453	24	0	0.533	0.000
	0.09	477	45	24	453	24	0	0.533	0.000
	0.10	477	45	26	451	25	1	0.556	0.002
	0.15	477	45	32	445	30	2	0.667	0.004
	0.20	477	45	32	445	30	2	0.667	0.004
	0.25	477	45	33	444	30	3	0.667	0.007
	0.30	477	45	37	440	31	6	0.689	0.014
	0.40	477	45	43	434	31	12	0.689	0.027
	0.50	477	45	55	422	32	23	0.711	0.051
	0.60	477	45	165	312	32	142	0.711	0.247
	0.70	477	45	251	226	38	213	0.844	0.330

TV – Threshold value of the SAM-threshold function (s_0 = 0.5); T - True positives, the number of identified proteins (Yeast and UPS1); P - Positives, all identified among the 48 UPS1 proteins; D^+ - The number of proteins that are positively tested; D^- - The number of proteins below the threshold; TP - True positives, the number of UPS1 proteins among the proteins that passed the threshold; FP - False positives, the number of UPS1 proteins that were identified, but below the threshold; TPR - True positive rate, the number of true positives with respect to all positives (all identified UPS1); FPR - False positive rate, the number of false positives with respect to all negatives, which is the sum of true negatives and false positives.
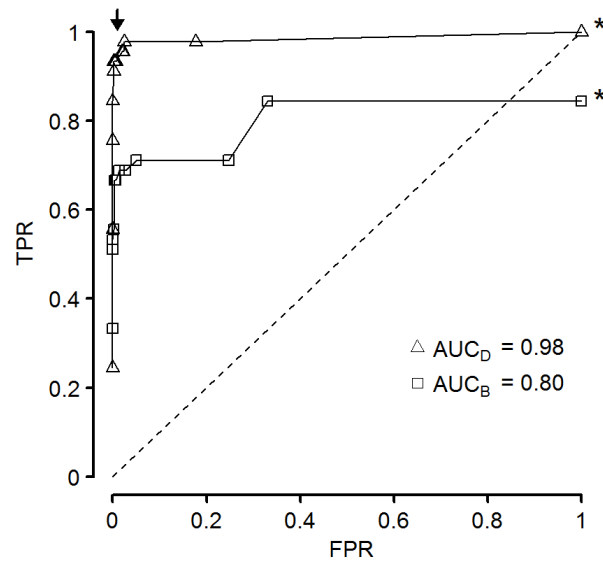
Figure 2.10: ROC-Analysis of experiments D and B (Indices correspond to Figure 2.7). Threshold values (TV) were varied from 0.001 to 0.7, with a constant slope value $s_0 = 0.5$. By increasing TV, the true positive rate (TPR) drastically increases with nearly no false positives (FPR - False positive rate) up to a TV of 0.25 (arrow) where FPR < 0.01. By further increasing the TV, only the FPR is increased. AUC - Area under the curve; Stars - Approximated points.

the assay for different threshold settings and the accuracy of the method, ROC-analysis (Receiver Operating Characteristic) was performed on two of the experiments, with UPS1-ratios of 3 (Table 2.1, Figure 2.10). The true positive rate (TPR) was calculated with respect to the 45 of the 48 UPS1 proteins which were identified in the overall analysis. The difference between both experiments was that although in both cases UPS1 ratios of three were analyzed, in experiment D both samples had 9-fold bigger UPS1 amounts as the respective samples in experiment B. In the ROC-analysis, the TRR is compared to the false positive rate (FPR) reflecting yeast proteins that wrongly passed the threshold. If pairs of yeast and UPS1 proteins (i.e. non differentially and differentially abundant proteins) are randomly drawn and analyzed, the area under the curve (AUC) in a ROC-analysis represents the percentage of pairs were the method correctly classifies the proteins. The AUCs in Figure 2.10 denote an excellent accuracy of experiment D and a good accuracy of experiment B (0.9-1: excellent; 0.8-0.9: good; 0.7-0.8:fair; 0.6-0.7: poor; 0.5-0.6: fail). This is, because the true positive rate in experiment B was also estimated with respect to the 45 UPS1 proteins identified across all experiments. One characteristic of the workflow is that MaxQuant interpolates peptide fragment information and hence protein identifications to samples processed in parallel were fragmentation was not triggered. Although 45 proteins were identified in such way, not all of them were in the dynamic range window and hence the TPR-maximum in experiment B is below 1.0. Beside the high accuracies to detect differentially abundant proteins, the curves reveal a high robustness with respect to the threshold values. In both cases the threshold values can be set as high as 0.25 before the false positive rate exceeds 1%. However, these analyses were performed in an ideal

situation with a very low variability in protein amounts of both, standards and background. Depending on the type of biological experiment, different grades of variability in levels of proteins were achieved within the replicates. Manual inspection of the proteins surpassing the threshold, let us choose threshold values far below 0.25 thus, i.e. 0.001 to 0.05. If doubts persisted that some proteins passing the threshold were representing true positives, the number of replicates was increased.

2.3 Application I – Analysis of plasma membrane proteins

Compositional proteomic analysis, enrichment analysis and differential proteomic analysis of membrane proteins before and upon biochemical enrichment

Stem cells possess unique proliferation and differentiation capabilities. They demonstrate prolonged self-renewal power, in contrast to mature (specialized) cells that have a more limited capability of dividing. All cells of the body can be generated from stem cells and the differentiation potential of a particular linage is classified into either totipotent, pluripotent or multipotent stem cells. Totipotent stem cells deriving from fertilized eggs can differentiate into a placenta and an embryo. Pluripotent stem cells can differentiate into any cells derived from three embryonic layers: mesoderm, endoderm and ectoderm, whereas multipotent stem cells are progenitors of a closely related lineage of cells (Lin, 2008). With the immanent ethic controversy in the use of embryonic stem cells, adult stem cells are a promising alternative source of stem cells for research and therapeutic applications. Two types of stem cells can be isolated from bone marrow, hematopoietic and mesenchymal stem cells. An additional multipotent mesenchymal stem cell population has been isolated from the dental pulp and termed dental pulp stem cells (DPSC) (Gronthos, 2000). These cells are capable of differentiation into chondrogenic, endothelial, osteogenic, neuronal and smooth muscle linages. Traditional isolation procedures for dental pulp stem cells do not appear much sophisticated and hence indeed lead to heterogeneous subpopulations. It is at first, a procedure based on their high proliferation capacity, in which single cells are separated and the clones with the highest growth rate are selected (Kawashima, 2012). A second approach uses the plastic adhesion characteristic of DPSC cells to isolate those. A more advanced approach is based on the presence (or absence) of specific cell surface markers used for fluorescence-activated cell sorting. Once DPS cells are isolated, they need to be cultivated. Karbanová et al. (Karbanová, 2010) showed that the chosen cultivation medium has an influence on the cell surface proteome. The work presented here is a continuation of the work of Dr. Jana Karbanová in collaboration with her in the group of Dr. Denis Corbeil (Biotec, Dresden). The depicted workflow is aimed to elucidate appropriate markers for a clear definition of a DPSC population and its isolation by FACS. The described cell surface proteome differences, caused by the cultivation media, should be analyzed on a molecular level.

Dental pulp stem cells were isolated and cultivated either in 10% fetal calf serum (FCS) 'standard medium' (as described in Karbanová, 2010) or in 'basic expansion medium', containing 2% FCS and epidermal growth factor (EGF) as well as platelet-derived growth factor BB (PDGF-BB) as supplements (Karbanová, 2010). Figure 2.11 illustrates the sample types that were generated for LC-MS analyses. Proteome compositions and the proteome differences, caused by the cultivation media, were investigated by the newly implemented proteomic workflows.
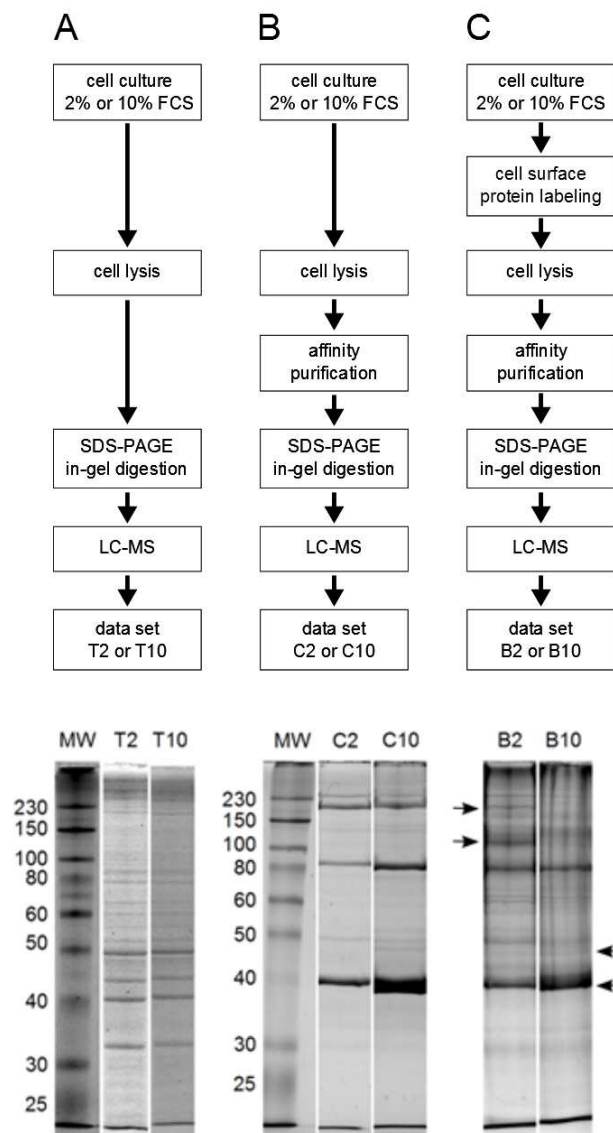
Figure 2.11: Experimental workflow. Dental pulp stem cells were cultivated either in 2% fetal calf serum, with PDGF-BB and EGF additives, or in 10% fetal calf serum (indices 2 or 10). A - Total proteome (samples T2 and T10). B - Control samples without labeling, but with affinity purification (samples C2 and C10). C - Cell surface proteome-enrichment with biotinylation and affinity purification (samples B2 and B10). Gel lanes from the total proteome samples were indistinguishable by their band patterns. Gel lanes of the controls exhibit small differences. Some obvious differences in the gel band patterns of cell surface proteome-enriched samples are indicated by arrows.

2.3.1 Enrichment analysis. The cell surface proteome-enriched samples were compared to the total proteome samples to assess the effectivity of the lysine-directed biotinylation and affinity chromatography enrichment procedure, which is one of three major CSP enrichment methods. The two other methods are density centrifugation, where we expect largest numbers of contaminant proteins due to the similar densities of plasma membranes, endoplasmic reticulum and mitochondria, and hydrazide capturing of reduced N-glycosidic carbohydrate side chains, which is claimed to be nearly free of false positives. Within the presented mass spectrometric workflow, we take advantage of the presence of a certain

degree of proteomic background, as this is used for the calculation of deviations and normalization. This is why the lysine-biotinylation method had been chosen.

We were comparing different methods to access and plot enrichment efficiency. Pie-charts are widely used to give an overview about sample composition by partitioning the samples according to the gene ontology classifications (GO) (http://www.geneontology.org). The total proteome, cell surface proteome and control samples are nearly indistinguishable in pie-charts (Figure 2.12). Several problems are associated with this representation. Gene

Figure 2.12: Pie-diagrams showing the grade of gene ontology assignments within the samples. Samples are partioned by the grade of gene ontology assignments.

ontologies are not exclusive, i.e. proteins can be part of more than one of the gene ontologies used in the diagram. This necessitates a selection of gene ontologies ensuring a minimum of such overlaps. Pie-chart representations do not utilize intensity values of proteins, and hence ignore protein abundances. The diagram resembles a percental representation making it difficult to compare samples of unequal sizes.

Venn-diagrams as in Figure 2.13 are popular to compare samples in a certain aspect, such as gene ontology. This diagram type circumvents the problems with different sample sizes and gene ontology overlap. As in the pie-charts the biotinylated samples as well as the total proteome samples show 27-30% cytosol annotation, comparing the numbers of proteins by Venn-diagrams elucidates a larger difference: 10% of all proteins were identified in the cell surface proteome only, whereas 43% were exclusively identified in the total proteome. The overall overlap is 40% and slightly more proteins (34%) were solely identified in the total proteome samples than were solely identified in the cell surface proteome sample (26%). About three times more proteins are assigned to the plasma membrane solely in the cell surface proteome sample than in the total proteome. Also for organelle membrane proteins, there are roughly one third more proteins unique in the cell surface proteome samples than in the total proteome samples. The question remains which of these differences are obvious, and also protein abundances are not utilized in Venn-diagrams as well. In high sensitivity and high accuracy mass spectrometry, huge numbers of proteins are identified and enrichment is less effecting the identification of proteins as their intensities.
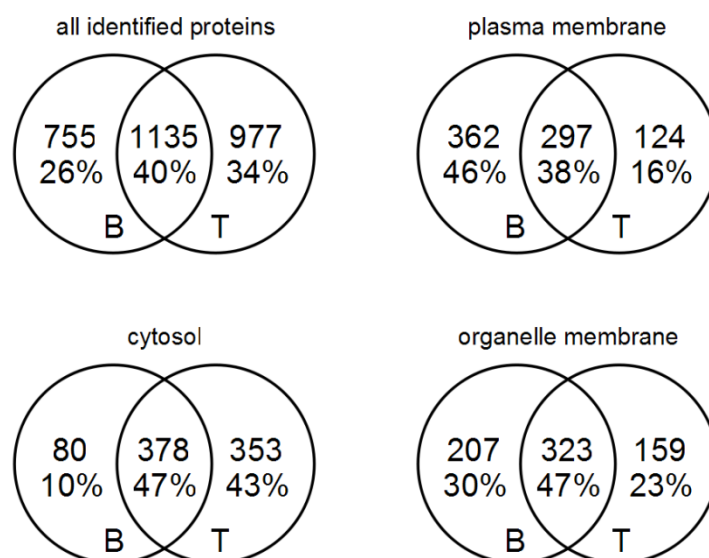
Figure 2.13: Venn-diagrams of cell surface proteome-enriched (B) and total proteome (T) samples. Around 40% of the 2867 identified proteins were found in both types of samples. The cell surface proteome-enriched sample exhibit an appr. three fold higher number of proteins with a gene ontology assignment to plasma membrane localization (GO:0005886), whereas the total proteome sample contain a more than four fold number of cytosol localizing proteins (GO:0005829). Proteins assigned to organelle membranes (GO:0031090) are not preferentially enriched in either sample.

Histogram diagrams as shown in Figure 2.14 can be used to embed intensity values in enrichment analysis by plotting the number of proteins in small intensity regions ('bins'). If the mean of a subpopulation of proteins sharing a certain gene ontology is greater than the overall mean, those proteins are stated to be enriched by Pan et al. (Pan, 2009). However, when proteins are enriched compared to a total proteome, those means not necessarily differ in the final sample. What can differ in addition, is also the area under the curve fraction of the subpopulation meaning that the Pan and Cox (Cox, 2012) methods ignore a relative increase in the number of different proteins assigned to a specific gene ontology.
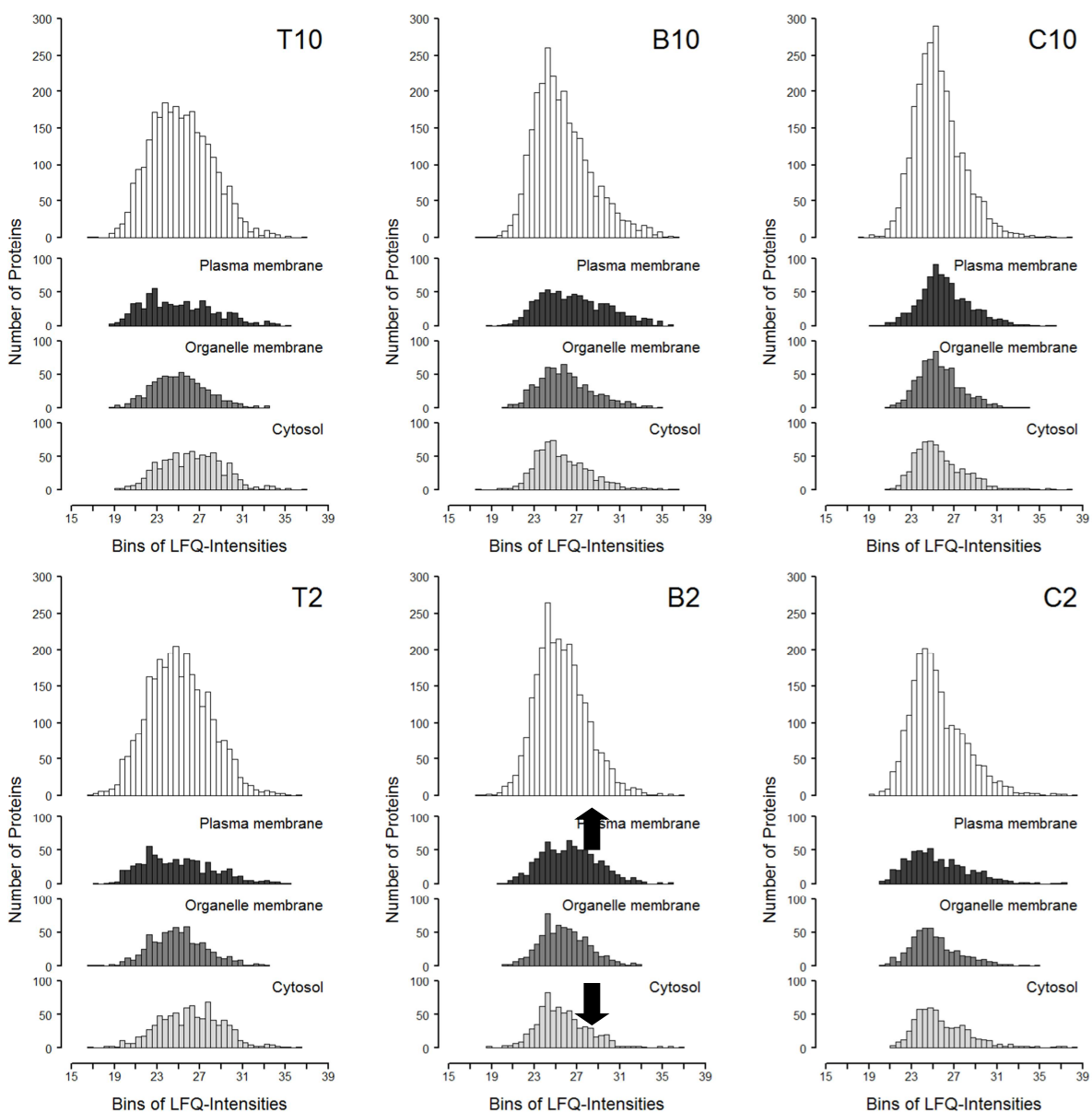
Figure 2.14: Histogram plots reflecting the number of proteins in small intensity-ranges. Distributions of protein subpopulations defined by gene ontology assignments barely differ in their means from the overall mean.

Figure 2.14 demonstrates that the means of all subpopulations, i.e. proteins with gene ontology assignments to plasma membrane, organelle membrane or cytosol, do not clearly differ from the overall means in any of the samples. What can be seen are slight redistributions in the intensity bins. As an example, a loss of proteins can be seen in the B2 bins 2^{27}-2^{31} when compared to the respective T2 bins, whereas plasma membrane proteins were added (arrows).

In our method, proteins are ranked by intensity before binning and the number of proteins classified to a specific gene ontology in a bin is then compared to a random draw by a hypergeometric distribution model. The resulting p-values reflect the possibility of receiving that number of specific proteins by chance. The negative decadic logarithm of the p-value is used to express enrichment thus and can be seen as a rescaling of the number of positive draws. This rescaling is a normalization of the respective gene ontology for its sparsity. The total number of proteins in a bin is always the same here (100), in contrast to the histogram plots. Figure 2.15 shows the application of the model to the experimental data. In the cell surface proteome samples, plasma membrane proteins accumulate in the first (high intensity) bins, whereas organelle membrane and cytosolic proteins are equally distributed over all intensities. In the total proteome samples, cytosolic proteins dominate in the higher intensity bins on the contrary and plasma membrane proteins and organelle membrane proteins are equally distributed. In the controls, proteins of all gene ontologies are equally distributed on a lower level. All three sample types clearly differ in this analysis. Enrichment of cell surface proteins seem to be effected by an increase of the number of plasma membrane proteins in high intensity bins and a simultaneous decrease of cytosolic proteins in those.

In the plots of Figure 2.15, redistributions could be overseen if they occur less drastic so that there is a smaller shift in intensity of the respective proteins. This would reposition the bars only slightly in proximity. As the scaling on the x-axes are always the same (bins of 100 intensity-ranked proteins), two plots can be combined into one comparing the p-values of the corresponding bins. This plot type, which is hereby named P-P plot, is applied in Figure 2.16 to the data of both cultivation media. Significance regions can be defined where high differences of enriched and total proteome samples get revealed by high differences in the p-values of the respective bins and gene ontologies. For the cell surface proteome enriched samples, those districts contain only data points representing plasma membrane protein enrichment, as on the other hand for total proteome samples, only cytosolic proteins appear to be enriched. The P-P plot illustrates all information to judge success of an enrichment procedure in a single mathematical analysis, numbers of proteins that are shifted into or out of the observation window as well as proteins that are decreased or increased in numbers but were always observed. Gene ontologies are normalized herein to insure their explanatory power.
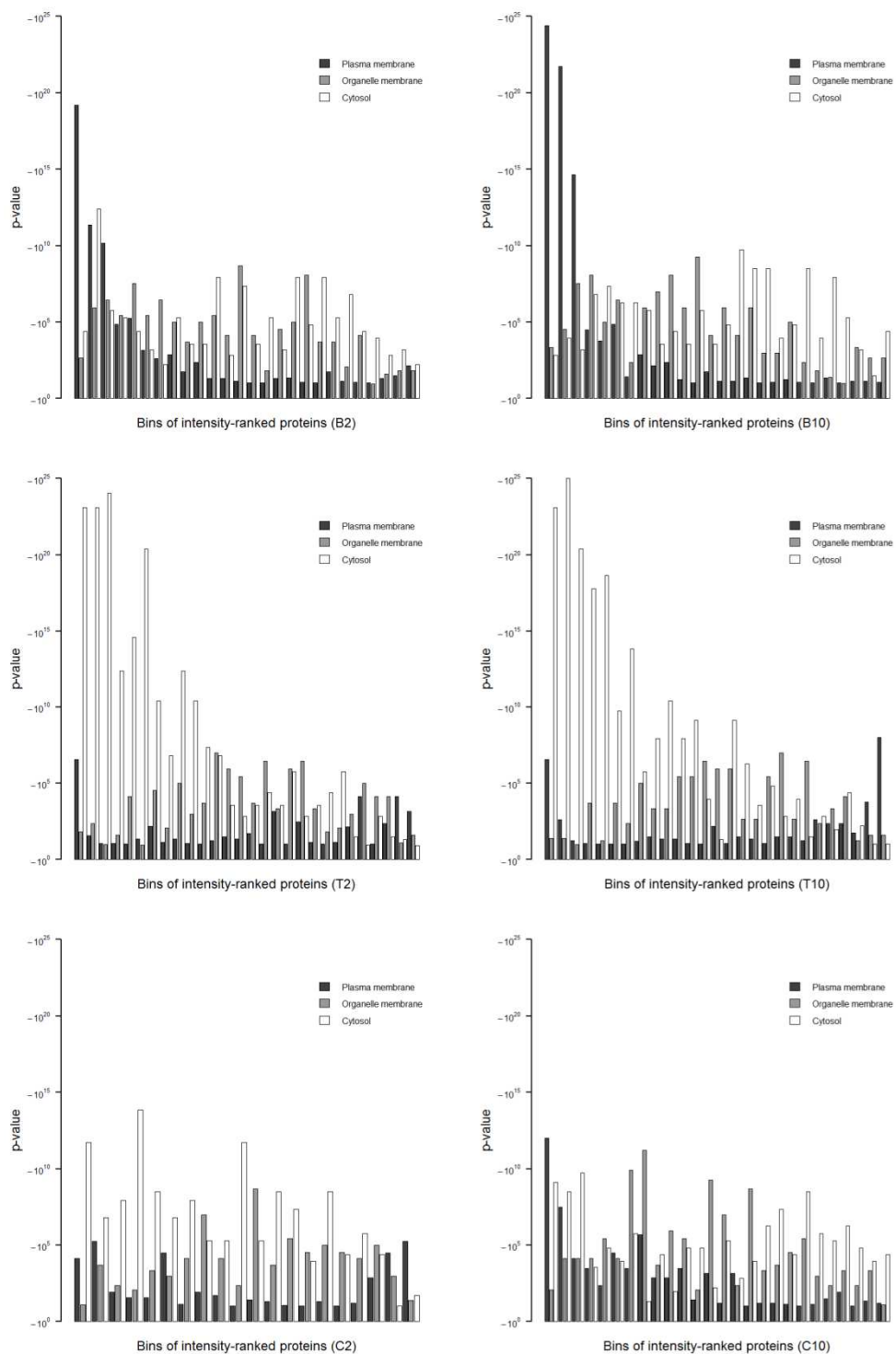
Figure 2.15: Enrichment analysis based on hypergeometric distribution. Proteins were ranked and binned by intensity. In each bin, the number of proteins with a gene ontology-assignment to cytosol, plasma membrane or organelle membrane is estimated. The possibility (p-value) of observing such a number with respect to the bin size, the number of proteins in the UniProt data bank in total and assigned to that gene ontology is calculated with a hypergeometric distribution model. As the significance of accumulation is increasing with decreasing p-values, a taller bar in the plot represents a more significant accumulation. Proteins in high intensity bins were partially identified in the controls as well, but were enriched in amount.
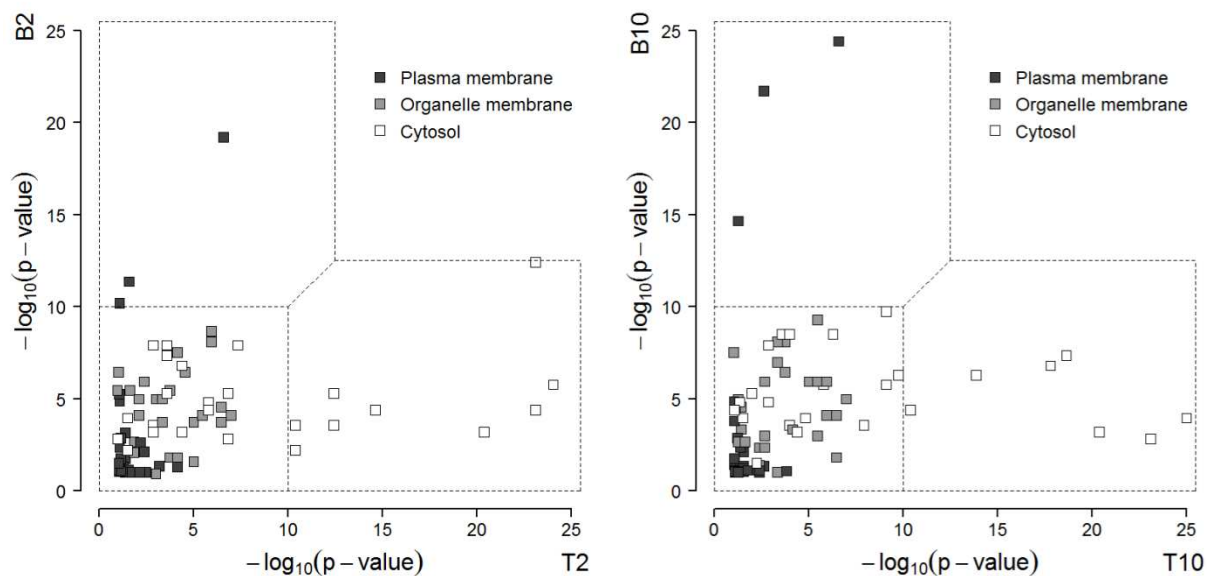
Figure 2.16: P-P plot of cell surface proteome enriched versus total proteome samples (B2,T2 or B10,T10). Regions of low p-values (p-value$^{-10..25}$) can be encircled where emerging labels spot respective enrichments.

2.3.2 Compositional analysis. The detailed compositional proteome analysis is based on the LFQ intensity values, because these values are normalized between samples and thus allow averaging. The $\log_2$-transformed LFQ intensities were simplified by dividing into five abundance classes, below 18 and between 18 to 23, 23 to 28, 28 to 33 and above 33 (Labeled as: -, +, ++, +++, ++++ in table S.1). Figure 2.17 shows a comparison of LFQ intensities for the T10 and T2 samples.
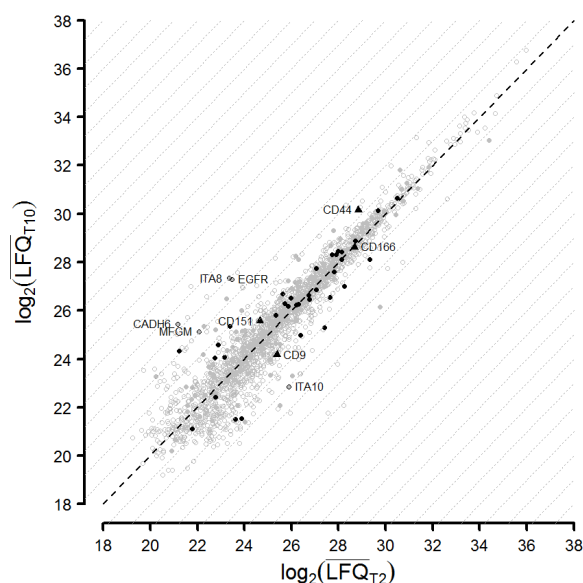
Figure 2.17: Compositional analysis of the total proteomic samples. Averages of LFQ-intensities of cells growing in BE-medium (T2) and S-medium (T10) were compared. Open grey circles – Non CSP proteins; grey circles – CSP-proteins; black circles – CD-marker proteins; triangles – selected CD-marker proteins for FACS validation; open black circles – CSP-proteins, found to significantly differ in t-tests between B10 and B2.

Common mesenchymal stem cell markers are those proteins that enable prolonged self-renewal - the stemness factors (Nanog, Oct-3/4, Sox-2) - and cell surface markers. Table 2.2 compares the respective cell surface markers to our findings. Stro-1 is omitted, because the nature and identity of the associated antigen is unclear. We found 7 of 8 proposed mesenchymal markers in high or highest intensities in the cell surface proteome enriched samples (B) and 5 of this as well in the total proteome sample (T). CD271 was identified in neither of the samples. It remains to be observed, if the absence of CD271 is a result of the isolation or cultivation methods used. None of the hematopoietic stem cell surface markers were identified indicating the purity of the stem cell isolation. Maybe of similar importance are not identified cell surface markers. As they proof to be present on co-occurring cells of the same niche, those markers can be used as negative selection markers in FACS. We did not identify 285 protein markers of the CD system in the total proteome or cell surface proteome enriched samples. This forms a sound base for further work where the cell surface proteome of co-occurring cell types will be elucidated. In addition to the expected mesenchymal stem cell surface markers, we identified 31 CD-marker in the total proteome samples (one was found exclusively here) and 73 CD-marker in the cell surface proteome enriched samples. Beside the CD-classified proteins for which commercial antibodies for FACS analysis are generally available, we identified 79 intense or high intense potential cell surface protein markers in the total proteome samples and 180 in the cell surface proteome enriched samples. This is the so far most comprehensive and rigorous quantitative compositional analysis of a mammalian cell surface proteome.

Table 2.2: Stem cell surface markers (Kawashima, 2012; Karbanová, 2010).

Marker	AccNo	B	T	Name
Mesenchymal stem cell markers				
CD29	P05556	++++	++++	Integrin beta-1
CD44	P16070	++++	++++	CD44 antigen
CD73	P21589	++++	++++	5-nucleotidase
CD90	P04216	++++	+++	Thy-1 membrane glycoprotein
CD105	P17813	++++	-	Endoglin
CD146	P43121	+++	-	Cell surface glycoprotein MUC18
CD166	Q13740	++++	++++	CD166 antigen
CD271	P08138	-	-	Tumor necrosis factor receptor superfamily member 16
Hematopoietic stem cell markers				
CD34	P28906	-	-	Hematopoietic progenitor cell antigen CD34
CD45	P08575	-	-	Receptor-type tyrosine-protein phosphatase C
CD117	P10721	-	-	Mast/stem cell growth factor receptor Kit
CD133	O43490	-	-	Prominin-1

B,T - average of LFQ intensities of all cell surface proteome enriched (B) or total proteome (T) samples.

The results are in good agreement with the previous FACS results of Karbanová et al. (Karbanová, 2010) (Figure 2.18). All hematopoietic markers negatively tested in FACS were also not identifiable in MS. Five mesenchymal stem cell markers positively tested in FACS were intense even in the total proteome. Only a small number of cells were positively tested

for CD146 in FACS, not identified though in the total proteome, and of high but not highest intensity after cell surface protein enrichment. The latter is an example for the bias of the enrichment procedure which is also reflected by a low Spearman's rank correlation coefficient (0.183). An identified cell surface marker should be present on every cell for its suitability, so one might prefer not to identify such an ambivalent protein. On the other hand, the cell surface protein enrichment procedure elucidated, in conjunction with the total proteome, this ambivalence and hence the protein from the negative marker
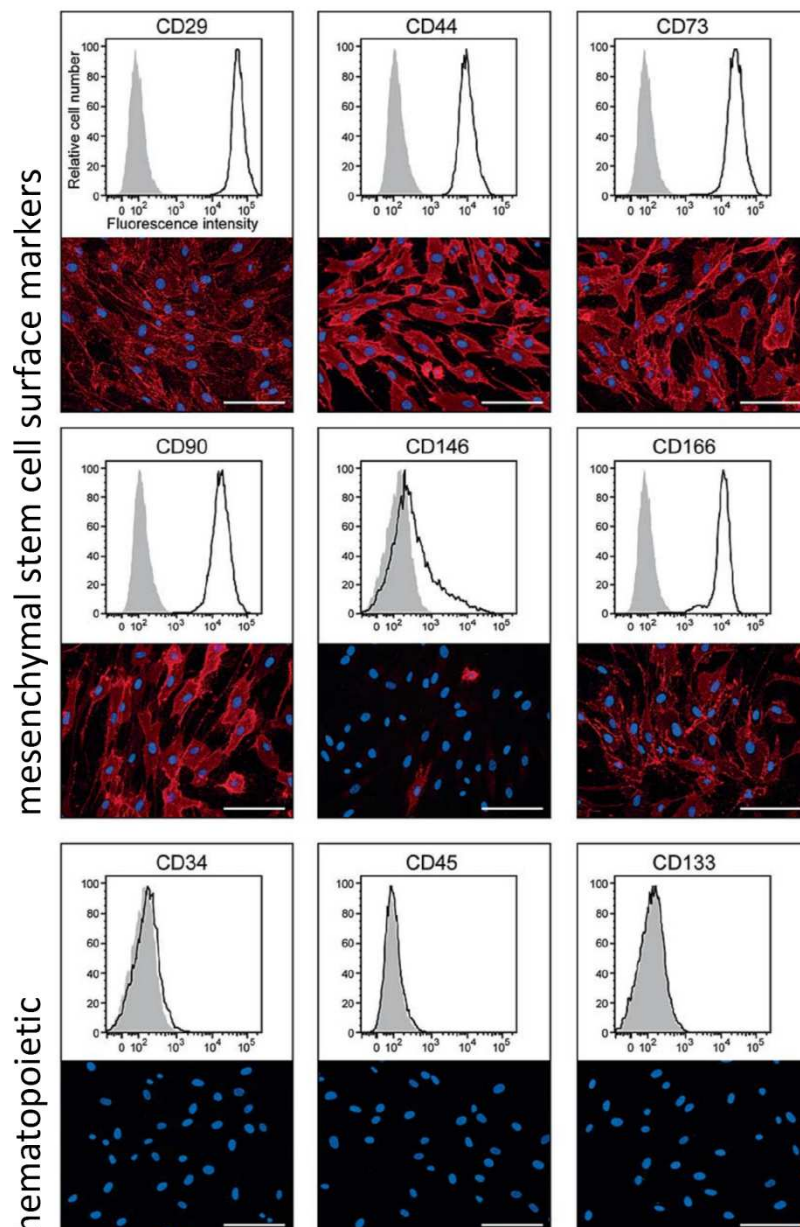
Figure 2.18: FACS analyses and immunofluorescence microscopic images of DPS cells cultivated in BE (B2) medium (from Karbanova2010, PMID 21071916). Unfilled area under the curve – antigen expressing cells. Grey area – appropriate isotype-matching controls.

candidate list. The total proteome sample's LFQ intensity values are yet reflecting the actual relative abundance of the cell surface markers, with the higher risk of false negative identifications.

To further validate the quantitative compositional analysis, six CD-marker proteins of different intensity classes (Table 2.3) were selected for FACS and immunostaining analyses (Figure 2.19). All four markers, high intense or intense in mass spectrometric measurements, show homogeneous occurrence among all cells. The two marker proteins that were only identified upon CSP enrichment appeared only sporadic. No differences were observed between samples of the two media.

Table 2.3: 6 CD-marker proteins, two of which showed high intensity in the total proteome, 2 intense and 2 were only detected after enrichment of cell surface proteins.

CD	AccNo	B	T	Name
CD44	P16070	++++	++++	CD44 antigen
CD166	Q13740	++++	++++	CD166 antigen
CD9	P21926	++++	+++	CD9 antigen
CD151	P48509	++++	+++	CD151 antigen
CD39	P49961	++	-	Ectonucleoside triphosphate diphosphohydrolase 1
CD106	P19320	++	-	Vascular cell adhesion protein 1

B,T - average of LFQ intensities of all cell surface proteome enriched (B) or total proteome (T) samples.
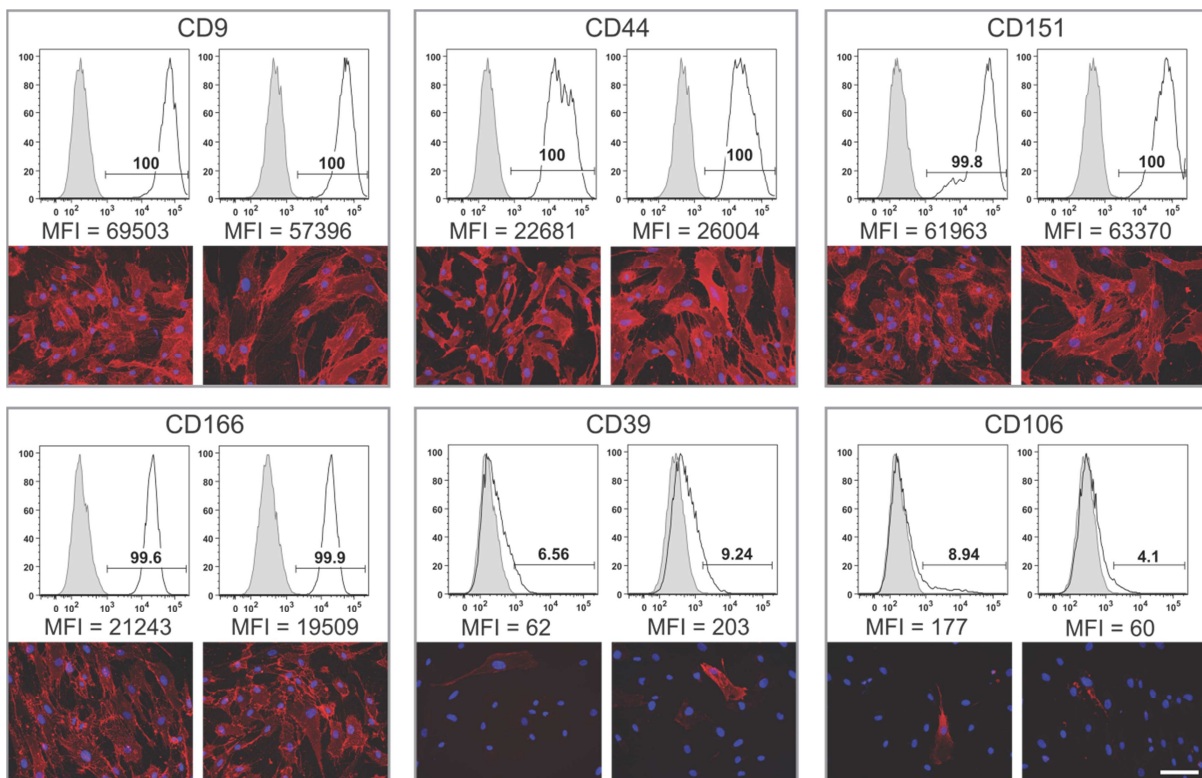
Figure 2.19: FACS analyses and immunofluorescence microscopic images of DPS cells cultivated in BE (B2) medium (Left images) or S (B10) medium (Right images). Unfilled area under the curve – antigen expressing cells. Grey area – appropriate isotype-matching controls.

2.3.3 Differential proteome analysis. Next, we were interested in elucidating differences in the cell surface proteomes of DPSC cultivated in the two different media. Such a procedure should be based on a statistical model. Most proteins in Figure 2.17 are close to the bisecting line, but some proteins deviate by up to 2^4. However, these are only mean values and standard deviations of both LFQ intensities have to be taken into account. Such a plot, with standard deviations in both directions, would be unreadable and the estimation of standard deviations of triplicate samples is vague. Therefore, plots that compare intensity means, like figure 2.17 are not suited for such analyses. To integrate a statistical assessment of the values, a modified t-test is used to calculate p-values and to define a significance threshold line. The negative decadic logarithm of those p-values is plotted over the logarithm of the mean LFQ intensity ratios in a volcano plot (Figure 2.18). Cell surface proteins that pass the threshold and thus represent significant differences between the samples are labeled in figures 2.18 and 2.17. Selected CD-marker proteins which were used for validation experiments, such as fluorescence-activated cell sorting (FACS), are also labeled in Figure 2.17. CD44 and CD166 belong to the most abundant cell surface proteins, CD9 and CD151 are of medium abundance and CD39 and CD106 were not detected in the total proteomic samples, but in the cell surface proteome enriched samples.
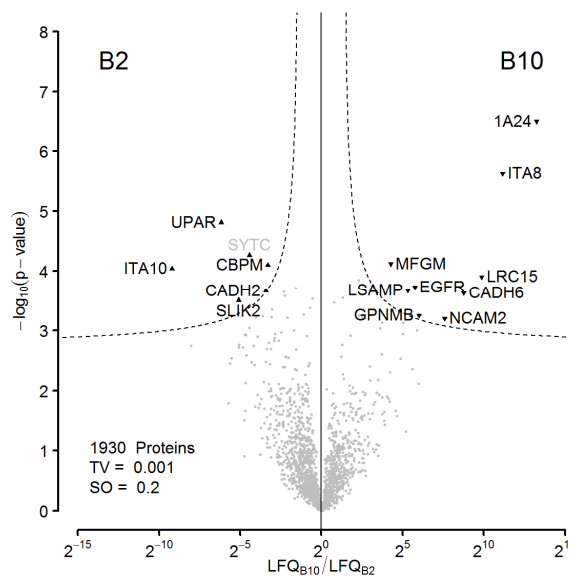
Figure 2.18: Differential proteome analysis. Cells were isolated by cell surface biotin labeling after cultivation, either in BE-medium (B2) or S-medium (B10). 9 proteins are passing a conservative threshold of 0.001 on the B10 side and 6 show preferential abundance in the B2 samples. Black labels – significant hits, assigned to the CSP by the Boolean algorithm. Grey labels – significant non CSP proteins.

The differential proteome analysis (Figure 2.18) unbosomed the mentioned influence of cultivation media on the cell surface proteome. Quantitative PCR and Western blotting (where antibodies were available) were used in subsequent experiments to validate those findings for a broader number of donors (Figure 2.19). Table 2.4 summarizes the results.
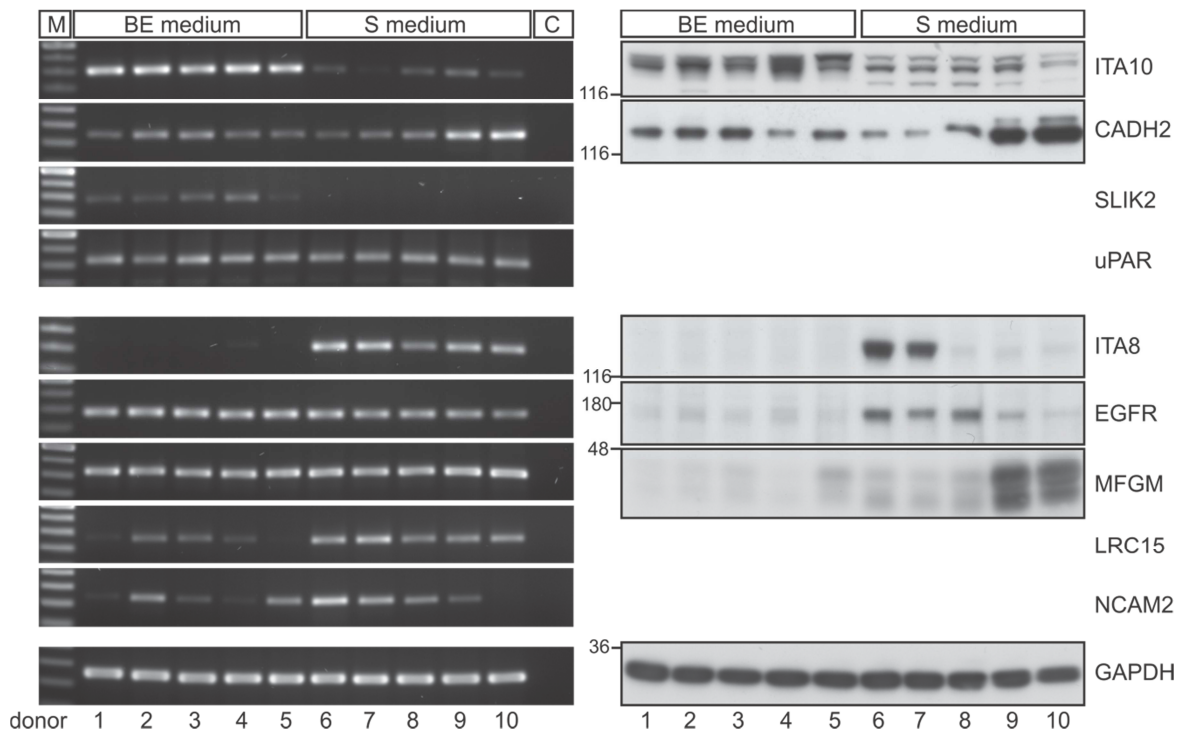
Figure 2.19: Validation experiments of the differential proteome analysis of DPS cells cultivated in two different media. Left - qPCR results. Right – Western blot results. BE – Basic expansion medium (B2), S - Standard medium (B10). Samples from donors 1 and 7 were also used for differential proteome analysis as shown in figure 2.18.

Table 2.4: Summary of the validation of the differential proteome analysis of DPS cells, cultivated in two different media by qPCR and Western blotting of total proteomic samples.

Protein	AccNo	CSP	WB	qPCR	Name
Differential MS-analysis: B2>B10					
UPAR	Q03405	+		0	Urokinase plasminogen activator surface receptor
ITA10	O75578	+	a	a	Integrin alpha-10
CBPM	P14384	+			Carboxypeptidase M
CADH2	P19022	+	h	0,h	Cadherin-2
SYTC	P26639	-			Threonine--tRNA ligase, cytoplasmic
SLIK2	Q9H156	+		a	SLIT and NTRK-like protein 2
Differential MS-analysis: B10>B2					
1A24	P05534	+			HLA class I histocompatibility antigen, A-24 alpha chain
ITA8	P53708	+	a,h	a	Integrin alpha-8
CADH6	P55285	+			Cadherin-6
NCAM2	O15394	+		h	Neural cell adhesion molecule 2
EGFR	P00533	+	a,h	0	Epidermal growth factor receptor
MFGM	Q08431	+	a,h	0	Lactadherin
LSAMP	Q13449	+			Limbic system-associated membrane protein
GPNMB	Q14956	+			Transmembrane glycoprotein NMB
LRC15	Q8TF66	+		a,h	Leucine-rich repeat-containing protein 15

CSP – Cell surface protein, B2 (BE – Basic expansion medium), B10 (S – Standard medium), WB – Western blotting, qPCR – quantitative polymerase chain reaction, a – affirmed difference, h – heterogeneous between donors, 0 – no difference detected, c – conflictive difference (none)

.

The validation discloses differences that occur due to donor-dependent deviations, but also cogent distinctions. The most conclusive distinctions are the integrins 10 and 8. In similar cultivation experiments, a significant increase in integrins ITB1 (CD29) and ITA6 (CD49f), upon cultivation in serum-free media, has been previously reported (Lorenz, 2009). The EGF receptor seems to be comparably expressed in B2 medium, but internalized and degraded, taken by its lower detection in mass spectrometry and Western blotting. As EGF was provided as a supplement in the B2 medium, internalization and degradation of the receptor are part of the normal EGF uptake mechanism (Kirisitis, 2007). The diminished cell surface localization of the receptor could demonstrate a mechanism of counter-regulation to the constant triggering of the EGF-signaling pathway. The threonine-tRNA ligase SYTC is believed to be a false positive finding. Although the t-test threshold was set very strictly, so that the false discovery rate can be assumed far below 1%, false positives can never be completely excluded. No conflictive results were yielded from the validation experiments. For a generalization of the findings, a huger number of donors should be used, preferably for FACS or Western blot analyses. Whilst the findings of Lorenz et al. were yielded by flow cytometry and immunofluorescence, those experiments were hypothesis-driven. The major advantage of the shotgun mass spectrometric approach is the unbiased, discovery-driven way to yield proteomic differences.

One of the purposes of this project was the selection of cell surface markers that can be used in fluorescence-activated cell sorting (FACS) to isolate and quantitate cells like stem cells or cancer cells. The supplemental table S.1 lists the identified CD-markers and intensity grades in the cell surface and total proteome samples. A comprehensive list of all other cell surface proteins is given in the supplemental table S.2.

2.4 Application II – Analysis of membrane-associated protein networks

Introducing a new coupling chemistry and differential proteomics into recruitment assays to study post-Golgi trafficking of a multitude of transmembrane proteins

2.4.1 Recruitment assay development. <u>Synthetic liposomes</u> allow in-vitro studies of membrane-binding processes and enable to elucidate the influence of distinct lipid compositions on this. After recruitment, liposomes can be separated from the cytosol by discontinuous density centrifugation. Catimel et al. (Catimel, 2008) used synthetic liposomes with embedded PI35P2 or PI45P2 (5 mol-%) to study phosphoinositide- dependent and independent recruitment of cytosolic proteins from a LIM1215 colonic carcinoma cell line. They used a somewhat differential proteomic approach by the use of a control sample lacking phosphoinositides. Proteins that were identified with this control were subtracted from the identification lists of the two phosphoinositide containing samples, irrespective of protein abundance changes. Also, because no replicates where measured, their results suffer from a huge false positive content. In shotgun proteomics, identification of proteins is a stochastic procedure meaning that single measurements are tend to lack to identify a minor number of proteins. Biological replicates not only overcome variability of the assay, but also fill-in the gaps in the identification list. If no replicates are done in a differential proteomic assay, many estimated differences can be just accidental. This can be seen in the work of Catimel et al. by many proteins that were classified as been recruited, although they are intrinsic membrane proteins and hence belong to the co-purified proteins that form part of the background. Moreover, these stochastic lacks of identification is likewise leading to a huge false negative rate. In addition, false negatives are also immanent to this approach, because many proteins bind to membrane surfaces unspecific, i.e. with low affinity. When additional factors, like the 'correct' phosphoinositide are presented, the affinity is increased leading to an increased abundance and mass spectrometric intensity-sum for this protein. As in both cases, i.e. in the sample and control, the protein may have been identified, this would lead to an elimination of the protein from the list of recruited proteins and thus a false negative. To complement this project, they repeated the analysis with PI345P3 (Catimel, 2009). Many membrane-associated processes are triggered by proteins that bind to more than a phosphoinositide in a cooperative kinetic. This so-called coincidence detection occurs for instance when adapter complexes like AP-2 assemble and bind in the initial steps of vesicle formation. AP-2 recognizes PI45P2 by two binding sites, at the N-terminus of the AP-2α subunit and within the cargo-recognizing AP-2μ subunit (Krauss, 2007). The binding affinity therefor depends on the endocytic sorting motif of the cargo protein presented in the context of a PI45P2 containing (plasma) membrane. In the liposome-experiments of Catimel et al., AP-2α2 was consequently the only part of adapter complexes found and could moreover indicate that AP-2 complexes itself form on top of membranes - as distinguished from Figure 2.20. However, it could also indicate that just a certain amount of AP-2α is monomeric in the cytosol.
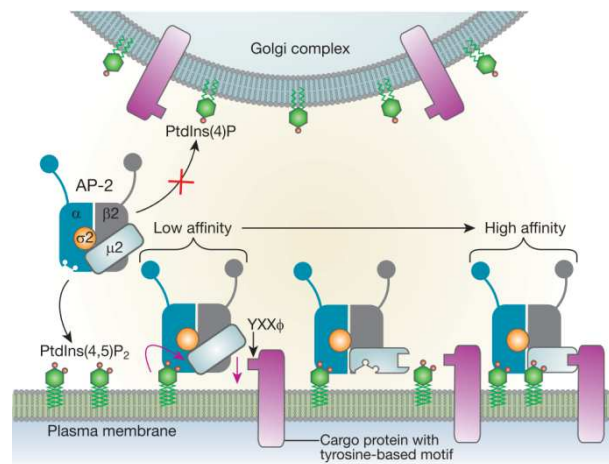
Figure 2.20: Coincidence detection when AP-2 recognizes sorting motifs in the context of a PI45P2-containg membrane (Di Paolo, 2006).

Proteo-liposomes present coupled proteins or parts of proteins in a membrane (phosphoinositide) context. Zhu et al. used liposomes with myristoylated Arf1-GTPγS to successfully recruit AP-1 complexes, even without integral membrane proteins (Zhu, 1999). Crottet et al. extended the approach with a coupled synthetic peptide carrying Lamp1A or TGN38A-derived sorting motifs (Crottet, 2002). Unfortunately, they neglected the opportunity to perform a thorough mass spectrometric analysis of their recruitment samples. Wenk et al. coupled recombinant synaptotagmin to liposomes (Wenk, 2003). Their protocols relied on an amino-terminal gluthathione tag, fused to the protein of interest. They also omitted an extended proteomic analysis. This was subsequently done by Baust et al. (Baust, 2006) in our group (Prof. Dr. Hoflack), who changed the coupling principle into hydrazone bonds, introducing a highly selective and efficient chemistry (Bourel-Bonnet, 2005).

However, this chemistry is also a crux as it relies on synthetic peptides which restrict the application to cargo with short cytosolic domains. Also, this type of modification (amino-terminal hydrazino acetyl) is not offered by commercial peptide synthesis services. Moreover, peptide synthesis is expensive conflicting its use in screening approaches. We therefor changed the coupling chemistry to a thiol-directed maleimide-anchor lipid allowing to couple peptides, either recombinant or synthesized, equipped with a free (nearly terminal) cystein. Baust et al. performed a single proteomic analysis without controls. The classification of recruited proteins into background and recruited proteins relied on manual selection and made unbiased findings impossible. In fact, we expected that, as it was shown by Catimel et al., the majority of proteins both in number and abundance bind unspecific or are co-purified, while a minority represents the truly recruited proteins. Such recruitment assays thus represented an appropriate task for sophisticated differential proteomic analyses with promising possibilities (Figure 2.21).
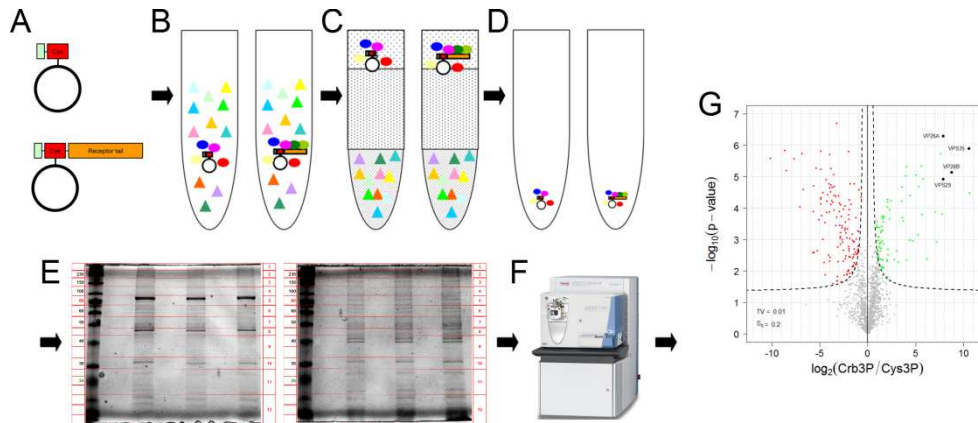
Figure 2.21: Recruitment assay with differential proteomic analysis. A - Preparation of liposomes for control and with receptor sequence. B - Recruitment with cytosol and drugs with triplicates of each. C - Separation of cytosol and liposomes by a sucrose density gradient centrifugation. D - Harvest and wash of the liposomes. E - Fractionation and in-gel digestion. F - High resolution mass spectrometry. G - Label-free quantitative analysis.

In the previous recruitment experiments of Dr. Thorsten Baust in our group (Baust, 2006; Baust, 2008), the adapter complexes AP-1 and AP-3 have been shown to be recruited to the liposomes. Along with that, accessory proteins were found that provided new insights into the mechanisms of trafficking. This motivated us to proceed with in-vitro recruitment assay experiments. The preparation of the liposomes basically followed an established protocol of Baust et al. (Baust, 2006), with the buffer composition adopted with respect to the altered coupling chemistry. The extrusion through a filter membrane was omitted, because the maleimide group of the anchor lipid is sensitive to hydrolysis and extrusion would have extended the exposure of the reactive group to water. Also, we did not want to produce unisize, unilamellar liposomes but a distribution of liposome sizes instead. We suspected that a size selection introduces unwanted biases for proteins sensitive for membrane curvature and we were, in contrast to Baust et al., not aimed to idealize for mimicking clathrin-coated vesicles in particular. The coupling efficiency was monitored by individual controls of the purified construct, Tev-secession of the His-MBP-tag, the supernatant of the liposome wash after coupling and the pelleted liposomes with coupled receptor domains. Tev-digestion and coupling was always successful and coupling of the full construct never observed. The elimination of the tag was not completely accurate, but remaining traces of the tag were incapable of coupling due to the total absence of cysteine. As to the subsequent exposure to cytosolic proteins, the presence of a minor fraction of tag-protein was tolerated. We were confident thus that we had the proper complex baits at hand for recruitment screening experiments.

The biochemical workflow is depicted in Figure 2.22. The recruitment time, the concentration of the cytosol, and the concentration of GTP-γS were optimized on PI4P-containing liposomes with coupled Varicella zoster virus glycoprotein E (gE) using Western blot analyses of clathrin heavy chain (CHC) and adapter protein 1 gamma (AP1G) (Figure 2.23). Optimal differences between gE-samples and Cys-controls were achieved by a

cytosolic protein concentration of 4 mg/ml, a GTP-γS concentration of 1.5 µM, and a recruitment time between 8 and 16 minutes.

```
express construct        prepare          isolate mouse
6xH-MBP-Tev-C-Rec.      liposomes         brain cytosol
        |                   |                   |
        v                   |                   |
  purify construct          |                   |
        |                   |                   |
        v                   v                   |
   Tev-cleavage  -->  couple cytosolic          |
                      receptor domain           |
                            |                    |
                            v                    |
                        /       \                |
   supernatant:        / spin down \             |
   6xH-MBP and  <----- and wash    |            |
   unbound C-Rec.      \ liposomes /             |
                        \       /                |
                            |                    |
                            v                    |
                        pellet:                  |
                      Lipo~C-Rec.                |
                            |                    |
  agents, e.g.:             v                    |
  GTP-γS        -->     Recruitment  <-----------+
  ATP
  Latrunculin B             |
                            v
                     Isolate Proteo-
                       liposomes
                      by density
                      centrifugation
                            |
                            v
                       spin down
                       and wash
                            |
                            v
                        SDS-PAGE
                            |
                            v
                       Gel-lane
                      fractionation
                            |
                            v
                    in-gel reduction,
                     alkylation and
                    trypsin-digestion
                            |
                            v
                         LC-MS
```

Figure 2.22: Development of an improved recruitment assay. Cytosolic domains of membrane receptors are coupled via an introduced cysteine to a maleimido-functionalized lipid anchor. As a control, a similar liposome preparation is saturated with free cysteine instead. Recruitment reactions are performed for receptor containing and control liposomes in parallel. For screening of receptor baits and their addiction to specific trafficking adapters, spectral counts of unique samples are compared. Triplicates of such pairs of samples and controls enable application of statistically validated mass spectrometric label-free relative quantification of recruited proteomes.
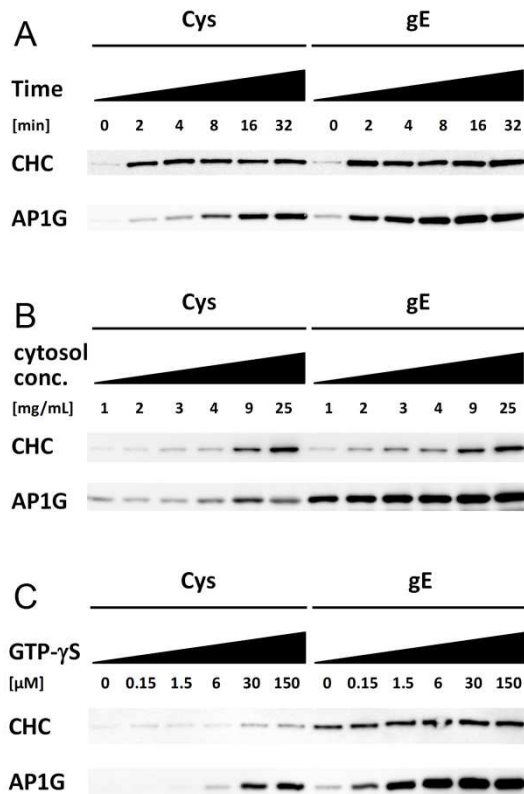
Figure 2.23: Optimization of recruitment parameters. Recruitment experiments were performed using Varicella zoster virus glycoprotein E (gE) on PI4P-containg liposomes with varying recruitment times (A), cytosolic protein concentrations (B), and GTP-γS concentrations. Clathrin heavy chain (CHC) and AP1G were detected by Western blotting.

2.4.2 Recruitment screen. The purpose of the screening experiment was to assess the recruitment assay with an array of different membrane proteins for which different sorting motifs and hence interacting proteins were expected. As in the previous studies, we used the varicella-zoster virus envelope glycoprotein E (gE), as it is a known cargo of the adapter complex 1 (AP-1). We used the vesicular stomatitis virus glycoprotein G (VSVG), as it is known for clathrin-dependent endocytosis to the endosomes. We included the beta amyloid precursor protein (APP), for which trafficking and localization is partially known, to further extend the validation of our approach and to enable unbiased investigation of this key player in Alzheimer's disease. The localization and trafficking of sortilin (SORT) have been studied in depth before and we included it into the screening because of its three different types of known trafficking motifs (Table 2.5). We used the transferrin receptor (TFR) as an example of a type-II membrane protein, neuroligin 1 (NLG1), because we were interested in the cytosolic interaction partners beyond trafficking and crumbs 2 (CRB) to elucidate trafficking of this cell polarity determinant. The cytosolic sequences of this membrane proteins and the supposed trafficking motifs therein are listed in table 2.5. We used different types of phosphoinositides (either PI3P, PI4P, PI45P2 or PI345P3), as phosphoinositides represent membrane labels for complexes, like trafficking complexes, that need to recognize and form on top of specific organelles.

Table 2.5: Cargos and sorting motifs.

Name	Motif	Sequence after cleavage by *Tev*
APP	YXXØ,NPXY, YXXØ, (Y)KFFE	GSCGKKKQ**YTSI**HHGVVEVDAAVTPEERHLSKMQQNGYE<u>**NPTYKFF**</u>EQMQN
CRB		GSCGRKRRQSEGTYSPSQQEVAGARLEMDSVLKVPPEERLI
gE	YXXØ, acidic cluster	GSIEGRCTKRMRVKAYRVDKSPYNQSMY**YAGL**PV**DDFEDSESTDTEEEF**GNAIGGSHGGSSYTVYIDKTR
NLG	P-rich	GSCGYYKKDKRRHDVHRRCSPQRTTTNDLTHAPEEEIMSLQMKHTDLDHECESIHPHEVVLRTACPPDYTLAMRRSPDDIP LMTPNTITMIPNTIPGIQPLHTFNTFTGGQNNTL**PHPHPHPH**SHSTTRV
SORT	[WF]L[MV]*, YXXØ ,DXXLL	GSCGKKYVCGGR**FLV**HR**YSVL**QQHAEANGVDGVDALDTASHTNKSGYHDDS**DEDLL**E
TFR	YXXØ	GSMMDQARSAFSNLFGGEPLS**YTRF**SLARQVDGDNSHVEMKLAVDEEENADNNTKANVTKPKRCSGSIC
VSVG	YXXØ, YXXØ	GSCGYL**YIKL**KHTKKRQI**YTDI**EMNRLGR

The varicella zoster virus envelope protein gE was not part of the screen. *retromer binding motif according to Seaman 2007 [Seaman2007, PMID17606993]. Ø - bulky hydrophobic residue.

The gel band patterns among experiments with the same phosphoinositide are related, but not identical suggesting a broad background of proteins binding to the liposomes irrespective of the receptor, but affected by the phosphoinositides used. Many of the trafficking complex proteins were identified in the majority of samples. Hence, the spectral count values were extracted from the data to enable a simple quantitative examination (Figure 2.25). All member of the AP-1 complex, except for AP1M2, were found in the controls, highly in the APP, sortilin and TfR samples, some in NLG and minor counts were received for crumbs and VSVG. A tendency for a preference of PI4P containing liposomes or others is not apparent. All members of the AP-2 complex were also widely identified across the samples, with minor counts in the controls, crumbs and VSVG and high counts for the PI345P3 containing liposomes decorated with the cytosolic domains of APP, NLG, sortilin or TfR. Members of the AP-3 complex were found only sporadic and in low amounts. The full complex was found in the controls, with higher counts for MBP and in the PI345P3-TfR experiment. Members of the AP-4 complex were only identified in the APP-samples with an unclear preference for the phosphoinositides. The AP-5 complex members were not identified. Members of the cargo-selective retromer core complex were found in smallest amounts in the controls, minor amounts in the crumbs and sortilin samples and high amounts in the TfR, APP, NLG and VSVG samples. A general preference for a specific phosphoinositide is not visible. Clathrin heavy chain was found with all receptors and controls, highest in sortilin for which also highest counts for AP-1 and AP-2 members were achieved. A preference for PI345P3 containing samples can be spotted in which partially also clathrin light chain proteins were identified. The clathrin coat assembly protein AP180 was found across all samples but doesn't seem to correlate with the clathrin proteins.
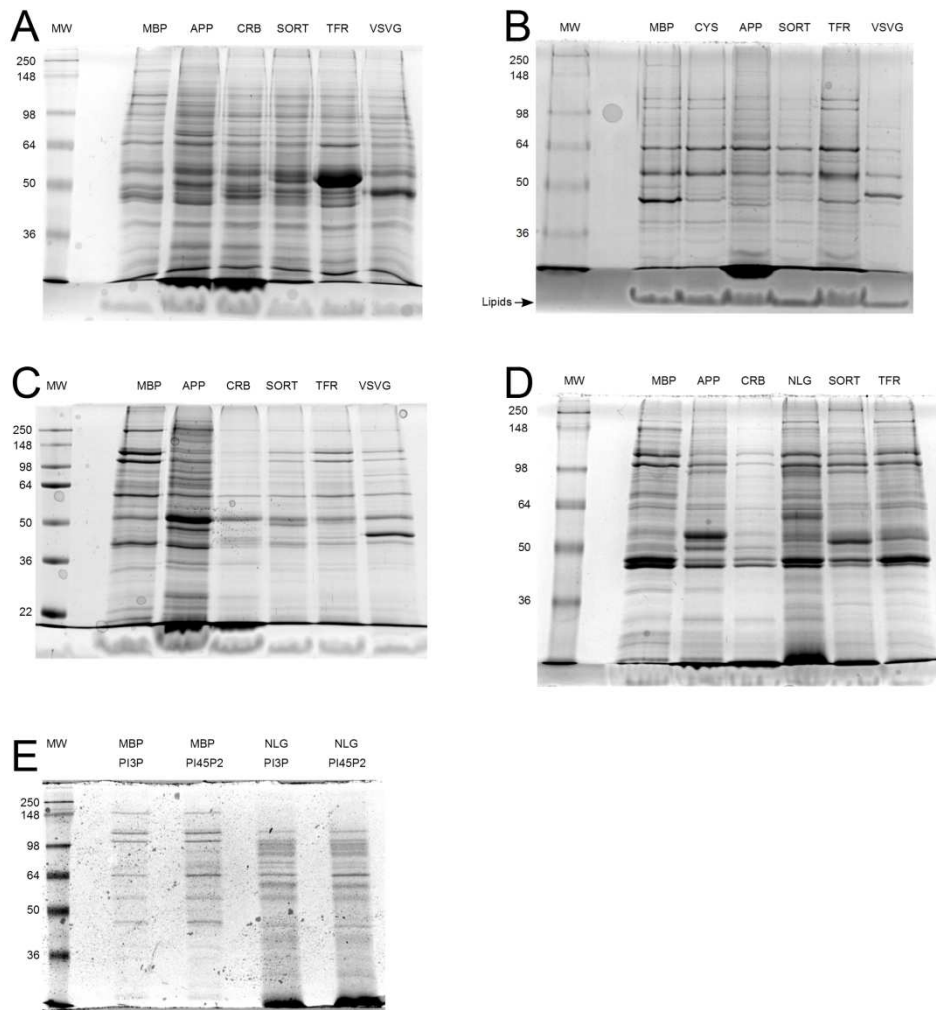
Figure 2.24: Recruitment screening. Recruitment experiments were performed with a library of cytosolic receptor domains and liposomes that differ in the phosphoinositides used. A – PI3P containing liposomes, B – PI4P, C – PI45P2, D - PI345P3, E – PI3P and PI45P2 as labeled. For all gels, 50% of the final samples were loaded. MW-Molecular weight marker, CYS – Cystein and MBP were used as a control, APP – beta amyloid precursor protein, CRB – Crumbs 2, NLG – Neuroligin 1, SORT – Sortilin, TFR – Transferrin receptor, VSVG – Vesicular stomatitis virus glycoprotein G.

When having a closer look on the retromer complex members, one can identify clear recruitment of the complex members in the APP, NLG, TfR and VSVG experiments with respect to the controls, but only relatively small increased counts in the crumbs-PI3P sample. It's not obvious how a significance threshold for a spectral-count ratio (crumbs vs. respective control) can be set. A better alternative is the use of a G-test analysis where the sum of all spectral counts of a sample is used for normalization and the test statistic can be approximated by a chi-square distribution. The model allows to classify quantitative comparative protein data into strongly significantly recruited, significantly recruited and not significantly recruited, and hence allows to compare different proteins, with different absolute values, to be compared. This is specifically of interest for complexes, like the adapter complexes, where proteins of fairly different sizes, and thus spectral count levels, occur.
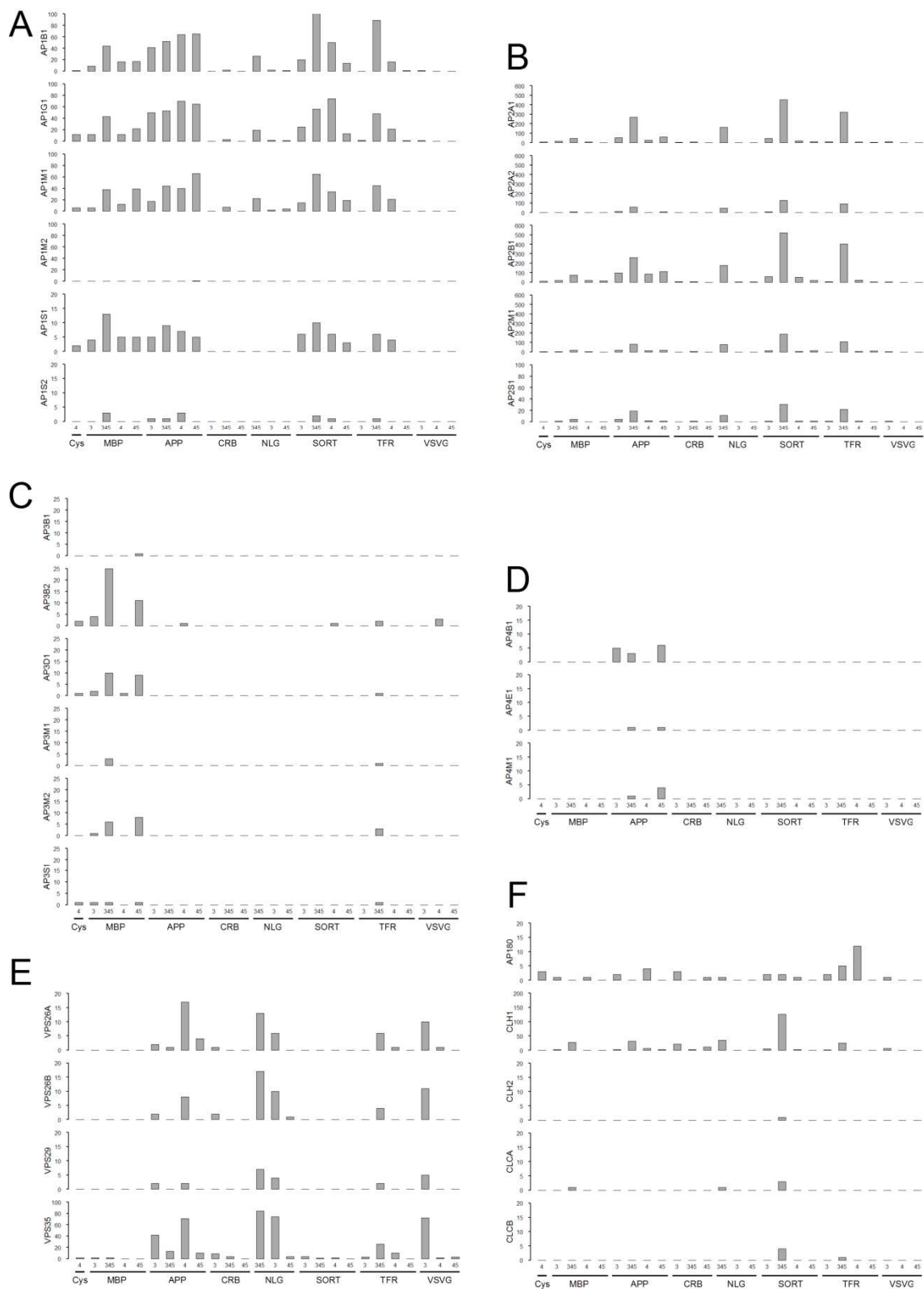
Figure 2.25: Spectral count analysis of trafficking complex members. A - AP-1 complex, B - AP-2 complex, C - AP-3, complex, D - AP-4 complex, E - Retromer complex, F - Clathrins and clathrin associated proteins. Labels: 3,4,45,345 refers to the respective phosphoinositeds in the liposomes, i.e. PI3P, PI4p, PI45P2 and PI345P3 (Other labels as in Figure 2.19).

Table 2.6: G-test applied to spectral count results of the recruitment screen. Spectral counts that were significantly enhanced in the receptor domain containing sample versus the respective control are labeled with + when passing a confidence threshold of p=0.01 or with (+) for p=0.05.

	APP				CRB			NLG			SOR				TFR				VSVG		
	3	345	4	45	3	345	45	345	3	45	3	345	4	45	3	345	4	45	3	4	45
AP1B1	+		+	+							(+)	+	+								
AP1G1	+		+	+							(+)		+								
AP1M1		+									(+)		+								
AP1M2																					
AP1S1																					
AP1S2																					

	APP				CRB			NLG			SOR				TFR				VSVG		
	3	345	4	45	3	345	45	345	3	45	3	345	4	45	3	345	4	45	3	4	45
AP2A1	+	+	+	+				+			+	+	+	+	+		+				
AP2A2	+	+		+				+			+	+			+						
AP2B1	+	+	+	+				+			+	+	+	(+)	+						
AP2M1	+	+	+	+				+			+	+	(+)	+	+		(+)	+			
AP2S1		+									+				(+)						

	APP				CRB			NLG			SOR				TFR				VSVG		
	3	345	4	45	3	345	45	345	3	45	3	345	4	45	3	345	4	45	3	4	45
AP3B1																					
AP3B2																					
AP3D1																					
AP3M1																					
AP3M2																					
AP3S1																					

	APP				CRB			NLG			SOR				TFR				VSVG		
	3	345	4	45	3	345	45	345	3	45	3	345	4	45	3	345	4	45	3	4	45
AP4B1	(+)			+																	
AP4E1																					
AP4M1																					

	APP				CRB			NLG			SOR				TFR				VSVG		
	3	345	4	45	3	345	45	345	3	45	3	345	4	45	3	345	4	45	3	4	45
AP180															(+)	+					
CLCA																					
CLCB																					
CLH1		+			+		+				+										
CLH2																					

	APP				CRB			NLG			SOR				TFR				VSVG		
	3	345	4	45	3	345	45	345	3	45	3	345	4	45	3	345	4	45	3	4	45
VP26A		+						+	+						(+)				+		
VP26B		+						+	+										+		
VPS29								+											+		
VPS35	+	+	+	+	(+)			+	+						+	+			+		

In table 2.6, the G-test is applied to the spectral count data for p-values of 0.01 and 0.05. This cleaned the data from dubious values and enhanced the explanatory power of the spectral counts. Herewith, AP-1 recruitment is achieved presumably with PI4P-liposomes and attached APP or sortilin. AP-2 recruitment seem to be slightly better with liposomes containing PI345P3 and no significant recruitment of AP-3 complex members can be observed, while there is still evidence of AP4B1 recruitment on APP-liposomes. While APP

still proves to recruit retromer, only VPS35 was significantly passing the test for crumbs with a relaxed p-value of 0.05 leaving an ambiguous statement about retromer complex recruitment under this conditions.

2.4.3 gE coupled to PI4P-liposomes. The varicella zoster virus (VZV), or human herpes virus 3 (HHV-3) is a member of the alphaherpesvirinae subfamily of the herpesviridae. VZV encodes eight membrane glycoproteins, gB, gC, gE, gH, gI, gK, gL and gM (Govero, 2007). One basic question in virology is how and thus where inside the cell a virus is assembled. This virus assembly requires the cell surface proteins being accumulated and in case of viruses like the VZV, that are surrounded by a lipid envelope, this means that accumulation has to take place in a membrane domain of a certain organelle. Hence VZV envelope proteins were objects of extended trafficking studies. On the surface of VZV-infected cells, gE is the most abundant VZV glycoprotein (Mo, 2003). On the contrary, most of the gE localizes not on the plasma membrane, but on the trans Golgi network. After their synthesis, transmembrane proteins are transported to and through the Golgi. It was shown that gE does not reside in the TGN, but is cycled between the plasma membrane and the TGN via an early endosomal compartment and it is in fact a steady-state distribution instead that localizes the majority of gE to the TGN (Alconada, 1996). As a hypothesis, the steady-state distribution offers a real time control of the amount of gE localized at the TGN in parallel to the assembly of the virus. The gE protein must contain trafficking signals for this routes. The sequence comprises two motifs that are presumably having such a function (Table 2.5). It is at first, an YXXØ motif. Such motifs are present in many cellular transmembrane proteins and in an extremely high number of herpes virus membrane proteins (Favoreel, 2006). Functions that have been presumed for such tyrosine-based motifs of alphaherpesvirus envelope proteins are the AP-2-mediated endocytosis, COP-II-mediated ER to Golgi transport, AP-1B-mediated basolateral sorting, AP-3-mediated sorting to axons and the target for signaling by phosphorylation through tyrosine kinases like Src (Favoreel, 2006) (Figure 2.26). The second trafficking motif present in gE is an acidic cluster with embedded potential serine and threonine phosphorylation sites. The calcium-independent mannose 6-phosphate receptor is a cellular protein with an acidic cluster motif and it was shown that this motif is critical for receptor sorting at early stages of intracellular transport following endocytosis (Totorella, 2007). PACS-1 is a known effector of acidic clusters controlling the correct subcellular localization of integral membrane proteins. It associates with both AP-1 and AP-3, but not AP-2, and forms a ternary complex with its cargo (Crump, 2001). Other parts of the gE-sequence underlie the evolutional pressure of economy and their thus obligatory functions remain to be discovered.

We used the cytosolic domain of gE in recruitment experiments to assess the assay for its ability to reflect accepted trafficking events. Figure 2.27A shows the results of the coupling. After cleavage with the protease Tev, the construct is split into an appr. 40 kDa cystein-free
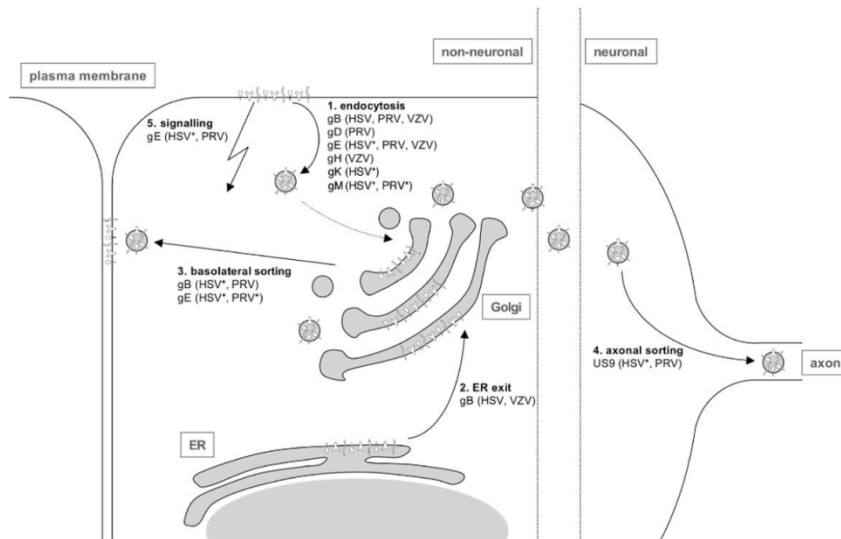
Figure 2.26: Trafficking ways operated by the YXXØ motif in herpes virus envelope proteins (Favoreel, 2006).

His-MBP tag and an appr. 10 kDa peptide representing the cytosolic domain of the type I membrane receptor with an amino-terminal introduced cystein for coupling. A small increase of roughly 1.5 kDa according to the gel's molecular weight marker demonstrates coupling to the anchor lipid, which was functionalized with a maleimide head group. The remaining uncoupled peptide indicates saturation of the anchor. After washing of the liposomes, the liposome-free supernatant lacks the gel-band of the coupled gE, as expected. The comparison of gel bands of a cystein-control and a gE-sample after recruitment show related, but not identical patterns (Figure 2.27B) pointing out a broad background proteome. Also, in addition to the debut of gel bands in gE with respect to the control, prominent bands resign suggesting a displacement of pure membrane binders, when protein networks are formed.
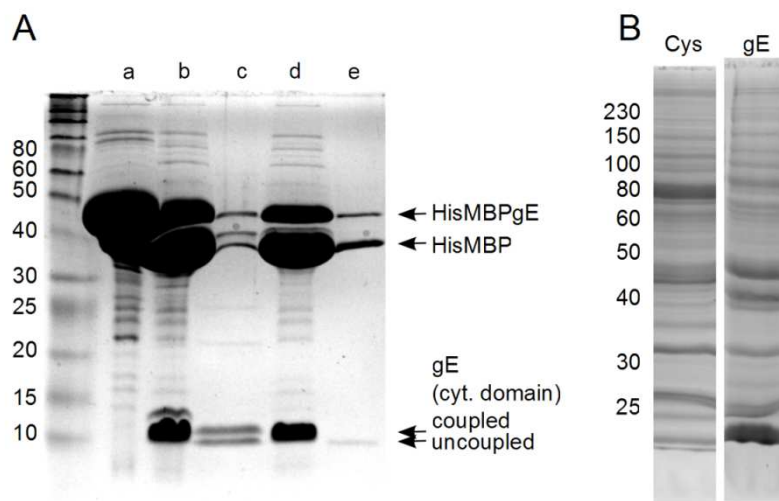
Figure 2.27: Recruitment experiment with the cytosolic domain of gE coupled to PI4P-liposomes. A - Coupling controls: a-full construct, b-construct cleaved by Tev, c-gE-peptide after coupling to the liposomes, d-supernatant after first washing of the liposomes, e-supernatant after second wash; B - Comparison of recruitment sample (gE) and control (Cys).

Differential proteomic analysis on recruited protein networks was performed then to introduce firmer statistic models and hence yield more rigorous lists of recruited proteins. The main experimental difference to the screening experiments is - beside the complex bioinformatic data processing - that controls and replicates (at least triplicates) are generally required. Also, mouse cytosol instead of pig was used, because of the greater data base (UniProt) annotation and hence identification probability of tryptic peptides and proteins. However, for a screening experiment the use of pig brain cytosol is justified, because the proteome is largely similar to human or mouse, but much easier to prepare in huger amounts.

We expected to recruit AP-2 and AP-1 along with clathrin because of the YXXØ motif and PACS as an indicator for trafficking caused by its acidic cluster. We had no evidence to expect other adapter complexes, retromer or ESCRT complexes to be recruited.

All members of the heterotetrameric complexes AP-1 and AP-2, including two isoforms of AP-1S (S1 and S2) and AP-2A (A1 and A2), were identified (Figure 2.28). The AP-2 complex was even recruited on PI4P-containing liposomes. It remains elusive if the reason for the seeming unspecificity of AP-2 with respect to the phosphoinositide was caused by the design of the experiment, with a large amount of concentrated cargo and too less PI4P, or if a loss or conversion of PI4P occurred. Interestingly, all complex members were found in tight ratio ranges, which could be due to equal affinities to the pure liposome membrane and a stoichiometric recruitment with the receptor. The complete AP-3 complex was also present, irrespective of the presence of the receptor and thus not recruited specific. AP-4 and AP-5 complex members were not identified. The retromer complex was not recruited, as VPS35 and VPS26B were present independent of gE. Clathrin heavy chain (CHC) and the clathrin coat assembly proteins AP180 AND PICA (Phosphatidylinositol-binding clathrin assembly protein) were recruited. CHC was hereby recruited in a similar ratio as for instance the AP-2 members. The supplemental table S.3 lists all proteins that passed the t-test threshold. There is no evidence that clathrin light chain was also recruited. The absence of the small (25 kDa) clathrin light chain is most probably founded by a lack of identification in the probabilistic fragmentation procedure of shotgun mass spectrometric methods. This could be overcome in a mass spectrometric experiment by adding more replicates or by the addition of another dimension of peptide fractionation prior mass spectrometry. However, it should be mentioned that this would highly increase measurement time and costs. Western blotting can often be used alternatively to answer such targeted questions in recruitment studies. The experiment further elucidated the recruitment of PACS2. As it is closely related to PACS1 (500 identical amino acids on the respective positions of the 961 and 862 amino acids of PACS1 and PACS2) and AP-3 was not recruited, PACS2 presumably aids in the AP-1 mediated trafficking of gE. On the contrary, PACS1 together with PLIN3/TIP47 was also found to be involved in retromer-mediated trafficking, which is clearly not related to gE according to our data (Seaman, 2005; Bonifacino, 2006).
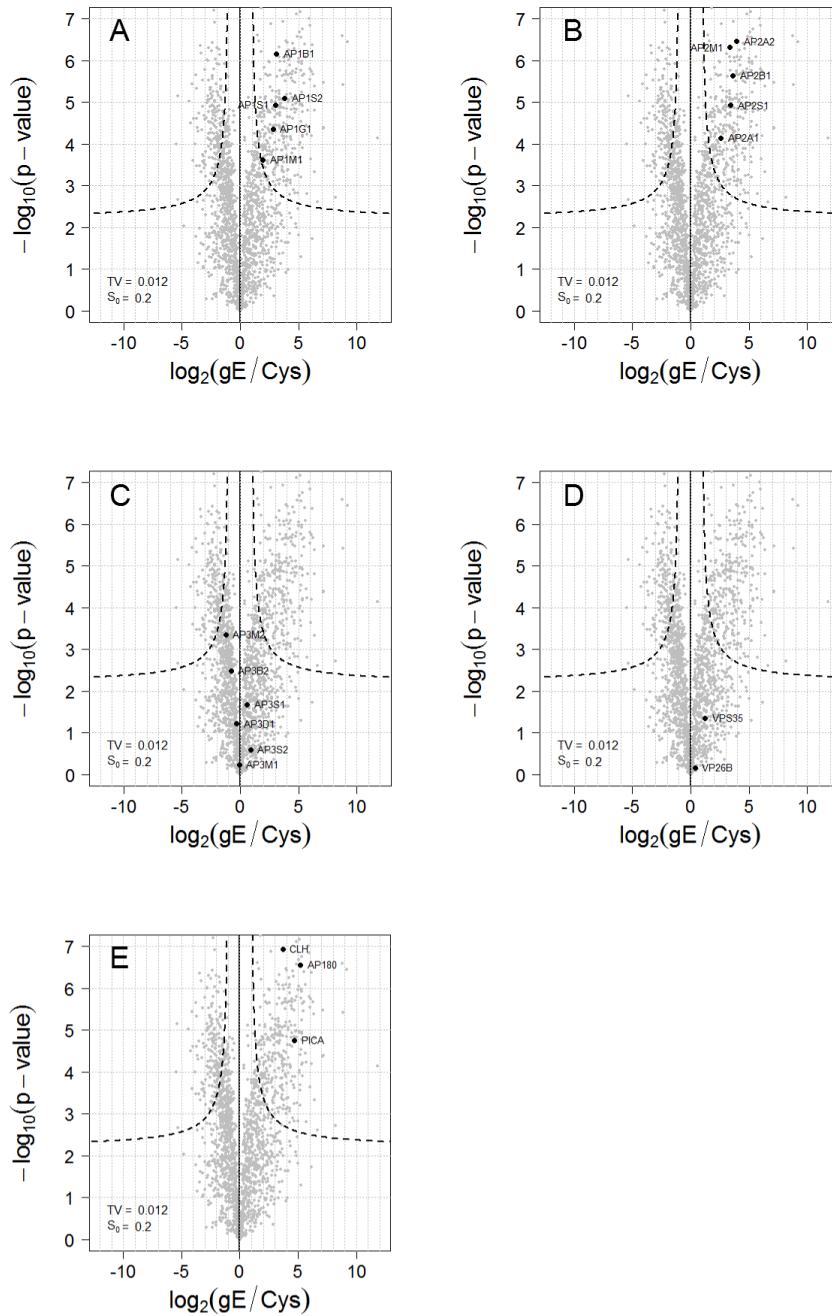
Figure 2.28: Post-golgi trafficking complexes identified in the differential proteomics analysis. A - AP-1 complex, in mouse composed of AP1G1/2 (either G1 or G2), AP1B1, AP1M1/2, AP1S1/2/3; B - AP-2, composed of AP2A1/2, AP2B1, AP2M1, AP2S1; C - AP-3, composed of AP3D1, AP3B1/2, AP3M1/2, AP3S1/2; D – Retromer cargo recognition subcomplex composed of VPS26A/B, VPS29, VPS35; E - Clathrin and clathrin assembly proteins, CHC1, CLCA/B, AP180, PICA; Members of the AP-4 (composed of AP4E1, AP4B1, AP4M1, AP4S1, AP4AT) , AP-5 (composed of AP5Z1, AP5B1, AP5M1, AP5S1), or ESCRT-0 (STAM1/2, HGS), ESCRT-1 (TS101, VPS28, VP37A/B/C/D, MB12A/B), ESCRT-2 (VPS25, VPS36,SNF8), ESCRT-3 complexes were not identified.

The achieved results were in very good agreement with the known mechanisms in the trafficking of gE. This encouraged us to use the workflow for the analysis of unknown trafficking routes and interactors of other membrane proteins.

2.4.4 Crumbs2 coupled to PI3P-liposomes. Crumbs is an evolutionary conserved apical determinant. It is essential for many epithelial tissues for maintaining apicobasal polarity and integrity (Pocha, 2011). As a part of a group of apical polarity regulators it is functional related to a serine/threonine kinase (aPKC), a lipid phosphatase (PTEN), a small GTPase (Cdc42), FERM domain proteins (Moesin, Yurt), and several adaptor or scaffolding proteins (Bazooka/Par3, Par6, Stardust, Patj) (Tepass, 2012). Par3, Par6 have been implicated in endocytosis and recycling. Moreover, Cdc42 and aPKC have been shown to regulate trafficking mechanisms and crumbs was presumed to be the primary target of endocytic regulation by Cdc42, its downstream acting aPKC and the Par proteins (Tepass, 2012).

The cytosolic domains of all crumbs proteins are highly conserved suggesting that their trafficking mechanisms may also be conserved (Pocha, 2011). We chose crumbs2, because it is the predominant crumbs gene expressed in the vertebrate brain, which allowed us to perform recruitment experiments with the respective brain cytosol. Crumbs lacks all known trafficking motifs leaving its trafficking routes so far elusive (Table 2.5).

The G-test in table 2.6 unbosomed no recruitment of any of the AP-1 to AP-5 complex members in experiments with the crumbs-like protein 2 (crumbs, Crb). But nevertheless, clathrin heavy chain was recruited. The involvement of retromer in the trafficking of crumbs was indicated through recruitment of VPS35, but premised to accept a p-value of 0.05. These results inspired us to a more detailed analysis for two reasons. At first, it was so far unknown that crumbs is trafficked by the retromer complex. Secondly, the smallest indication of the recruitment of a complex is a demanding approach to benchmark the sophisticated differential proteomic workflow. Therefore, triplicate recruitment experiments were performed on PI3P-liposomes that were either decorated with crumbs or not (Figure 2.29). As an additional control, PI4P-liposomes were used without crumbs.

A String-analysis (www.string-db.org) of CRUM2_MOUSE results in potential interactors that were previously experimentally identified. Of this, MPP5 and MPDZ were recruited in the recruitment experiments (Figure 2.30), hereby validating the approach. We aimed to observe, if the implementation of a sophisticated differential proteomics approach delivers clearer indication of retromer recruitment. This implies at first, a positive result with a lower rate of false positives allowed and at second, evidence of recruitment of other retromer core complex members. We fully succeeded with this approach and yielded VPS35, VPS29 and VPS26A/B clearly recruited, along with clathrin heavy and light chains (Figure 2.31-A) (Niehage, 2013, in press). The four sorting nexins (SNX1/2/5/6) were also recruited, but as a comparison of the cargo-free liposomes containing PI3P or PI4P showed, predominantly because of the PI3P-content (Figure 2.31-B). However, the small abundance-increase of this sorting nexins in presence of the crumbs-cargo and the retromer core complex can be interpreted as a coincidence detection mechanism.
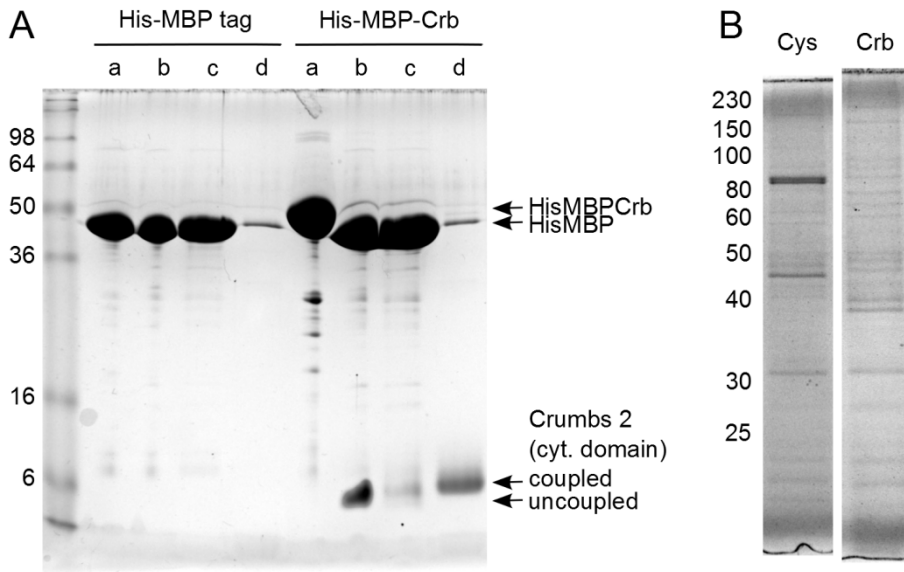
Figure 2.29: A - Coupling controls of crumbs and the cysteine-free His-MBP tag. a: Purified expressed protein construct. b: Construct after Tev-cleavage. c: Supernatant on top of pelleted liposomes after coupling. d: Liposome-pellet. Arrows: a small offset between the gel bands indicates success of coupling. B - Comparison of gel lanes from recruitment experiments without (Cys) and in presence of crumbs (Crbs). Each lane was fractionated into 10 slices for separate mass spectrometric analysis.
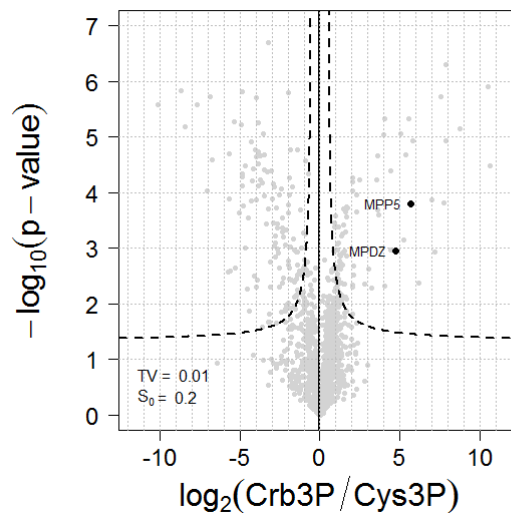
Figure 2.30: Volcano plot of CRB vs. Cys PI3P. All two identified known direct crumbs 2-interactors emerge on the upper right side, representing significantly recruited proteins.
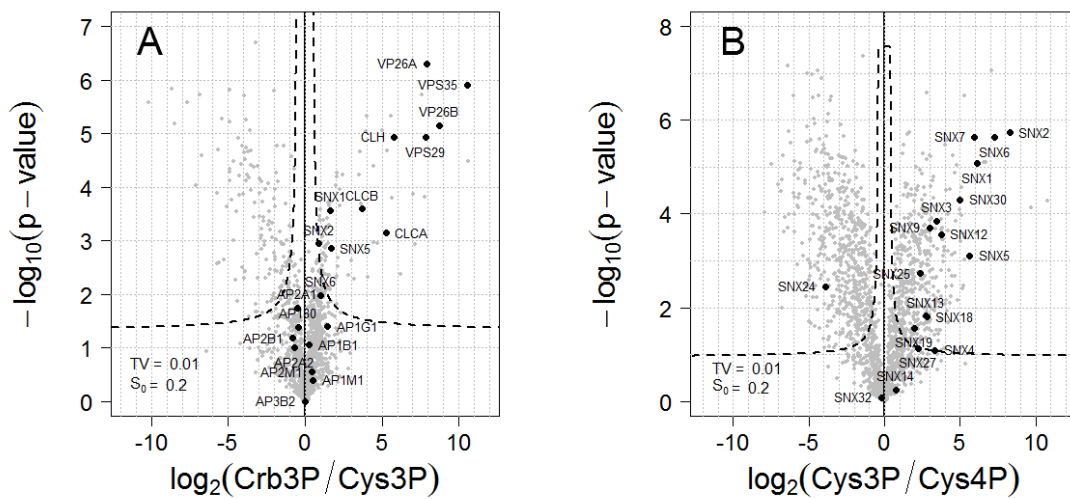
Figure 2.31: A: Volcano plot of CRB vs. Cys (PI3P). The cargo-recognizing components of the retromer complex VPS26A, VPS26B, VPS29, VPS35 are recruited due to the presence of the cytosolic domain of crumbs 2. B: Volcano plot of Cys3 vs. Cys4. The phosphoinositide-recognizing components of the retromer complex SNX1, SNX2, SNX5, SNX6 are recruited due to the presence of PI3P. Recruitment of those sorting nexins is further enhanced by the presence of the tail indicating coincidence detection of PI3P and VPS-proteins.

We validated the recruitment of VPS35 by western blotting before phenotypes were studied on *Drosophila* larvae (Figure 2.32) (Pocha, 2011).
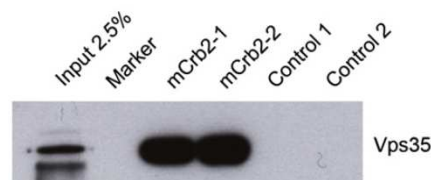
Figure 2.32: Duplicate recruitment experiment with PI3P-liposomes with crumbs2 or without (controls). The western blot was probed with an anti-VPS35 antibody. VPS35 was highly abundant in the cytosol (input) and recruited in the presence of crumbs, while not detectable on control liposomes.

With this study at hand we had clear evidence that crumbs is trafficked by the retromer complex.

The data bare many more interesting proteins that have been recruited in that context, like members of the WASH complex, which is known to associate with retromer and others with less obvious functional relations, like ELAV3, a factor in neuronal differentiation and maintenance. All proteins found to be significantly recruited by crumbs are listed in the supplemental table S.4. Supplemental table S.5 denotes sorting nexins found to be recruited by bare PI3P liposomes with respect to PI4P liposomes.

2.4.5 APP coupled to various phosphoinositide-containing liposomes. The function of the amyloid precursor protein (APP) is still a matter of discussion. It is ubiquitously expressed and has been implicated to diverse processes like the import of copper (Treiber, 2009) or synapse maintenance (Priller, 2006). The major focus on APP research, however, is its role in the formation of amyloid plaques in brains of patients with Alzheimer's disease. APP is a member of the small, extremely well-conserved APP family of type-I membrane proteins, consisting of APP, APLP1 and APLP2 (Suzuki, 2000). APP is the only protein among this paralogues that is causing the amyloidogenic Aβ peptide. It is ubiquitously-expressed with various isoforms, with the APP695-isoform as the predominant expressed in the brain (Figure 2.33-A). Various post-translational modifications are executed on APP. The most profoundly are proteolytic cleavages in the transmembrane region. Two cleavages, caused by α-secretase or β-secretase occur alternatively before a γ-secretase-cleavage. The amyloidogenic Aβ peptide is the result of the combined β- and γ-secretase cleavage (Figure 2.33-B) and an overproduction of the Aβ peptide is thought to be a major factor in the development of Alzheimer (Burgos, 2010). Various functions have been elucidated for the cleavage products, such as the targeting of the intracellular domain (AICD) to the nucleus and its involvement in gene regulation due to the binding of a diversity of factors.
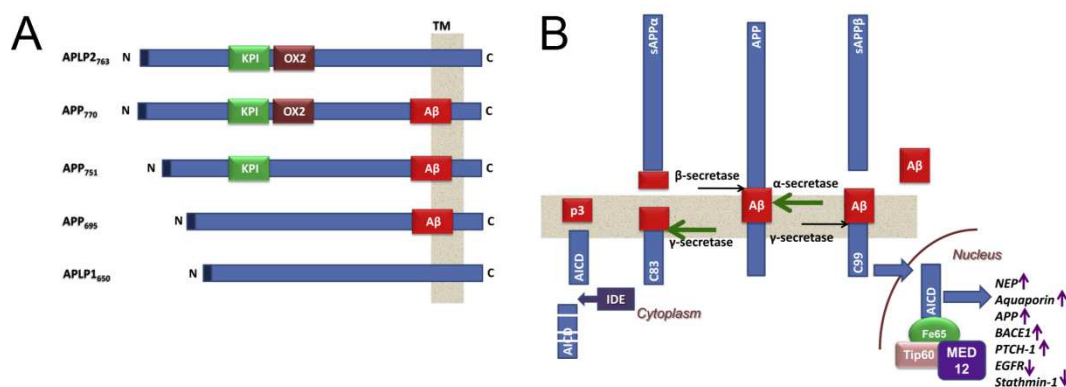
Figure 2.33: A – APP-family, consisting of APP and APP-like proteins 1 (APLP1) and 2 (APLP2). 3 isoforms of APP occur: APP770, APP751, APP695. APP695 is the major isoform in the brain. B – Proteolytic processing of APP695 (Nalivaeva, 2013).

APP co-localizes predominantly with EEA1 although there is also a population co-localizing with TGN46 (Burgos, 2010). This distribution of APP is not static, but a steady-state distribution as a result of APP-trafficking (Figure 2.36) (Caporaso, 1994; Haass, 1992). The two primary secretases are believed to act on different membranes, the plasma membrane (α-secretase) and the endosomal system (β-secretase) (Rajendran, 2012). Proteomic studies of Alzheimer patient's brains revealed also a deficiency in the retromer sorting pathway (Fjorback, 2012). The elucidation of trafficking mechanisms involved in Alzheimer's disease, as the trafficking of APP or the different secretases, is crucial though to understand the disease itself.

The G-test in table 2.6 indicates AP-4 mediated trafficking of APP. When we did this experiment, this was unknown and we chose to study APP in different membrane contexts. Expression, purification and coupling of APP were successful (Figure 2.34).
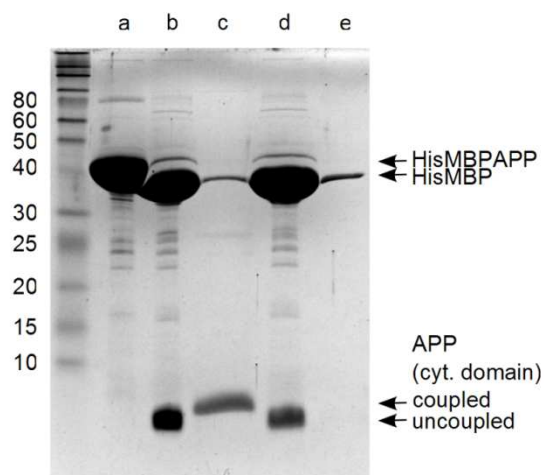
Figure 2.34: Coupling controls of APP. a: Purified expressed protein construct. b: Construct after Tev-cleavage. c: Final liposome pellet used for recruitment. d: Supernatant on top of pelleted liposomes after coupling. e: Supernatant after wash.

Recruitment experiments were performed with liposomes containing either PI3P, PI4P, PI35P2, PI45P2 or PI345P3 (Figure 2.35).
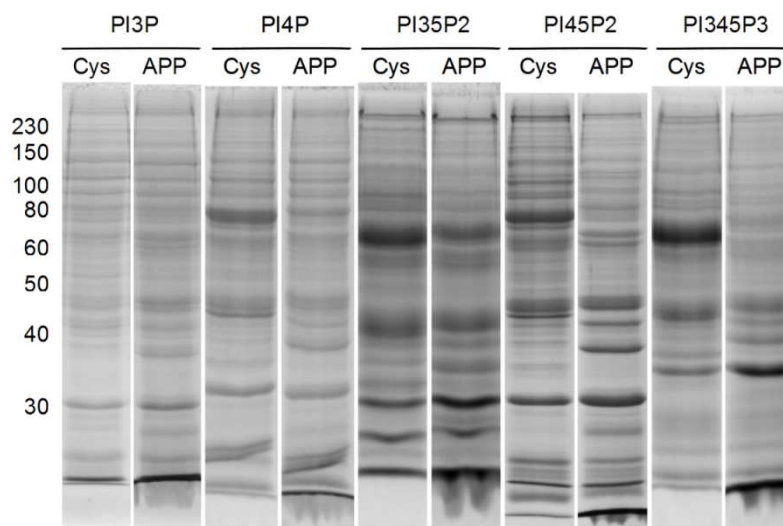
Figure 2.35: Comparison of recruitment samples. While triplicates were indistinguishable (not shown), gel lanes of samples and respective controls vary in several bands. Gel lanes of different phosphoinositide-experiments differ also to some degree.

Differential proteomic analyses were conducted, comparing samples of the APP-containing liposomes with the respective controls. All results are listed in the supplementary table S.6. Figures 2.38 to 2.47 show selected proteins that are associated with certain processes or complexes.

We observed our results at first for known general interactors of the APP-intracellular domain (AICD). The amyloid precursor protein possesses the most intricate combination of putative trafficking signals among the transmembrane proteins tested. The AICD incorporates three sites that are known to interact with other proteins, an YTSI site representing a binding YXXØ motif of trafficking-adapter complexes, a NPTY site following
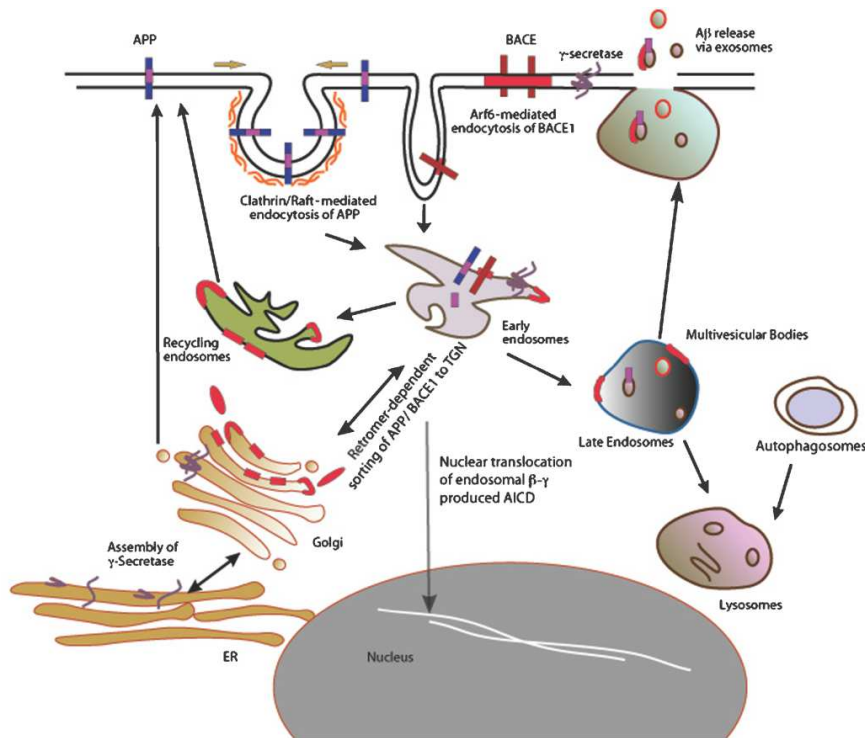
Figure 2.36: Membrane trafficking pathways in APP processing as summarized by Rajendran et al. (Rajendran, 2012).

the NPXY recognition motif of AP-2μ and Dab2 and overlapping with the latter, an (Y)KFFE binding site for AP-4μ (Table 2.5) (Burgos, 2010). The NPXY-motif is also part of the YENPTY motif, the binding site for APBA1 (X11), APBB1 (Fe65) and JIP1. This is remarkably due to their phosphotyrosine binding domain (PTB) (King, 2004), although the Y^{687}-residue (numbering according to the major APP695-isoform of the brain) inside the YENPTY motif is not phosphorylated in the brain (Suzuki, 2008).

We found APBA1, APBA2 and APBB1 clearly recruited in a phosphoinositide-independent manner, but we did not identify JIP1 though (Figure 2.37). This is most probably, because of an unmodified tyrosine on the -14 position upstream of the YKFFE motif, that has been shown to be necessarily phosphorylated for binding of JIP1 (Muresan, 2005). It was further shown that a phosphorylation of this specific threonine (T^{668}) diminishes the interaction with APBB1 (Fe65) by a conformational change in the AICD, while
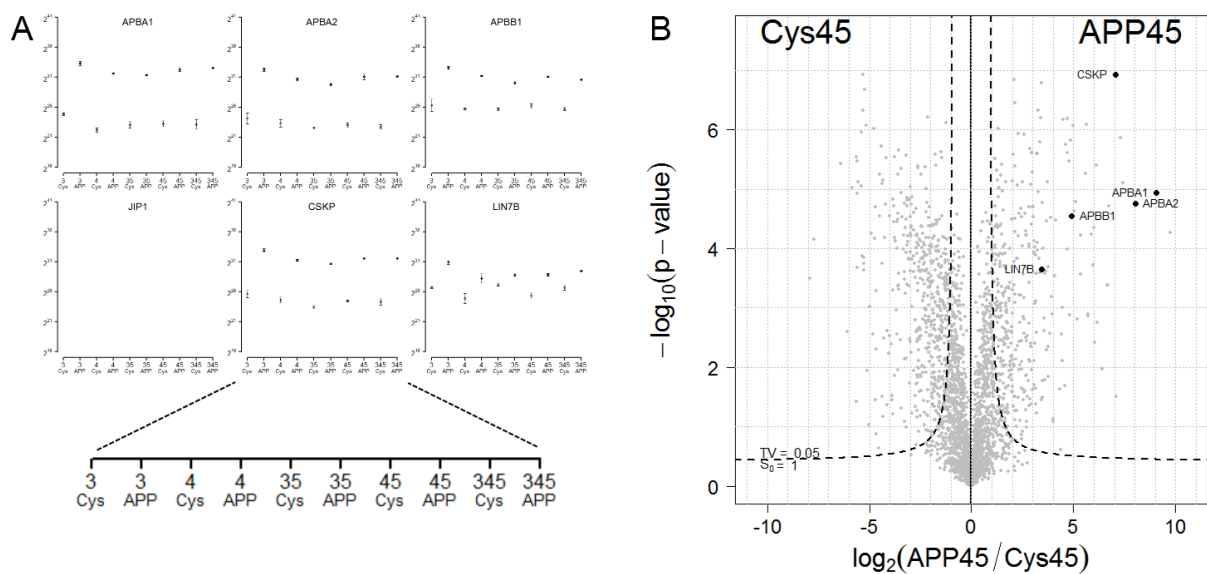
Figure 2.37: Known interactors of AICD. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI45P2-liposomes with and without APP.

it had no influence on the binding of X11 (Ando, 2001). It remains to be investigated, which parameters could be altered to enable in vitro phosphorylation of the T^{668}-residue during recruitment experiments. The interaction data base of the european bioinformatics institute (EMBL-EBI, www.ebi.ac.uk/intact) lists 85 direct or indirect interactors, or co-localizing proteins of APP. As the AICD can be targeted to the nucleus, interactions of the AICD can occur with proteins from the cytoplasm or nucleus. Filtering of this interactor list for such proteins results in a list of 57 putative interactors. Another popular resource for interaction information is the STRING-data base (http://www.string-db.org). 4 of the 57 in the EBI-list are also annotated in STRING as 'experimental evident'. Another 2 are missing in the EBI-list. A comparison with our experimental data showed that 8 proteins of this 59 proteins in total were recruited with APP (APBA1, APBB1, CALR, EXOC6, GSK3A, NF1, NSF, TBB5), while one (TPD54) yielded ambiguous results (recruited on PI3P-APP, diminished on PI35P2-APP liposomes). Recently, Chakrabarti et al. used in vitro pull-down experiments, followed by 2D-gel electrophoresis and MALDI-MS to identify interactors of the AICD (Chakrabarti, 2012). The experiment was constructed as a differential proteome analysis, comparing AICD pull-down samples with those obtained from control pull-down experiments with bare bead material. None of the accepted interactors as APBB1 (Fe65), APBB3, the Mint proteins APBA1/2/3 (X11α/β/γ), or AP-4 complex members (particularly AP-4 μ) were identified herewith suggesting a severe false negative rate. In addition, many likely false positive hits appeared among the mere 20 identified proteins, like secreted proteins. Notably, none of the potential AICD-interactors stated in Chakrabarti et al. was found to be recruited in our experiments, except for a vague recruitment of DYHC2/Dynein. Some proteins like APOA4 that showed binding to bare liposomes were even highly diminished in presence of APP. This demonstrates the benefit of both improved in vitro experiments as well as sophisticated mass spectrometry and post-acquisition workflows.

In addition to APBA1 and APBA2, we found also a considerable recruitment of CASK (CSKP) and LIN7B. CASK is present at low levels in all cells, but expressed primarily in the brain (Hata, 1993) where it localizes in presynaptic and postsynaptic densities (Hsueh, 1998). CASK (CSKP), Velis (LIN7A, LIN7B or LIN7C), and Mint 1 (APBA1) form a tripartite complex (Tabuchi, 2002) which is therefor a synapse-specific heterotrimeric complex. This specificity is ensured through a binding of CASK to the cytoplasmic tails of several cell-surface proteins, like neurexins (Figure 2.38) (Tabuchi, 2002). A direct binding to APP is so far not reported. An alternative for APBA1 in the complex is Caskin-1, which we did not identify (Tabuchi, 2002). If APBA2 could compete with APBA1 is not known, but both proteins share huge sequence similarities, mainly in the carboxy-terminal halves. An evolving model is that of APBA1 and APBA2 binding to the NPXY-motif of APP, with which CASK and LIN7B is forming a subcomplex that is linked in the pre- and post-synaptic regions of nerve cells via CASK to neurexins and other transmembrane proteins. This network could be further stabilized, by the binding of neurexins to APBA1 (Rogelj, 2006).
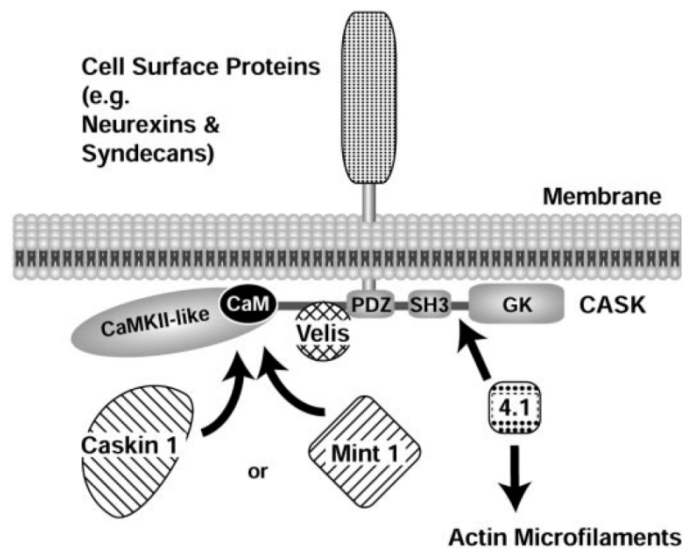
Figure 2.38: Heterotrimeric complex of Mint 1 (APBA1), CASK (CSKP) and Velis (LIN7A, LIN7B or LIN7C) bound to transmembrane receptors, like neurexins (Tabuchi, 2002).

The experiments with APP also revealed vast recruitment of NUMB and NUMBL (Figure 2.39). Both are known interactors of APP (Roncarati, 2002) and impart clathrin-dependent and -independent endocytosis by binding of AP-2α and various proteins of the epsin15 homology domain family (Gulino, 2010). Clathrin heavy and light chains (CLH,CLCB) as well as clathrin-associated proteins were allways present, but not recruited by APP.

We exploited our results further for post-Golgi trafficking complex members. We received ambiguous results from the screening experiment and the sophisticated recruitment experiment for the recruitment of AP-1 and AP-2 (Figures 2.40 and 2.41). The major differences between both types of experiments - except for the mass spectrometry
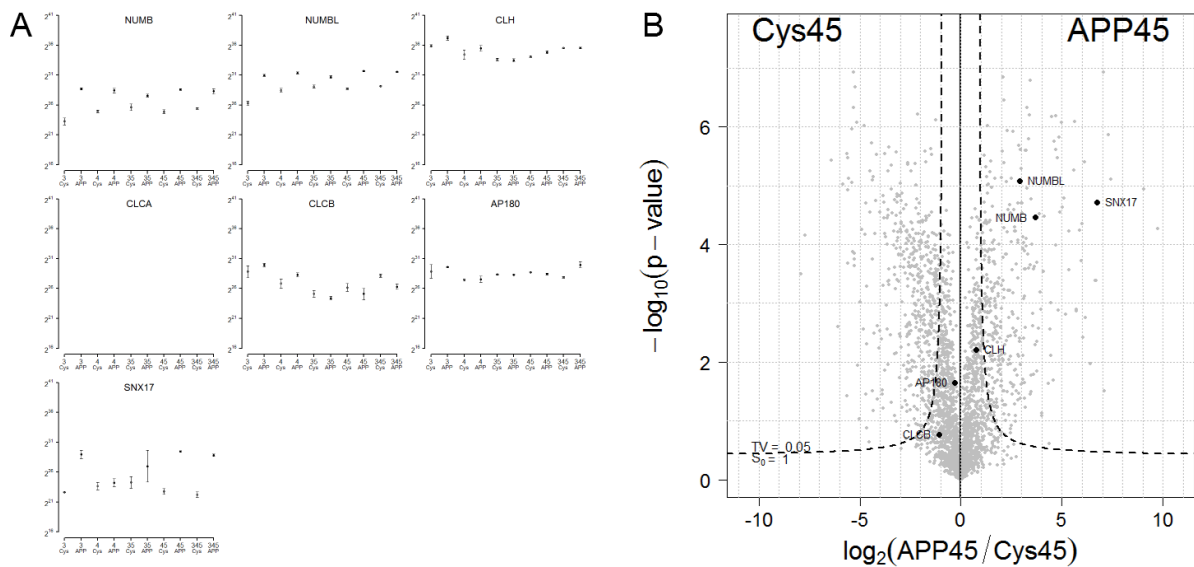
Figure 2.39: Known interactors of APP (NUMB, NUMBL) and other proteins functioning in endocytosis. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI45P2-liposomes with and without APP.

and post-acquisition workflows - was the use of pig brain cytosol in the sreen and mouse brain cytosol in the sophisticated approach, a recruitment time of 1 hour versus 20 minutes, 150 µM versus 1.5 µM GTP-µS and the addition of latrunculin B in the sophisticated approach to prevent actin polymerization which hampers isolation of the liposomes from the cytosol. While we achieved recruitment of AP-2 (and AP-1) in the screen, this was only slightly, but not significantly the case in the sophisticated experiment. However, the desribed experimental differences did not interfere with the capability of gE to
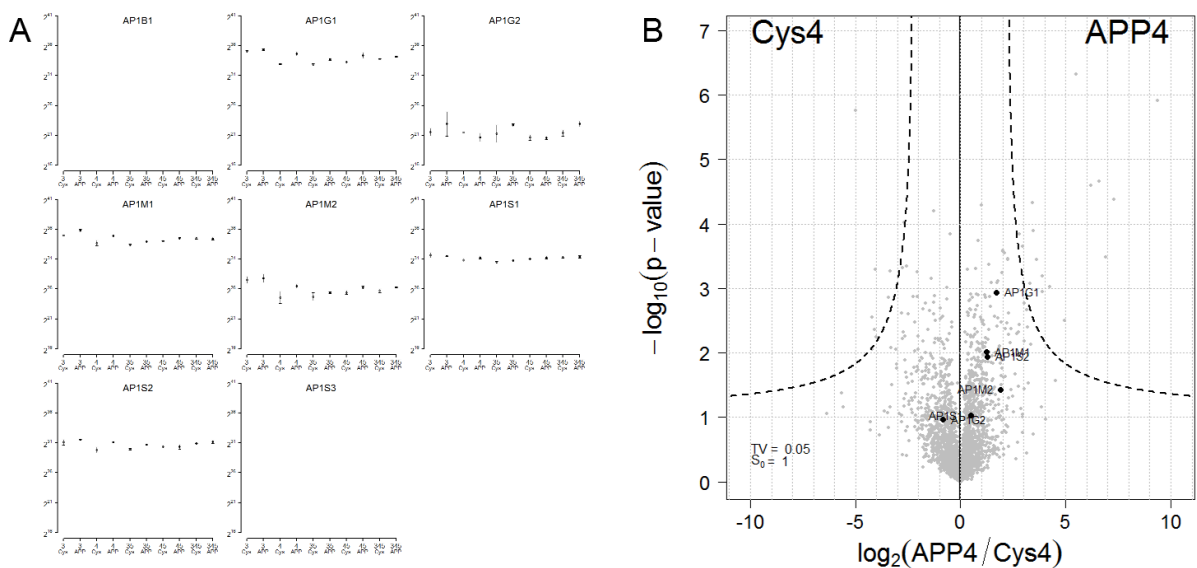
Figure 2.40: Proteins of the AP-1 trafficking complex. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI4P-liposomes with and without APP.
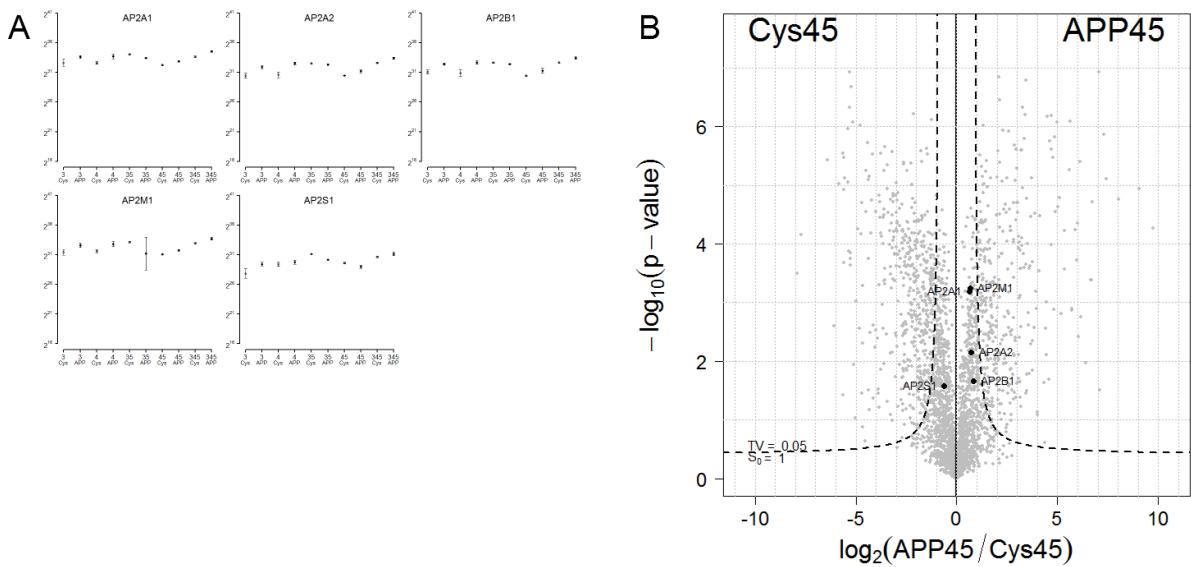
Figure 2.41: Proteins of the AP-2 trafficking complex. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI45P2-liposomes with and without APP.

recruit AP-1 and AP-2. The much higher concentration of GTP-µS in the recruitment screen on the other hand could have added up with a weak capability of APP to recruit AP-1 and AP-2. Members of the AP-3 complex were not recruited (Figure 2.42). In the recruitment screening experiment, we found indication of AP-4 recruitment with the identification of AP4B1, AP4E1 and AP4M1 with one of these (AP4B1) significantly recruited (Figure 2.43). When we did this study, no AP-4 cargo was identified so far. A short time later, a publication of Burgos et al. (Burgos, 2010) demonstrated that APP is trafficked by AP-4, providing us with the biological validation experiments. The improved recruitment assays elucidate
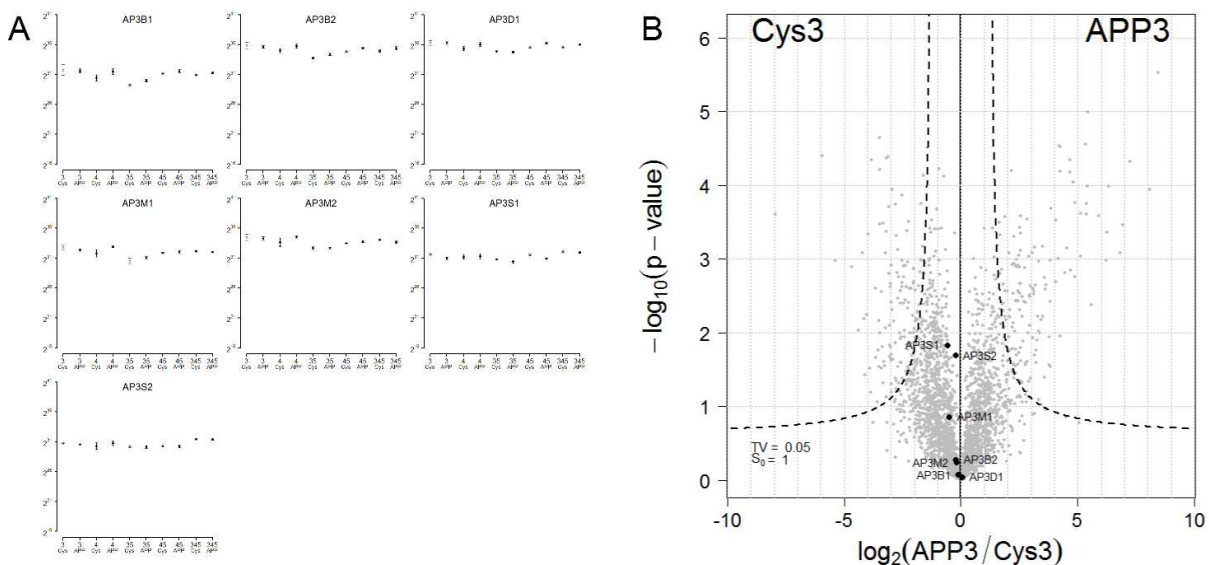
Figure 2.42: Proteins of the AP-3 trafficking complex. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI3P-liposomes with and without APP.
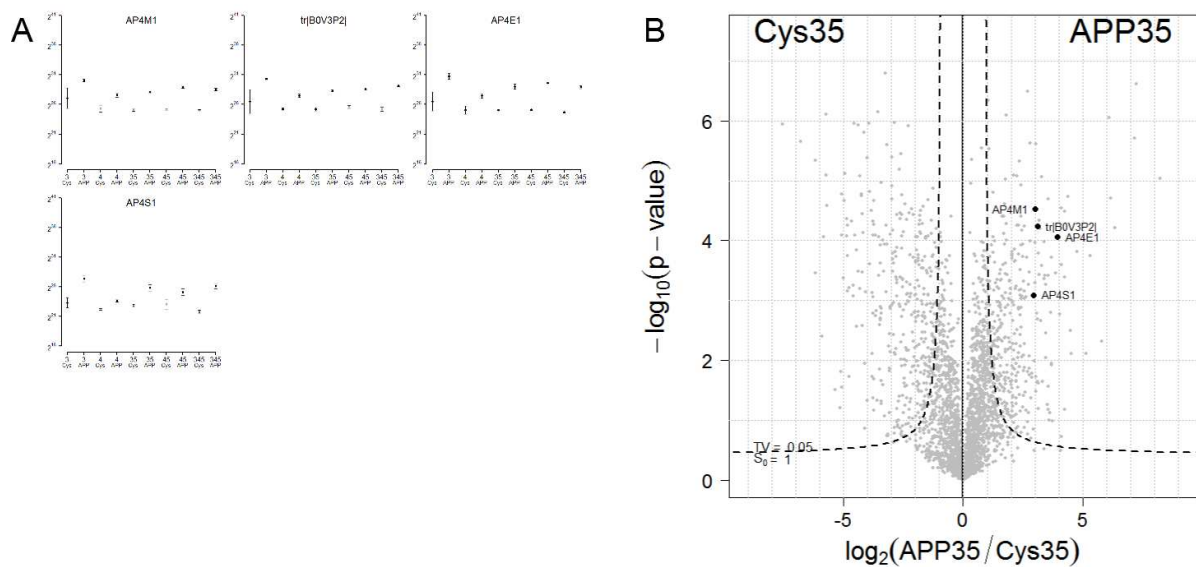
Figure 2.43: Proteins of the AP-4 trafficking complex. The former UniProt-Trembl entry B0V3P2 is meanwhile annotated as AP4B1. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI35P2-liposomes with and without APP.

clear recruitment of the whole AP-4 complex with no predilection for a specific phosphoinositide. AP-4 localization is known to be largely restricted to the trans-Golgi network, whereas under normal conditions APP localizes predominantly to endosomes. Upon disruption of the APP-AP-4 binding, APP is redistributed to the TGN, suggesting a TGN to endosome AP-4 mediated trafficking route of APP (Burgos, 2010). Burgos et al. discovered the AP-4 involvement in the trafficking of AP-4 in a hypothesis-driven discovery approach when they used yeast-two-hybrid assays to prove for interaction. The strength of the recruitment assay is that it is discovery-driven and unbiased, except for the selection of the receptor and cytosol. A fifth adapter complex (AP-5) was recently found in a genetic screen (Hirst, 2011). No cargo has been identified so far, but lately two additional members of the AP-5 complex, SPG11 (SPTCS) and SPG15 (ZFY26), were discovered (Hirst, 2013). We found YK046 (AP5B1) together with SPTCS and ZFY26 recruited with APP (Figure 2.44). None of the other three expected AP-5 complex members were identified (AP5Z1, AP5M1, AP5S1). This could indicate, that the three complex members found represent a cargo-binding core complex and that experimental factors (maybe PI5P ?) are missing which would allow to assemble the holo-complex. If AP-5-mediated trafficking of APP could be confirmed by additional experiments, this would thus not only add to the understanding of APP-trafficking but also represent the first AP-5-cargo identified and open new possibilities to study this adapter complex. We also found all members of the retromer complex recruited with APP (VPS26A/B, VPS29, VPS35, SNX1/2/5/6) (Figure 2.45). Here, the phosphoinositides show great influence. The sorting nexins seem not to be recruited when PI3P was used, but this is clearly due to the enhanced binding to the control liposomes (Figure 2.45-A). Recruitment was lowest when PI4P was used. Involvement of the retromer complex is known for long, though the underlying mechanism remains a matter of discussion. Recently, Fjorback et al.
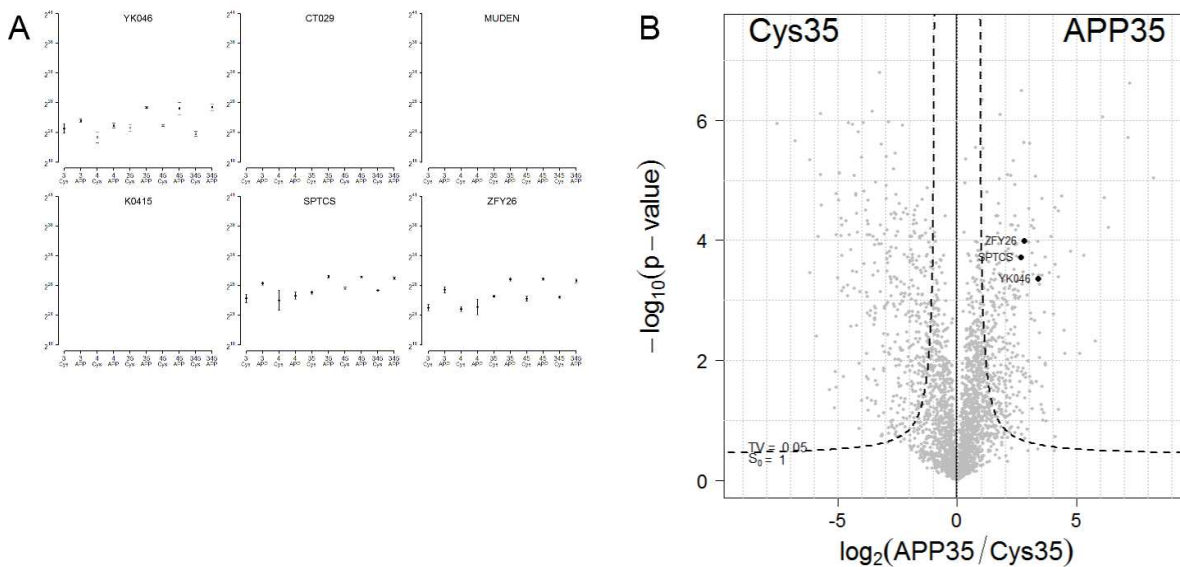
Figure 2.44: Proteins of the AP-5 trafficking complex. The former UniProt-Trembl entries are meanwhile exchanges by AP5B1 (YK046), AP5S1 (CT029), AP5M1 (MUDEN) and AP5Z1 (ZFY26). Former synonyms for SPTCS and ZFY26 were SPG11 and SPG15. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI35P2-liposomes with and without APP.

proposed SorLA, a type-I transmembrane protein as a molecular link for retromer-dependent sorting of APP (Fjorback, 2012). The absence of other transmembrane proteins in the recruitment confutes the necessity for such a mechanism. Retromer is believed to traffick APP from the endosomes to the trans-Golgi network in a long-range axonal transport and a VPS35-deficiency was associated with Aβ-increase (Bhalla, 2012) underlining the importance of retromer-mediated APP trafficking. Sorting nexin 17 (SNX17) has also been shown to associate with AICD, Fe65 and X11 (Lee, 2008). In addition, Lee at al. identified Dab2 to bind to the same YXNPXY-motif of APP and they proposed that both proteins regulate APP endocytic trafficking. SNX17 possesses a PX domain that imparts binding to PI3P. We found SNX17, but not Dab2, recruited with APP, particularly on PI45P2, PI345P3 and PI3P-liposomes that represent the early stages of endocytosis.

In vesicular trafficking, also membranes are transported and fuse with the target membrane. Embedded phosphoinositides that represent the identity of the membrane need to be converted thus. We found the PIKfyve complex (FYV1, FIG4, VAC14), that facilitates the interconversion of PI3P and PI35P2, recruited on APP-liposomes (Jin, 2008) (Figure 2.46). We propose that APP in turn regulates the PIKFyve complex through this recruitment and thus regulates the endosomal phosphoinositide metabolism (Balklava, 2013a, submitted).
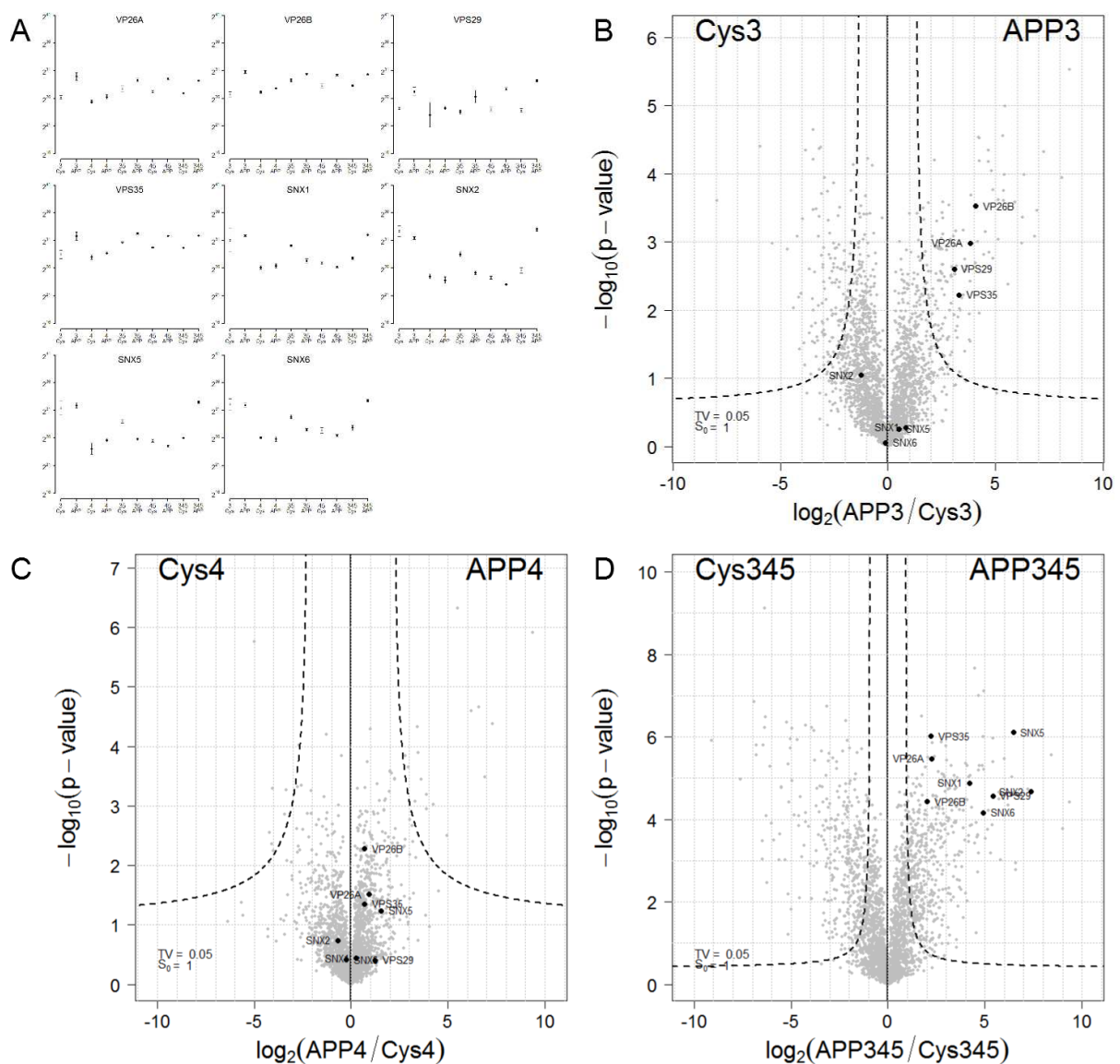
Figure 2.45: Proteins of the retromer complex. A - Distribution of LFQ Intensities. B,C,D - volcano plots of the t-tests of PI3P, PI4P and PI345P3-liposomes with and without APP.

Vesicular trafficking and signaling pathways are often intertwined (Clague, 2001; Shilo, 2011). We found the mTOR complex 1 (MTOR, LST8, RPTOR) recruited on APP-liposomes (Balklava, 2013b, submitted) (Figure 2.47). Interestingly, localization of the mTOR complex to endomembranes has been proposed as a prerequisite for Rheb-association and activation of downstream targets (Buerger, 2006). No direct liaison between APP and the mTOR complexes 1 or 2 have been hypothesized so far, although functional bonds between Aβ overproduction in Alzheimer's disease and mTOR-signaling have been described (Caccamo, 2010).
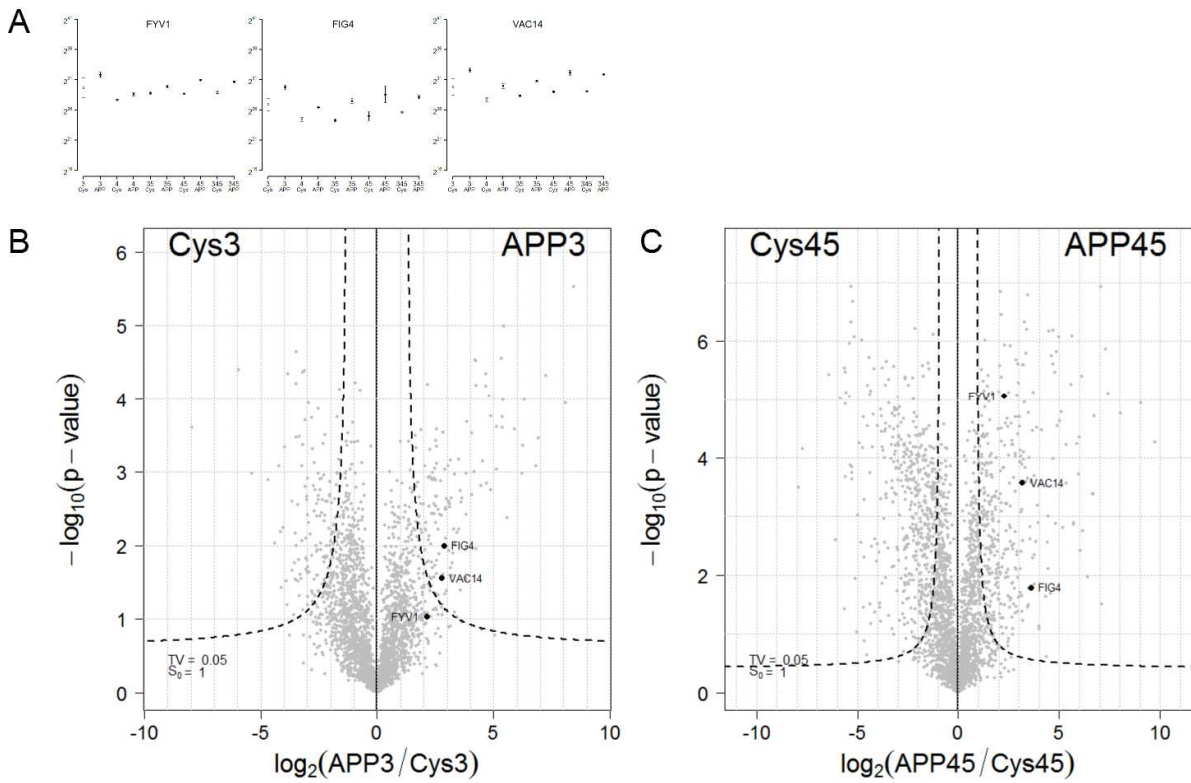
Figure 2.46: Proteins of the PIKfyve complex. A - Distribution of LFQ Intensities. B,C - volcano plots of the t-tests of PI3P and PI45P2-liposomes with and without APP.
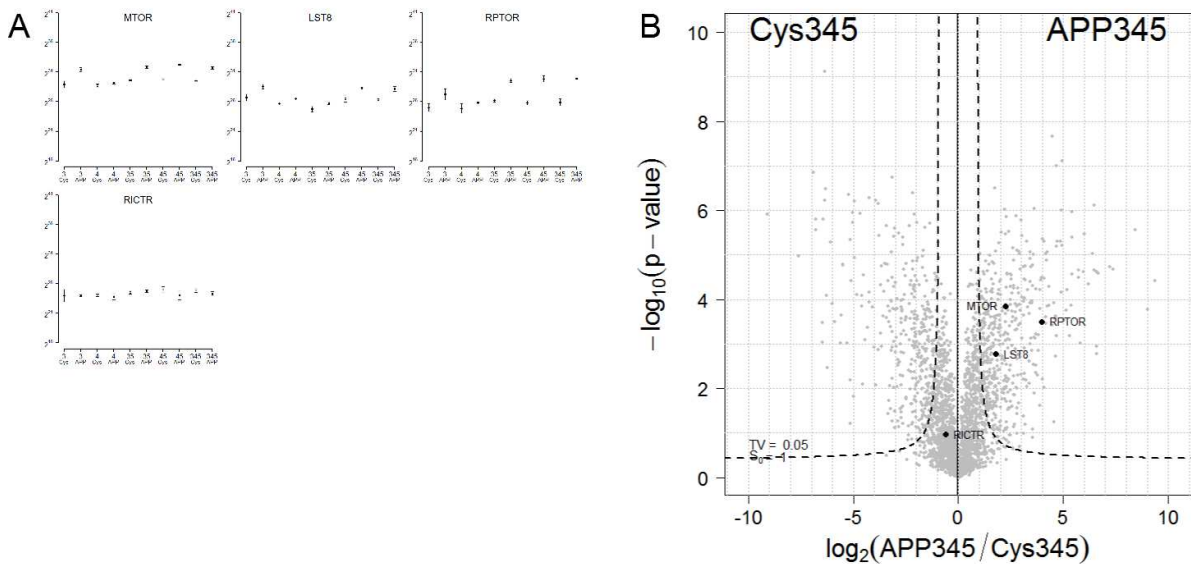
Figure 2.47: Proteins of the mTOR1/2 complexes. A - Distribution of LFQ Intensities. B - volcano plot of the t-test of PI345P3-liposomes with and without APP.

3. DISCUSSION

3.1 Sample composition

3.1.1 Label-free mass spectrometric parameters. Proteomics is the science and associated technics of analyzing the proteome, the entity of all proteins present in a given cell type or sample. The matter of proteomic analyses are essentially protein identifications and quantifications, but can also be post-translational modifications or protein-protein interactions. Probably the most important high throughput technique in proteomics is mass spectrometry as it allows for unbiased identification and quantification of a few thousand proteins in parallel. With the instrumental developments of the last 10 years or so, the achievable resolution was increased and so was throughput and reliability of protein identification. However, protein quantification was hampered by physico-chemical effects that distorted direct derivation of protein abundance from corresponding tryptic peptide measurements. A successful strategy to overcome these obstacles is the chemical labeling of proteins and since its introduction in 2002 notably the metabolic labeling by SILAC (Ong, 2002). On the other hand, labeling approaches are expensive and often impede mass spectrometric analyses or screening experiments. They introduce biases due to imperfect labeling and are sometimes not applicable. A multiplicity of label-free approaches was additionally developed. Those approaches are based on intensity values or indirect parameters which are subsequently normalized in different aspects and thus represent engineered 'rule of thumb'-parameters. To implement a label-free mass spectrometry platform, we saw the need for a systematic comparison of label-free mass spectrometric parameters. We found that the intensity (MQ) as it is estimated by MaxQuant performed best in terms of dynamic range and linearity over four orders of magnitude, even in presence of a complex proteomic background. When we assayed the accuracy with which amounts of standard proteins have been estimated, the MaxQuant-intensity and its normalized values, the label-free quantification intensity (LFQ) and the intensity-based absolute quantification values (iBAQ) were clearly superior to the purely engineered values peptide number (PN), spectrum counts (SC), exponentially-modified protein abundance index (emPAI) and normalized spectral abundance factor (NSAF). These experiments also showed that the accuracy was even better when the values were compared to mass amounts instead of molar amounts, suggesting another level of normalization. Moreover, outliers were minimized allowing even a reliable absolute quantification of all proteins tested. In a detailed view, iBAQ values slightly outperformed LFQ values here and both were clearly dominating before pure MQ values. However there is a basic difference between iBAQ and LFQ values. LFQ-values are normalized in a way that reduces sample-to-sample variations, allowing the comparison of different samples. To achieve this with iBAQ values, they must be further normalized with spiked-in UPS standards as was demonstrated in the initial publication of Schwanhäusser et al. (Schwanhäusser, 2011). The costs of these spike-in standards are comparable to SILAC experiments. Therefore, we use LFQ values to estimate protein abundance for both, compositional analyses of single samples and differential proteome analyses of two or more samples. While we cannot completely

exclude single principle outliers through atypic physico-chemical effects when different proteins are compared within a sample, comparing different groups of proteins seems justifiable. This is for instance the case when the success of a certain biochemical enrichment procedure has to be demonstrated.

3.1.2 Enrichment analysis. In mass spectrometry based proteomics, sample composition can be studied by either targeted approaches, where only predefined proteins are monitored by the mass spectrometer, or shotgun approaches where as many proteins as possible are traced. Targeted approaches allow label-free absolute quantification of proteins with high sensitivity even in complex proteomes. It is the method of choice when specific questions are addressed like in doping analysis. The approach is technically restricted to the parallel monitoring of up to 100 proteins per LC-MS run. For each protein of interest a method has to be established were a set of tryptic peptides that are unique for this protein isoform have to be selected and their retention time and mass spectrometric detection and fragmentation behavior have to be studied on synthetic analogs. When constructing a method for the parallel monitoring of different proteins those peptide probes should also differ in retention time and mass, as low resolution mass spectrometers are used, like triple quadrupole instruments. Such a method development is time and cost intensive although major progress took place in the last years (Picotti, 2012). Shotgun approaches are commonly appointed unbiased, meaning that phenotypes can be studied – in theory – on the whole proteome. In a shotgun experiment up to 4000 proteins are identified from a complex proteomic sample (Nagaraj, 2012) leaving the majority of expressed proteins unidentified. The typical shotgun approach uses a TopN-program that conducts the mass spectrometer to automatically select the N most intensive peptides in the MS1 spectrum to be fragmented for MS2. As the most abundant proteins are more likely generating higher intense MS1 peaks they are more likely being identified. This results in a bias of identified proteins for the more abundant proteins in a proteomic sample. Not all of the proteins that are identified can be quantified. The reasons for this are manifold and differ in the quantitative methods. In shotgun experiments, low intense proteins are often not identified in replicates and hence are deselected because statistical tests are mandatory. This results in a further bias in the selection of the more abundant of the identified proteins for being quantified as well (Figure 3.1-A). If specific questions are addressed to an experiment like the presence or absence of specific marker proteins, a chemical enrichment procedure can be used to increase the abundances of those proteins within the proteomic sample and either elevate those proteins into the observation window or increase their intensity within all the hitherto identified proteins (Figure 3.1-B). A question that arises is how to define and detect enrichment success. Pie charts are broadly used to demonstrate sample composition reflecting the relative numbers of proteins assigned to certain gene ontologies. All parts of this pie should add up to 100%.
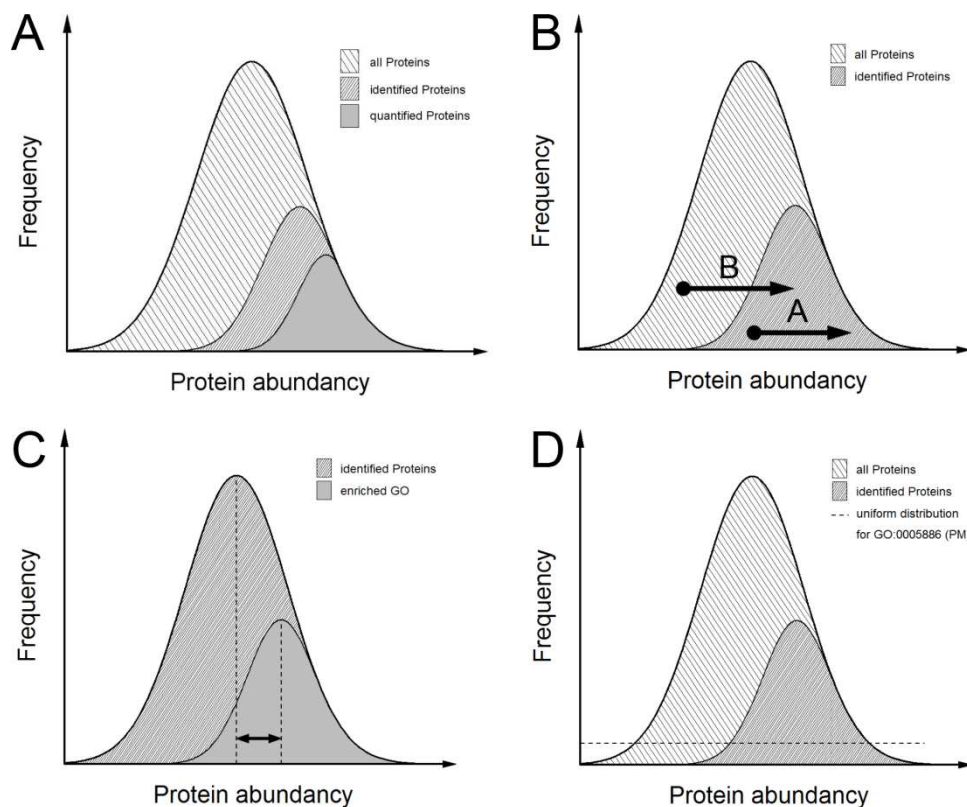
Figure 3.1: Histograms to access protein enrichment. The proteins are grouped for their abundances and the relative frequency at which different proteins in the abundancy bins occur is plotted. A - Bias of detection. In shotgun proteomic experiments not all proteins present can be detected. The subpopulation of identified proteins is composed of more abundant proteins than the average of all proteins present. Likewise, the subpopulation of quantified proteins is biased to the more abundant proteins among the identified. The reason for this prevalence to identify more abundant proteins is that more abundant proteins are more likely yielding more intense peptide precursors and are hence more likely selected for fragmentation. Likewise, more abundant proteins are more likely generating more and higher scoring peptide identifications, explaining the bias (modified from: Bantscheff, 2007). B - Observation of enrichment. Upon enrichment proteins can be shifted either within the observation window (A) or into (B). Moreover, proteins which are not promoted by the enrichment procedure can be decreased in intensity and even leave the observation window. C - Enrichment analysis by comparing the mean intensity of a subpopulation of identified proteins with the overall mean (Pan, 2009; Cox, 2012). This leads to information about the sample composition but as a chemical enrichment procedure does not necessarily lead to such an intrasample shift of the intensity distribution the method's axiom seems improper. As an example, if an absolute enrichment would have been achieved leading to identify proteins from a specific gene ontology only, this method would yield no enrichment at all. D - Schematic representation of all 20256 proteins annotated in UniProt/Sprot and the 22.4% of this which are assigned as plasma membrane proteins (PM). As there is no prior information about abundances in the data bank, p-value-based enrichment calculations assume uniform distribution which is very unlikely to reflect the real situation. The strategy presented here is to calculate p-values for cell surface proteome enriched samples as well as for total proteome samples and compare p-values of respective intensity bins.

However, as proteins are assigned to many different gene ontologies there are redundant occurrences of proteins in different groups. Also, there is no information if a size of a group reflects enrichment or just a probabilistic sampling. When samples before and

after enrichment are compared, only those proteins that were unobserved in either sample effect the size of a sample group while proteins that are altered in abundance but were always observed do not. The latter is also the case for Venn diagrams which are commonly used to compare the number of proteins assigned to a specific group between two samples. The more sensitive the mass spectrometer and the more broadly a proteome is covered in the identification pipeline, the more indistinguishable two samples are in a Venn diagram. Cox et al. (Cox, 2012) uses a ranking of proteins by intensity in a histogram plot and defines enrichment as a subpopulation of proteins sharing a specific gene ontology with a mean value deviating from the mean of all other proteins. This is similar to Pan et al. (Pan, 2009) were the subpopulation mean is compared to the mean of all proteins identified (Figure 3.1-C). This approach has its weakness when proteins of a specific GO-assignment enter the observation window in a number that could not have happened by chance, but do not deviate in their distribution mean from the overall. For instance, if an absolute enrichment would have been prepared leading to identify proteins from a specific gene ontology only, this method would yield no enrichment at all. Cox et al. expanded their approach to a two dimensional form which allowed comparisons of different samples and even from different types of experiments. They converted the distance from the mean into a ranking value and prepared multivariate analysis of variance (MANOVA). In that way the approach still relies on intrasample distributions and does not implicate comparisons with the 'gene universe' (Falcon, 2007), i.e. the entry of proteins into the observation window.

Our approach is based on intensity-ranked even-sized contingents of proteins ('bins'), were accumulation of certain gene ontologies is tested for their likelihood to occur by chance. We define enrichment as the negative decadic logarithm of the resulting p-value and replace the simple number of proteins in histogram plots herewith. This enables different gene ontologies to be compared as GO categories are weighted for their explanatory power. A problem that arises with such an approach is that we do not know the actual intensity distribution of a certain gene ontology within all proteins. If for instance, a gene ontology for the glycolytic pathway is studied, the assumption can be made that involved proteins generally occur in high copy numbers. But without prior information about all the protein abundancies we have to assume a uniform distribution for the calculation of the p-values (Figure 3.1-D). The problem is circumvented in an extended two dimensional form of the approach by comparing ranked intensity bins of samples after chemical enrichment and total proteomic samples. The approach comprises both, proteins that were entering the identification window and that once intensified inside due to an intrinsic ranking comparison.

3.2 Differential proteome analysis

3.2.1 Imputation. Missing values in shotgun proteomic data are very common and limit the reproducibility of protein identifications. The reasons for missing values are multifaceted [Karpievitch2012]: a protein is not present in an experimental state, a protein is not detected, although present in a detectable amount, or a protein is present in an amount below the detection limit. There are several ways to handle missing values: related protein hits could be deleted, which could enhance the false negative rate. On the other hand, if a protein is deteced in only a small number of samples this protein could be coincidental present through variabilities in the biological experiment, sample preparation or representing a contaminant, e.g. from previous runs on the LC-MS system. In this case it would be advisable removing the protein from the list to avoid a false positive hit. Missing values could also be ignored, but this causes a diversity of possible errors. Ignoring could exclude proteins from t-testing leading to false negatives. It could lead to overestimation of the relative presence in one state compared to another, or it could lead to underestimate standard deviations and hence overestimate the significance in a t-test. The latter would cause false positive interpretations. If a fixed value would be used for imputation, this could as well lead to an underestimation of standard deviations and likewise cause false positives. Thus, a sophisticated imputation algorithm should be used. We estimated the distributions of standard deviations for different intensity classes to create a Gaussian distribution around the detection limit with a standard deviation estimated from the low intensity region to impute missing values from random draws.

3.2.2 Two-sample comparison. The simple analysis of a biological sample in a single condition doesn't offer much detailed information. It is hence more frequently of interest to compare the effects of one condition with another, like for instance when cells treated with a certain drug are compared with untreated control cells. A common problem in experiments that do not require a differential analysis on the first view, such as pull-down experiments, that are aimed to identify protein interaction partners, is the co-occurrence of unspecific background proteins. This is, because such experiments include the use of solid matrices to which many proteins bind and often represent the huge majority of identified proteins. A differential proteome analysis of pull-down samples in presence and absence of the bait is the most effective way to define true interaction partners then.

When recruitment experiments were previously performed (Crottet, 2002; Baust, 2006), supposed 'true' interaction partners were manually selected from a background proteome. We started our recruitment experiment project with a screening of different receptors in different conditions and used blank liposomes without receptor peptides as controls. We used spectral counting to compare protein levels in samples and controls, as spectrum counts are simple to acquire. We found that a definition of thresholds for minimal required spectrum counts and spectrum count ratios was not obvious and we thus implemented a

G-test-based approach to put this on a statistical base. It was previously shown that a sampling statistic based on the G-test is a robust approach to detect differential levels of proteins between two samples (Zhang, 2006). The method first normalizes the spectrum counts within sample and control and allows a Chi-square distribution to access p-values. The method is fairly good working, an improvement of the spectral counting-based differential proteome analyses and the method of choice for screening experiments. Nevertheless, we found that the sensitivity was low and suspected a severe false negative rate as many known multicomponent protein complexes appeared only partially. One reason for this might be the random sampling effect, meaning that peptides identified in one mass spectrometric run could be missed by another run, although present. This can happen for instance when co-eluting peptides suppress others during ionization, or during precursor-selection. The omitting of single or view peptides reduces the spectrum counts and thus reported quantitative level of the corresponding protein. The random sampling effect can be minimized by performing replicate runs and the use of statistical tests that can benefit from replicates, such as t-tests (Liu, 2009). As we favored MaxQuant's LFQ-intensity values for assessing protein levels, we evaluated the label-free differential proteomics workflow that is implemented in MaxQuant's accompanied statistical data evaluation package 'Perseus'.

The test statistic used in Perseus is a modified t-test. It is reported to be based on the significance analysis of microarrays (SAM) (Tusher, 2001). A constant (s_0) is added to the denominator of the t-test function that prevents indefinite high values when the standard deviation tends to values close to zero, as can coincidentally happen with small numbers of replicates:

$$d(i) = \frac{\bar{x}_S(i) - \bar{x}_C(i)}{s(i) + s_0}$$

$d(i)$ - t-test value of protein i ; $\bar{x}_S(i), \bar{x}_C(i)$ - means of expression values of protein i in states S (sample) or C (control) ; s_0 - fuse constant of SAM ; $s(i)$ - pooled standard deviation of intensity values of protein i, calculated as:

$$s(i) = \sqrt{\frac{\frac{1}{n_S} + \frac{1}{n_C}}{n_S + n_C - 2} \cdot \left\{ \sum_{j=1}^{n_S} \left(x_{S,j}(i) - \bar{x}_S(i) \right)^2 + \sum_{k=1}^{n_C} \left(x_{C,k}(i) - \bar{x}_C(i) \right)^2 \right\}}$$

n_S , n_C - number of replicates in states S or C, with $v = n_S + n_C - 2$ degrees of freedom; $x_{S,j}(i), x_{C,k}(i)$ - j-th or k-th intensity value of protein i in states S or C

In the SAM-procedure the test statistic is calculated for every protein first. In a second step, the grouping of values to state S (sample) or C (control) is changed in a random and nonsense way. The test statistic is calculated again. This is repeated for all possible permutations of data-group associations. A significant changed protein should exhibit a

much higher test-value as the average of all nonsense test-values. Such occurrences are elucidated by a comparison of the test-value distributions in a Q-Q plot (Figure 3.2).
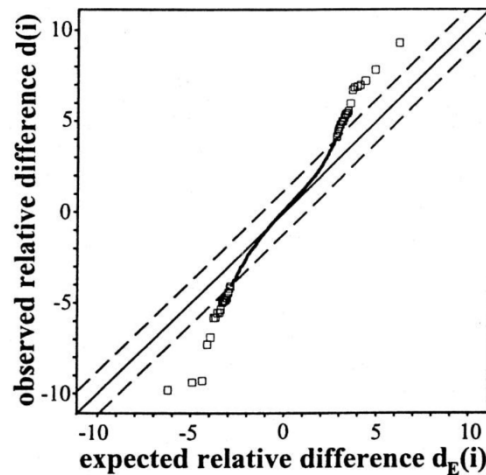
Figure 3.2: SAM-method to identify significant changing proteins. The distribution of the modified t-test values (d(i)) is compared to the distribution of corresponding median t-test values from all nonsense permutations ($d_E(i)$) of data-group associations in a Q-Q plot. Symmetric threshold lines in parallel to the bisecting line define significant changing proteins with real t-test values that largely differ from the random counterparts [Tusher2001, 11309499].

Threshold lines in parallel to the bisecting line are chosen to define significant changing proteins. These thresholds can be set in a way that a certain maximum FDR is not exceeded. The FDR is estimated from the permutations by counting the number of proteins classified as significant from all nonsense permutations that thus present false positives, calculating the median across all permutations and divide this by the number of all proteins in the original data (Draghici, 2013).

The precision of the permutation-based estimation of the FDR depends on the number of permutations, which is in turn depending on the number of replicates and can be calculated as follows:

⇨ The intensity values are written as a table header, e.g.:

1.4E11	2.1E11	9.2E10	2.4E10	8.3E09	1.8E10

⇨ The group-classifiers, S or C are written into the table and are permuted from line to line.

1.4E11	2.1E11	9.2E10	2.4E10	8.3E09	1.8E10
S	S	S	C	C	C
S	S	C	C	C	S
S	C	S	C	C	S
…	…	…	…	…	…

⇨ This represents a permutation of 6 elements with two times three equal elements

$$P_{n,n_S,n_C} = \frac{n!}{n_S! \cdot n_C!} \qquad \Rightarrow \qquad P_{6,3,3} = \frac{6!}{3! \cdot 3!} = 20$$

Thus, 3 replicates of controls and samples allow for 20 permutations only and a low precision of the estimated FDR, accordingly. Much more confident FDRs can be yielded with 5 replicates each (252 permutations), but on the costs of a nearly doubled measurement time and sample consumption.

Moreover, a fixed FDR is just another representation of a constant p-value threshold, which would appear in a volcano plot as a constant horizontal threshold line. Non-changing proteins with low ratios might pass this threshold as well and it is hence required to add additional vertical thresholds for minimal accepted ratios. An intuitive way to unify these orthogonal threshold lines are hyperbolic curves. This is basically what the modified t-test in Perseus does (Figure 3.3). The exact algorithm of Perseus' modified t-test is not published and a systematic assessment of this quantitative method is also not available. We used the universal proteomic standard 1 (UPS1) consisting of 49 equimolar proteins and spiked it into a total yeast proteome in a series of different amounts. An array of pairwise comparisons with the Perseus' modified t-test was performed to assess the test's accuracy when different ratios and overall amount of standard is varied. These experiments revealed a very good accuracy even for low standard amounts. The quantitative accuracy of the method was good for all ratios tested from 3x to 81x, although fewer proteins passed the significance thresholds for lower amounts or ratios.
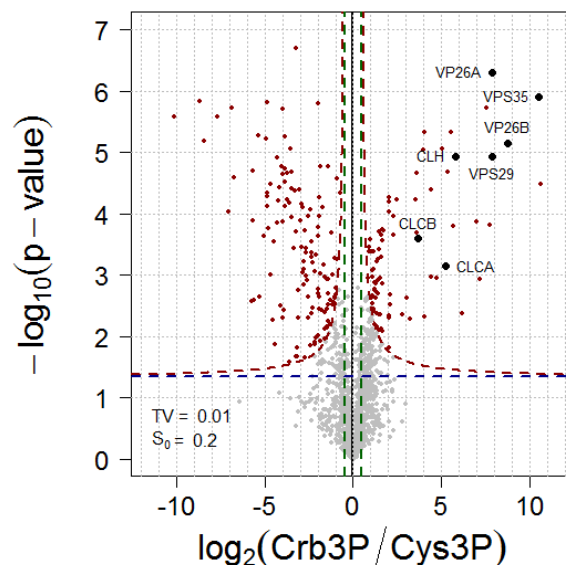
Figure 3.3: Threshold lines from Perseus' modified t-test in a volcano plot. The hyperbolic threshold lines (red) combine a horizontal p-value threshold (blue) with vertical ratio thresholds (green). Significant changing proteins are labeled in red or black (Retromer complex components and clathrin), while the majority of proteins that represent the background proteome is labeled in grey.

As shown in the results part, the standard deviations of LFQ intensities differ for different intensity regions and is largest for smaller values below 2^{24}. This raises the questions, if the reliability of the differential proteomic workflow to detect significantly changing proteins can be further enhanced by taking the absolute intensities into account. In his doctoral thesis, Christian A. Luber (LMU Munich, 2009) performed an experiment with lysates of HeLa cells and E. coli, where the E. coli lysates were used in an one to three ratio, while HeLa protein levels kept constant. He analyzed the overlap of ratio-distributions for the top-75% most intensive proteins (that represent the top-66% of decades of the dynamic range), the top-50% and the top-25% (representing nearly half of the dynamic range) and calculated the 3-σ deviations for this quantiles as 2.16-fold, 2.01 and 1.64. Thus a nearly complete separation of 3-fold ratio E. coli proteins and the unchanged HeLa proteins was only successful for the top-25% most intensive proteins, the upper half of the dynamic range (Figure 3.4). This experiment assessed the accuracy of the quantitative values, i.e. the LFQ-intensities, but not the accuracy of the Perseus' modified t-test. This means that only mean LFQ-values were used and standard deviations of individual proteins in the ground or lifted state were neglected. The Perseus' modified t-test on the other hand exploits this protein- and sample-specific scattering through its permutation-based estimation of p-values, thus leading to lower p-values for lower intense proteins that show greater variability in absolute LFQ-intensity values. A lower p-value could hinder these proteins from passing the threshold and could hence lower the true positive rate.
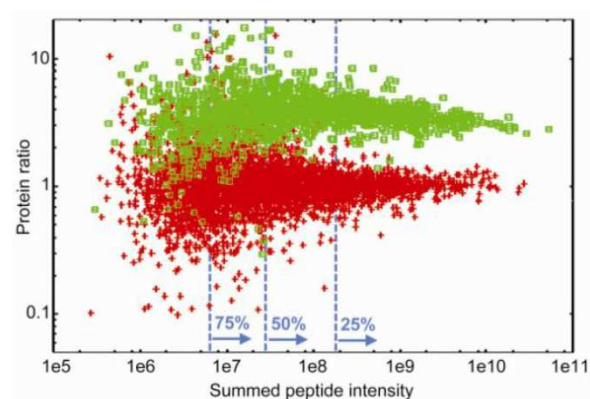
Figure 3.4: Accuracy of protein ratios estimated by LFQ-intensities and its dependency of the height of the sum intensity. Red – proteins from Hela cells in a 1:1 ratio, green – proteins from E. coli in a 1:3 ratio. [Dissertation Christian A. Luber, Ludwig-Maximilians-Universität München, 2009]

Another question to address was the influence of the test's parameters, the s_0-value and the TV (threshold value). The t-test and its derivatives, like the Welch'-modified t-test or SAM, are binary classifier systems and their performance can be studied in a receiver operating characteristic (ROC) analysis where the discrimination threshold is varied. A prerequisite to ROC-analysis is to exactly know which proteins are actually changing and which are actually not changing. This is the case in our experiments on the accuracy of Perseus' modified t-test where all positives are proteins of the universal proteome standard

(UPS1) whereas all yeast proteins are negatives. There are certain forms of ROC-analysis according to which parameters are compared. Probably the most common one is the comparison of the true positive rate to the false negative rate. The true positive rate is the fraction of true positives among all that are actually positives. In other words it's the fraction of all UPS-proteins that the test classified as changing among all UPS proteins identified. The true positive rate is sometimes also called sensitivity or recall. The true negative rate is the fraction of false positives among all that are actually negative. In our experiment, it is the fraction of yeast proteins that the test classified as changing among all yeast proteins identified. Sometimes, the sensitivity is compared to the specificity, which is equivalent to 1-FPR and represents the fraction of total negatives among all actual negatives. Other studies compare recall with precision. That is the fraction of true positives among all classified positives. Another term used is accuracy, which in this context is defined as the sum of true positives and true negatives divided by the number of all proteins identified.

We found that the s_0-value has a minor influence on the threshold curve and hence modified the TV for a receiver operating characteristic (ROC) analysis. Surprisingly, the assay exhibits an excellent characteristic, meaning that the TV can be selected in a great range before false positives increase. It should be adjusted not to conservative on the other hand, to allow as many true positives as possible. However, within 'real' biochemical experiments, as for instance the recruitment assays, broader background variability is common and the TV was found to be best selected between 0.001 and 0.05.

As a summary, the modified t-test implemented in Perseus was proven to be reliable in the detection even of subtle proteomic changes with excellent accuracy over a broad dynamic range. The ROC analysis with a varying threshold value showed that no thorough experiment-specific adaptation of test parameters is required and thus proved the methods usability for sophisticated screening experiments.

The gold standard of quantitative mass spectrometric approaches so far is SILAC. This is because label-free methods were thought to be generally impractical due to the different responses of the mass spectrometer for different peptides. Among the labeling approaches, SILAC is the most accurate one, because samples and controls are mixed early in the sample processing workflow and no chemistry is involved that introduces biases in the efficiency to label different peptides. But there is still a labeling efficiency below 100%. 100% pure labeled arginine or lysine is not available and in many cases costs hinder multiple passages of cultured cells in the labeled media. Impurities are introduced by fetal calf serum and arginine to proline (and back) interconversion occurs widely. The labeling efficiency has a huge impact on the quantitative accuracy. A labeling efficiency of 98% does not mean that the final ratio is underestimated by just 2%, as is demonstrated in the following calculation:

⇨ We assume, that a SILAC-experiment had been performed where cells of the control are not labeled, whereas cells that were perturbed are labeled with ($^{13}C_6$, $^{15}N_2$)-Lys.

H_E and L_E are the estimated peak heights of the heavy and light monoisotopic peaks. The theoretical height of the heavy peak H_T is underestimated due to imperfect labeling. On the other hand, the theoretical height of the light peak L_T is overestimated, because some molecules that were not labeled in the perturbed sample, appear in the light peak (Figure 3.5).
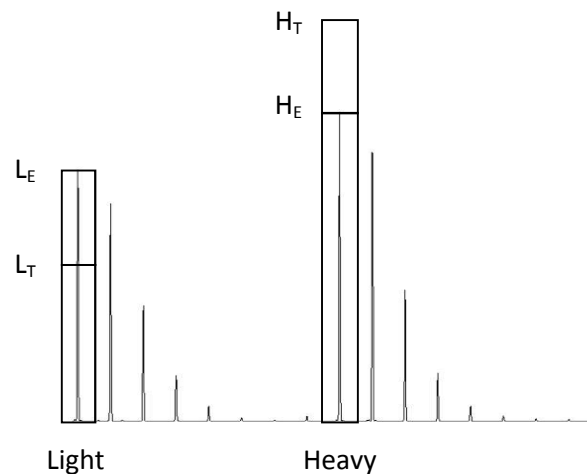
Figure 3.5: SILAC-Pair of heavy and light monoisotopic peaks. The estimated peak heights of the heavy and lights peaks under- and overestimate the theoretical peak heights because of a not achievable 100% labeling efficiency.

⇨ We like to find a relation, where the estimated ratio is a function of the theoretical ratio and the labeling efficiency E [in %].

(1) $$\frac{H_E}{L_E} = fct(\frac{H_T}{L_T}, E)$$

⇨ The estimated peak height is the theoretical peak height, reduced by the labeling efficiency.

(2) $$H_E = H_T \cdot \frac{E}{100\%}$$

⇨ The peptides that are not labeled due to the imperfect labeling appear in the light peak.

(3) $$L_E - L_T = H_T - H_E \Leftrightarrow L_E = H_T - H_E + L_T$$

⇨ Equation (2) and (3) combined:

(4) $$\overset{(2)\& (3)}{\Longrightarrow} L_E = H_T \left(1 - \frac{E}{100\%}\right) + L_T$$

⇨ The relation in (1) is yielded by dividing the equations (2) and (4):

$$(5) \xRightarrow{(2)\&(4)} \quad \frac{H_E}{L_E} = \frac{H_T \cdot \frac{E}{100\%}}{H_T\left(1-\frac{E}{100\%}\right)+L_T} = \frac{\frac{E}{100\%}}{1-\frac{E}{100\%}+\frac{L_T}{H_T}} = \frac{1}{\frac{\frac{L_T}{H_T}+1}{\frac{E}{100\%}}-1} \quad q.e.d.$$

⇨ This equation allows for a graphical analysis of the behavior of the estimated ratio, when different theoretical ratios are analyzed with a theoretical 100%, a maximal 98% and an often accomplished 93% labeling efficiency (Figure 3.6).
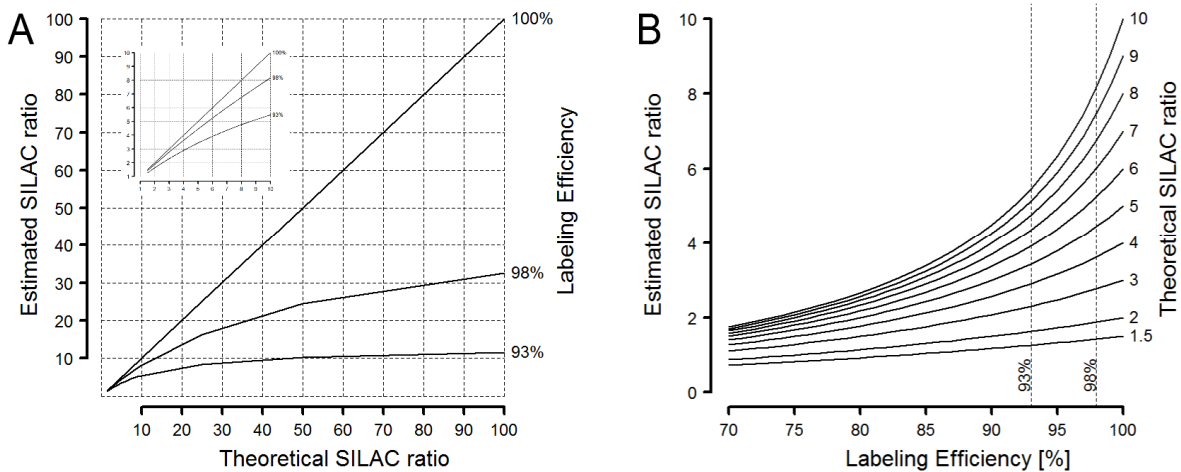
Figure 3.6: The accuracy of SILAC largely depends on the achieved labeling efficiency. A - Comparison of the estimated and theoretical SILAC ratios. B - Estimated SILAC ratios as a function of the labeling efficiency.

This problematic behavior of the SILAC method (and other methods based on stable isotopes), caused by imperfect labeling efficiencies is so far not a matter of discussion in the literature, but indirectly mentioned as in Altelaar et al. (Altelaar, 2012). The authors refer to Ong et al. (Ong, 2005) while stating : "...exact levels mostly depend on the purity of the amino acids". In the context of e.g. a 98% labeling efficiency and a huge, 100-fold actual ratio, this appears to be rather an understatement as the estimated fold-change is only one third here. Altelaar et al. performed experiments to assess the accuracy of SILAC to reflect actual ratios of 2- and 10-fold, resulting in 2.8- and 3.7-fold estimated ratios and hence overestimated (!) the actual ratios by 40% and 30% in the small ratio range (standard deviations were 0.35 and 0.68). Unfortunately, the reasons for these inaccuracies remained undiscussed there. Our assumption is that, the large underestimation of huge actual ratios has a great impact of the dynamic range of the SILAC method even for the best possible labeling efficiency of 98%. Quantitative analyses are thus restricted to small proteomic differences. The label-free approach on the contrary delivered reliable ratios in the dynamic range tested from 3x to 81x. Another flaw immanent in the SILAC method is the multiplying of MS[1] features. This reduces the identification of peptides in a shotgun proteomic approach and hence also reduces the number of proteins identified and quantified.

3.3 Application I – Analysis of plasma membrane proteins

Since we published the cell surface proteome of mesenchymal stem cells (Niehage, 2011), such cell surface proteome studies have been often requested by other groups. Their projects were either addressed to stem cells or cancer cells. In any case this is, because the mass spectrometric approach elucidates cell surface markers extensively and untargeted. These markers can be used within antibody-based technics, like fluorescence-activated cell sorting (FACS) or enzyme-linked immunosorbent assay (ELISA) that are commonly utilized in hospital labs. This opens new prospects in cancer diagnosis and pioneers regenerative medicine.

From a mass spectrometrist's point of view, such samples are challenging, as the hydrophobic membrane proteins are imbedded in a complex environment of lipids and a huge proteomic background. Moreover, sample amounts are very low and the degree of variation between replicates is high. In our initial project, we used spectral counting to register frequent cell surface proteins of mesenchymal stem cells. The suitability of identified proteins serving as stem cell markers was studied in broad FACS analyses hereafter. In a subsequent collaborative project, we studied the extracellular matrix generated by mesenchymal stem cells (Prewitz, 2013). Here, we were confronted with the necessity to elucidate the sample composition meaning the relative levels of different proteins. Moreover, we wanted to demonstrate the accumulation of proteins sharing certain characteristics, like their known affiliation to the extracellular matrix as judged by their annotations of respective gene ontologies. We used label-free quantification intensity values (LFQ) for the comparison of protein levels, but were not satisfied with the common representation of sample composition based on pie-diagrams. This was basically for three reasons. First, gene ontologies are not exclusive meaning that proteins can be associated to more than one gene ontology which cannot be reflected by pie-diagrams. Second, the levels of individual proteins are neglected. If 5% of the proteins would be annotated as extracellular matrix (ECM) and would represent 99% of the molecules in the sample, a pie-diagram would describe a 5% ECM content and thus wrongly present a non ECM sample. Third, the frequency of gene ontology annotations within all proteins in the UniProt data base is largely diverse. As an example, 1/8 of all 20272 human proteins annotated is assigned to the cytosol (2545 proteins, GO:0005829), but only 1/50 of the proteins is assigned to the extracellular matrix (395 proteins, GO:0031012). If 100 proteins were identified in a sample and 13 of which are assigned to the cytosol, 13 to the ECM, the probabilities of such identifications being coincidental findings differ largely: 49% for cytosolic proteins, but 0.0000074% for ECM proteins (estimated from a hypergeometric distribution). Such different explanatory powers of different gene ontologies are not accessed with pie-diagrams. We developed a two-dimensional mathematical model (see chapter 3.1.2) that is based on a hypergeometric distribution and implies individual protein abundances by their LFQ-values to assess accumulation of gene ontologies. As p-values are compared, the model allows comparing different gene ontologies and different sample

sizes. In a collaborative project with Dr. Jana Karbanová (group of Dr. Denis Corbeil, Biotec TU Dresden), we applied the compositional proteome analysis, enrichment analysis and differential proteome analysis to samples from a CSP-enrichment of dental pulp stem cells (CPS cells). The partners previously noted that such cells exhibit differences in the cell surface proteome when cultivated in different media. We elucidated and validated these differences. This study is the most thorough study on a cell surface proteome so far (two publications in preparation). A current project is extending this approaches to elucidate potential differences in the cell surface proteomes of leukemic cells of patients with chronic lymphocytic leukemia, before transplantation of hematopoietic cells and after relapse (in collaboration with Dr. Falk Heidenreich, group of PD Dr. Johannes Schetelig, Uniklinikum Carl Gustav Carus, TU Dresden).

3.4 Application II – Analysis of membrane-associated protein networks

As membrane proteins are challenging to analyze, transient protein networks that form on top of membranes in the early steps of trafficking are even more demanding. These proteins are masked by an overwhelming amount of proteins binding to the membrane irrespective of the presence or absence of a short, moderately conserved sorting sequence. We assume a high variability between replicates in both, proteomic background and specific transient binders. To enable us to use a variety of type-I and –II membrane proteins, we changed the coupling chemistry to a cysteine-targeted maleimido-functionalized lipid head group. The coupling was always successful. The introduction of label-free differential mass spectrometric approaches also succeeded and replaced the manual classification of identified proteins to specific binders or background proteome. It is this, what enabled unbiased findings while keeping false positives low.

However, we did not always observe a selective effect of the phosphoinositides used on the formation of recruitment complexes. In recruitment experiments with gE on PI4P containing liposomes, the AP-2 complex was recruited beside AP-1. One would expect to recruit AP-2 in presence of PI45P2 or PI345P3 rather than PI4P. This could be because of very high affinities of the sorting signals to components of various coats. It remains elusive thus, if the reason for the seeming unspecificity of AP-2 with respect to the phosphoinositide was caused by the design of the experiment with the sorting motifs analyzes, with a large amount of concentrated cargo and too less PI4P, or if a loss or conversion of PI4P occurred. It would be a further improvement to accompany lipidomic studies to recruitment experiments to explore this. The synthesis and use of phosphoinositide derivatives containing phosphothioesters instead of phosphoesters and thiols or hydrogen instead of alcoholic groups that might be still recognized by phosphoinositide-binding proteins, but kept unaltered, could be also a step forward to ensure phosphoinositide-specificity and to dissect steps of trafficking events. Also, we probe recruitment only with different phosphoinositides, but other lipids can also influence

trafficking. Even close lipid deviates can cause tremendous discrepancies in trafficking, for instance by altering membrane curvature or aggregating a subset of lipids into membrane microdomains. Potentially rapid conversion of phosphoinositides has to be considered. It might be circumvented by optimizing recruitment times as well as the concentration of GTP-γS for every type of recruitment route and mechanism separately. Another inexactness which could lead to improper conclusion about trafficking mechanisms is the use of heterologous expressed proteins. Inaccurate protein conformations or missing post translational modifications can occur and impact trafficking mechanisms. The absence of transmembrane domains, of non-cytosolic domains or co-trafficking membrane proteins, as are proposed for SorLa or sortilin and APP (Yang, 2013; Gustafsen, 2013), can also puzzle the identification of trafficking routes and mechansims.

Despite that, we were able to selectively recruit, detect and statistically validate AP-1, AP-2, AP-4, AP-5, retromer and ESCRT complexes in nearly stoichiometric amounts, thus mimicking respective trafficking ways. Moreover, we found interconnection to a diversity of signaling processes and links to structural determinants as the presynaptic complex. The concurrent discovery of AP-4 mediated trafficking of APP by Burgos et al. (Burgos, 2010) was based on a hypothesis driven study. The discovery of AP-4 and AP-5, demonstrated that unbiased, discovery driven recruitment experiments yield strikingly more information in shorter time. We are currently using recruitment experiments to mimic trafficking of other receptors, like integrin beta 1 (CD29), which was identified in exosomes, thus serving as a model for ESCRT-mediated, exosome-directed sorting (Hakulinen, 2008).

3.5 Epilogue

A systematic assessment of mass spectrometric parameters and quantification methods allowed us to establish workflows for a sound and serious quantitative compositional proteomic analysis, to substantiate enrichment success and to define proteomic differences between samples on an accurate statistical foundation. The in-depth analysis of cell surface proteome-enriched samples further validated the workflows and demonstrated their usefulness and general suitability. A notably demanding application is the analysis of recruitment experiments, where proteomic differences between samples are negligible with respect to the huge proteomic background and variabilities of the replicates. Nonetheless, the procedures succeeded to unravel interacting proteins and protein complexes even in near stoichiometric ratios. Such improved recruitment assays yield global and reliable information about trafficking routes, signaling pathway interrelations and more.

The achievements gained through the development of quantitative mass spectrometric parameters, of statistical methods and error models showed that a mass spectrometry lab has to be constant innovative. This is also true for the installed instrumentation. As the newest orbitrap system, Thermo Scientific's Orbitrap Fusion Tribrid, offers resolutions of 450,000 ($m/z=200$), new parameters and models will take advantage of this making accessible lower intense signals. This gain in resolution is accompanied by increase in accuracy, sensitivity and reliability and hence in the dynamic range and it widens the quantification window. Faster scanning is another attribute of new instruments, which will allow accessing more proteins for identification and quantification. The next milestone in proteomics will be the identification and quantitative description of a complete mammalian proteome in a single sample. Lately, the yeast proteome was analyzed that way (Nagaraj, 2012) with an estimated number of approximately 4000 proteins expressed. An important aspect that often repels the use of mass spectrometry in research projects is the costs of the approach. With the ability to analyze more complex samples fractionation steps can be reduced and so are the number of samples to be analyzed. The introduction of reliable label-free methods is yet another step that reduces costs twice. It is firstly the waiving of expensive labels and secondly the omitting of tedious adaptation processes. The work presented here, demonstrates that label-free approaches can be universally used.

3.6 Publications

Parts of this thesis have been used for following publications:

Niehage, C., Steenblock, C., Pursche, T., Bornhäuser, M., Corbeil, D., Hoflack, B. (2011) The cell surface proteome of human mesenchymal stromal cells. PLoS ONE 6, e20399.

Pocha, S.M., Wassmer, T., Niehage, C., Hoflack, B., Knust, E. (2011) Retromer controls epithelial cell polarity by trafficking the apical determinant Crumbs. Curr. Biol. 21, 1111–1117.

Prewitz, M.C., Seib, F.P., von Bonin, M., Friedrichs, J., Stißel, A., Niehage, C., Müller, K., Anastassiadis, K., Waskow, C., Hoflack, B., Bornhäuser, M., Werner, C. (2013) Tightly anchored tissue-mimetic matrices as instructive stem cell microenvironments. Nat. Methods 10, 788–794.

Niehage, C., Stange, C., Anitei, M., Hoflack, B. (2014) Liposome-based assays to study membrane-associated protein networks. Methods Enzymol. 534, 223-243.

Upcoming publications:

Balklava, Z., Guscott, B., Niehage, C., Currinn, H., Hoflack, B., Wassmer, T. (2013a) Identification of the intracellular interactome of the Amyloid Precursor Protein establishes its role as a regulator of mTOR. Submitted to PLOS Biology.

Balklava, Z., Niehage, C., Currinn, H., Guscott, B., Hoflack, B., Wassmer, T. (2013b) The amyloid precursor protein is a regulator of the cellular phosphoinositide metabolism via a direct interaction with the PIKfyve complex. Submitted to Curr. Biol.

Niehage, C., Karbanová, J., Corbeil, D., Hoflack, B. (2013) Cell surface marker selection by label-free quantitative mass spectrometry. In preparation.

Steenblock, C., Heckel, T., Espírito Santo, A. I., Czupalla, C., Niehage, C., Hoflack, B. (2013) The Cdc42 guanine nucleotide exchange factor FGD6 coordinates cell polarity and endosomal membrane recycling in osteoclasts. Submitted to J. Biol. Chem.

3.7 Talks

April 2013, Conference talk
'Label-free mass spectrometric analysis of in vitro constituted trafficking processes'
Annual meeting of the DFG transregio-project TRR83: 'Molecular Architecture and Cellular Functions of Lipid/Protein Assemblies', Dresden

Sept. 2013, Invited talk
'Label-free mass spectrometric approaches in proteomics'
Novo Nordisk Foundation Center for Biosustainability (Technical University of Denmark)

4. MATERIALS AND METHODS

4.1 General methods

4.1.1 SDS-PAGE

SDS-PAGEs were performed with a separation lane length of app. 4 cm using a Protean 2 system (BIO-RAD, CA USA) according to Sambrook et al. 2001. Replicates of different types of samples which were dedicated for quantitative comparison were allways separated on different gels and kept in different staining containers to avoid cross contamination by diffusion (Knaust, 2012).

4.1.2 Gel-staining/destaining

Gels were stained with 40% MeOH / 10% acetic acid / 0.25% (w/v) coomassie brilliant blue G-250 for 1h at room temperature and destained twice with 40% methanol / 10% acetic acid for 15 minutes and over night hereafter. The gels were transferred into 10% acetic acid (without methanol) and were allowed to destain further and swell for at least 4 hours.

4.1.3 Western Blotting

Unstained gels were electro-blotted in a Mini Trans-Blot Cell tank blot system (BIO-RAD, CA USA) assembled according to the manufacturer's protocol. The transfer buffer used was a 4:6 mixture of methanol and 25 mM Tris base / 190 mM glycine / 0.1%(w/v) SDS with an intrinsic pH of 8.3. Proteins were blotted onto a nitrocellulose membrane with a constant electric current of 250 mA at 4°C within 90 minutes. First antibodies used were anti clathrin heavy chain antibody (BD Biosciences, CA USA) and anti AP-1 gamma antibody (in-house preparation) in a dilution of 1:500 and 1:2000 for 1 hour at room temperature and detected with an 1:5000 dilution of an anti rabbit second antibody conjugated to a horse radish peroxidase. The light emitting conversion of a Luminate Forte Western HRP Substrate (Merck Millipore, Germany) was observed with a LAS-3000 (Raytest, Germany) imaging system. Tiff-format images were uploaded into the free GelAnalyzer-Software (www.GelAnalyzer.com) and gel band intensities were quantified by integration of illumination intensities over gel band areas.

4.1.4 Protein concentration assay

Protein concentrations were estimated using an RC/DC-Assay (BIO-RAD, CA USA), according to the manufacturer's protocol, skipping the pre-pecipitation (RC) step if applicable.

4.1.5 In-gel digestion after SDS-PAGE

Gel-separated protein samples were trypsin-digested in a modified procedure from Shevchenko et al., 2006. The coomassie-stained gel lanes were cut into 10 slices, each slice was further cut into cubes of app. 1 mm and all cubes of a slice transferred into an 0.5 ml Eppendorf tube. 400 µl H_2O was added, incubated for 10 minutes, vortexed, spun down and removed. This step was repeated first with 150 µl 25 mM ammonium bicarbonate in

acetonitrile and hereafter with 150 µL acetonitrile. The samples were reduced 150 µl 10 mM dithiothreitol / 100 mM ammonium bicarbonate was added, incubated for 45 minutes at 56°C, vortexed, spun down and removed. Cysteins were alkylaled by the addition of 150 µl 55 mM iodoacetamide / 100 mM ammonium bicarbonate and incubation for 20 minutes in the dark. The gel pieces were vortexed, spun down and solution was removed. 150 µl 25 mM ammonium bicarbonate in acetonitrile was added, incubated for 5 minutes, vortexed, spun down and removed. This step was repeated with 150 µl. Remaining acetonitrile was evaporated in a speed-vacuum system (Eppendorf, Germany) for 15 minutes. Trypsin-digestion was performed by the addition of 30 µl (0.15 µg) of trypsin (Trypsin Gold, Promega, WI USA) in 5 mM ammonium bicarbonate to the gel pieces, incubation for 5 minutes, addition of 30 µl 5 mM ammonium bicarbonate and storage of the closed tubes for 16h at 37°C. To stop digestion 60 µl of 0.5% trifluoroacetic acid in acetonitrile was added and tubes were sonicated in an ultrasonic bath for 10 minutes and spun down. The peptide eluates were transferred into empty LC-vials. 60 µl acetonitrile was added to the gel pieces and tubes are sonicated in an ultrasonic bath for 10 minutes and spun down. The eluates were added to the previous ones. Peptides were dried in a speed-vacuum system. 20 µl of LC-buffer A (0.1% formic acid) was added. The glass vials were closed and sonicated for 10 minutes in an ultrasonic bath and spun down.

4.1.6 In-solution digestion

Filter-aided sample preparation (FASP) was adapted from the Prince-Lab (http://openwetware.org/wiki/Prince:FASP). Dry proteins or protein pellets after recruitment assays were resuspended in 30 µl of 4%(w/v) SDS / 0.1 M DTT / 0.1 M TrisHCl, pH 7.6 and sonicated in an ultrasonic bath for 10 minutes at 65°C. The suspension was allowed to cool down to 20°C, mixed with 200 µl 8 M urea / 0.1 M TrisHCl, pH 8.5 and transferred to a 30K filter insert (Vivacon500 V01H22, Sartorius, Germany) placed in a 2 ml Eppendorf tube. The solution was concentrated for 15 minutes in a table-top centrifuge (Eppendorf, Germany) at 14000xg, 20°C. Twice, 100 µl of 8 M urea / 0.1 M TrisHCl, pH 8.5 was added and centrifugated for 15 minutes at 14000xg, 20°C. 100 µl 55 mM iodoacetamide in 8 M urea / 0.1 M TrisHCl, pH 8.5 was added and alkylation was performed for 20 minutes at room temperature in the dark. The reagent was eliminated by centrifugation for 15 minutes at 14000xg, two washes of 100 µL 8 M urea / 0.1 M TrisHCl, pH 8.5 followed by two washes of 100 µl 50 mM ammonium bicarbonate. Digestion was performed by addition of 50 µL trypsin in a ratio of 1:40(w/w) to the overall protein amount in 50 mM ammonium bicarbonate. The filters were vortexed and pulse-centrifugated to 2000xg, tubes closed and incubated for 4h to 18h at 37°C. The filter units were transferred into new collection tubes and centrifugated for 15 minutes at 14000xg. The filters were washed once with 40 µL 50 mM ammonium bicarbonate and the eluates acidified to 1% formic acid. Peptides were desalted using C18-SepPak cartridges (WAT023590, Waters, Great Britain). The cartridges were washed by one column volume (CV = 200 µl) of methanol, 1 CV of 80% acetonitrile / 0.1% trifluoro acetic acid and 3 CV of 0.1% trifluoro acetic acid. Peptides were loaded on top

of the column material and slowly passed three times through it. The column was washed with 3 CV of 0.1% trifluoro acetic acid and eluted into a LC-vial with 1 CV of 80% acetonitrile / 0.1% trifluoro acetic acid. Peptides were dried in a speed-vacuum system and resuspended in 0.1% formic acid for 10 minutes in an ultrasonic bath.

4.2 Assessment of mass spectrometric quantification methods

4.2.1 Preparation of a yeast background proteome
1 g of a backer's yeast cube was dissolved in 10 mL 4% SDS / 0.1 M DTT / 0.1 M TrisHCl, pH 7.6 . 250 U benzonase (Merck, Germany) was added. Yeast cells were disrupted by 3 rounds of homogenization in an EmulsiFlex-C5 Homogenizer (Avestin, Canada) and cell debris was pelleted by centrifugation for 30 minutes, 10000xg at 4°C. The supernatant was filtered (0.45 µm) and kept on ice while the concentration of proteins was estimated. Aliquots of 80 µg total protein were used for in-solution trypsin digestion.

4.2.2 Preparation of dynamic range standards
Tryptic digests of 500 pmol bovine serum albumin (#217498, Bruker, Germany) and 500 pmol ß-casein (#217507, Bruker, Germany) were each dissolved in 50 µl 0.1% formic acid by 15 minutes of sonication in an ultrasonic bath and combined. A dilution series of 500 amol and 5, 50 and 500 fmol and 5, 50 and 500 pmol digested proteins each in 90 µl 0.1% formic acid was prepared and replicates of 9 µl were measured in the LC-MS system. A similar dilution series was prepared with additional 600 ng of yeast peptides in each vial as a complex proteomic background and replicates of 9 µl were measured in the LC-MS system.

4.2.3 Preparation of accuracy standards
1 vial universal proteomics standard 2 (UPS2, Sigma-Aldrich, Germany) consisting of 48 proteins in different molar amounts as listed in Table 4.1 (10.6 µg total protein) was reduced, alkylated and digested in-solution with 0.2 µg of trypsin. One tenth of the total sample was used for each of four replicate LC-MS runs.

4.2.4 Preparation of differential proteome standards
1 vial of universal proteomics standard 1 (UPS1, Sigma-Aldrich, Germany) consisting of 48 proteins in an amount of 5 pmol (6 µg total protein) was reduced, alkylated and digested in-solution with 0.1 µg of trypsin. This stem solution and a preparation of a yeast background proteome was used to construct a series of samples as follows (actual protein amounts injected into the LC-MS system, each sample contained 60 ng yeast): 20 fmol, 6.7 fmol, 2.2 fmol, 0.74 fmol, 0.24 fmol UPS1. All samples were measured in quadruplets.

Table 4.1: Universal Proteomic Standards (Sigma-Aldrich)

UNIPROT AccNo	UPS1 [fmol]	UPS2 [fmol]	Name	MW [kDa]
P00915	5	50000	Carbonic anhydrase 1	28.738
P00918	5	50000	Carbonic anhydrase 2	29.095
P01031	5	50000	Complement C5	8.266
P69905	5	50000	Hemoglobin alpha chain	15.127
P68871	5	50000	Hemoglobin beta chain	15.867
P41159	5	50000	Leptin	16.024
P02768	5	50000	Serum Albumin	66.393
P62988	5	50000	Ubiquitin	9.387
P04040	5	5000	Catalase	59.583
P00167	5	5000	Cytochrome b5	16.021
P01133	5	5000	Epidermal Growth Factor	6.211
P02144	5	5000	Myoglobin C	17.051
P15559	5	5000	NAD(P)H dehydrogenase	30.984
P62937	5	5000	Peptidyl-prolyl cis-trans isomerase A	17.947
Q06830	5	5000	Peroxiredoxin 1	22.106
P63165	5	5000	Small ubiquitin-related modifier 1	37.42
P00709	5	500	Alpha-lactalbumin	14.07
P06732	5	500	Creatine kinase M-type	43.07
P12081	5	500	Histidyl-tRNA synthetase	58.223
P61626	5	500	Lysozyme C	14.692
Q15843	5	500	Neddylin [Nedd8]	9.071
P02753	5	500	Retinol-binding protein	21.065
P16083	5	500	Ribosyldihydronicotinamide dehydrogenase	25.817
P63279	5	500	Ubiquitin-conjugating enzyme E2 I	22.907
P01008	5	50	Antithrombin-III	49.033
P61769	5	50	Beta-2-microglobulin	11.729
P55957	5	50	BH3 Interacting domain death agonist	21.978
P07339	5	50	Cathepsin D	26.624
P08263	5	50	Glutathione S-transferase A1	25.482
P01344	5	50	Insulin-like growth factor II	7.464
P01127	5	50	Platelet-derived growth factor B chain	12.286
P10599	5	50	Thioredoxin	12.424
P08311	5	5	Cathepsin G	26.751
P99999	5	5	Cytochrome c	11.608
P06396	5	5	Gelsolin	82.954
P09211	5	5	Glutathione S-transferase P	23.22
P01112	5	5	GTPase HRas	21.292
P02787	5	5	Serotransferrin	75.143
O00762	5	5	Ubiquitin-conjugating enzyme E2 C	20.473
P51965	5	5	Ubiquitin-conjugating enzyme E2 E1	22.222
P08758	5	0.5	Annexin A 5	35.782
P02741	5	0.5	C-reactive protein	23.03
P05413	5	0.5	Fatty acid-binding protein	14.716
P10145	5	0.5	Interleukin-8	8.381
P02788	5	0.5	Lactotransferrin	78.289
P10636	5	0.5	Microtubule-associated protein tau	46.81
P00441	5	0.5	Superoxide dismutase	15.8
P01343	5	0.5	Insulin-like growth factor IA	7.643
P01375	5	0.5	Tumor necrosis factor	17.35

4.3 Analysis of plasma membrane proteins

4.3.1 Cell surface protein enrichment

All steps were carried out at 4°C. Dental pulp stem cells were isolated and prepared by a project partner (Jana Karbanova, Corbeil-Lab, TU Dresden). 6 petri dishes (diameter: 15 cm) corresponding to app. $2.5 \cdot 10^7$ cells were washed three times with Ca/Mg-PBS (phosphate-buffered saline containing 1 mM $CaCl_2$, 0.5 mM $MgCl_2$). Cells were incubated with 2 mL 0.2 mM Sulfo-NHS-SS-Biotin (Pierce, IL USA) in Ca/Mg-PBS for 1h on a rocking platform. After three rounds of washing with Ca/Mg-PBS residual N-hydroxysuccinimide is quenched in a 10 minutes incubation with 20 mM glycine followed by three more rounds of washing with Ca/Mg-PBS. Cells were scraped in 4 mL of Ca/Mg-PBS and transferred ito a 15 ml falcon tube. After centrifugation for 5 minutes at 500xg, the supernatant was removed and cells were lysed by addition of 250 µl ice-cold solubilization buffer (1% Triton-X-100, 0.1% SDS, Complete Protease Inhibor Cocktail (Sigma, Germany) in PBS). The extract was diluted fourfold with Ca/Mg-PBS. 20 µl of streptavidin beads (GE Healthcare....) were washed three times with 1 ml of the same buffer and added to the cell extract. Binding of the biotinylated cell-surface proteins to the streptavidin beads were allowed to occur during 2h at 4°C while rotating the tubes. The beads were washed ten times with 0.1% Triton-X-100 in Ca/Mg-PBS. Hereafter, proteins were released from the beads by reduction of the crosslinker with 100 µl of 50 mM 2-mercaptoethane sufonate in 0.1% Triton-X-100 / Ca/Mg-PBS. Solution and slurry were deposited into an Ultrafree-MC spin filter (0.45 µm, Milipore, Germany) and spun down. An acetone precipitation was performed on the flow through: the fourfold volume of ice-cold acetone was added and incubated for 15 minutes on ice. After centrifugation for 10 minutes at 14000xg, the supernatant was discarded and the precipitation repeated once. Residual acetone was removed in a speed-vacuum system. 20 µl of twofold Laemmli sample buffer was added to the pellet and dissolvation was aided by 10 minutes of sonication in an ultrasonic bath at 65°C. Proteins were seperated on a 10% SDS-PAGE before staining/destaining and in-gel digestion.

4.4 Analysis of membrane-associated protein networks

4.4.1 Preparation of mouse brain cytosol

For mass spectrometric approaches using a mouse proteome is effective as mouse is the best anotated organism in the non redundant UniprotKB beside human. Brain cytosol is easy to prepare and trafficking complex proteins in neurons are well expressed. However, for the study of more special trafficking ways like basolateral trafficking via AP-1 beta (AP1M2) (Gonzalez and Rodriguez-Boulan, 2009) other cell types may be prefered. 30 mice (CD1, female ,12-14 weeks old) were decapitated and the brains were dissected. The brains were homogenized with a 7 ml - Dounce homogenizer (Gerresheimer, Germany) in 4 ml of recruitment buffer (20 mM HepesKOH pH 7.2, 125 mM K-acetate, 2.5 mM Mg-acetate)

supplied with complete protease inhibitor cocktail (Roche, Germany). The homogenate was cleared from cell debris by centrifugation for 60 minutes, 370000xg at 4°C and another centrifugation of the supernatant for 60 minutes, 170000xg at 4°C. The supernatant was collected and immediately used for recruitment assays. An aliquot was removed for estimating the protein concentration.

4.4.2 Expression and purification of 6xHis-MBP-Tev-Cys-tagged proteins

The pET28-derived expression plasmid was transormed into an E. coli expression strain (BL21). The cytosolic domains of type I membrane receptors were amino-terminal flanked by a 6xHis-MBP dual tag and an interjacent TEV protease cleavage site followed by a cysteine. This allowed to remove the double tag with a high sequence-specific, high activity protease and to receive a cysteine for coupling to the maleimide lipid anchor. In case of type II membrane receptors this cysteine was placed at the carboxy-terminal site of the peptide. The 6xHis-tag allowed easy purification of the construct whereas a MBP-tag was introduced for enhancing solubility. A 5 ml over night culture was prepared from the respective expression strain in LB-medium with the requisite antibiotic 30 mg/l kanamycin at 37°C. It was transferred into a 2.5 l flask containing 500 ml LB medium plus kanamycin and cultivated at 37°C until the relative optical density to uninfected medium at 600 nm reached 0.6 (1 cm optical path). The culture is cooled to 15°C and the construct expression was induced with 0.1 mM IPTG. The expression was performed over night at 15°C. Bacteria were harvested by centrifugation at 2500xg for 20 minutes at 4°C. The pellets were resuspended in 15 ml cell disruption buffer (50 mM TrisHCl, 100 mM NaCl, protein inhibitor mix M (Serva, Germany) without EDTA, 20 mM imidazole, pH 7.4) on ice. 250 U benzonase (Merck, Germany) was added. The bacteria were disrupted by 3 rounds of homogenization in an EmulsiFlex-C5 Homogenizer (Avestin, Canada). Cell debris was pelleted by centrifugation for 30 minutes, 10000xg at 4°C. The supernatant was filtered (0.45 µm) and kept on ice. An empty 10 ml Poly-Prep column (BIO-RAD, CA USA) was filled with 1 ml IMAC sepharose (GE Healthcare, Sweden). The sepharose was washed 3 times with 10 ml H_2O. Metal was loaded with 10 ml NiCl and the sepharose was washed 3 times with 10 ml H_2O. The sepharose was washed once with cell disruption buffer. The supernatant was loaded onto the column, flow through was kept on ice. The column was washed 3 times with 10 ml cell disruption buffer. All flow throughs were kept separately on ice. The first elution step was performed by 3 washes with 1.8 ml cell disruption buffer containing 250 mM imidazole, all flow throughs were kept separately on ice. The second elution step was performed by 3 washes with 1.8 ml cell disruption buffer containing 500 mM imidazole, all flow throughs were kept separately on ice. An SDS-PAGE was performed from aliquots of all steps and checked for the elution peak of the construct (usually the first elution step with 250 mM imidazole). The respective sample was dialyzed over night vs. coupling buffer (20 mM HepesKOH pH 7.2, 125 mM K-acetate, 1 mM EDTA, no protease inhibitors). The sample was harvested, protein concentration estimated and adjusted to 10 mg/ml. Aliquots of 50 µl were stored at -80°C.

4.4.3 Preparation of liposomes

Liposomes were prepared in following molar ratio of lipids: 42% PC, 11% PS, 33% PE, 11% C, 1% PI, 2% anchor (PE) (see Table 4.2). The use of an extruder is optional. Extrusion causes monolaminar vesicles with a restricted size distribution. To keep size heterogeneity and to minimize hydrolysis of the maleimido anchor extrusion was omitted. For a full differential proteomic approach - 3 recruitments plus 3 controls - two of the following preparations were performed. For three different samples lipids were mixed as follows: 18.0 µl PC, 33.2 µl PE, 12.1 µl, 5.75 µl C, 32 µl of the respective phosphoinositide and 6.6 µl PE-MCC in a siliconized 1.5 ml Eppendorf tube. The mixture was thoroughly vortexed and dried down under a stream of nitrogen to yield a thin and homogeneous film of lipids. Lipids were resuspended in 330 µl coupling buffer (20 mM HepesKOH pH 7.2, 125 mM K-acetate, 1 mM EDTA). Six cycles of 4 minutes freezing in liquid nitrogen and 4 minutes of thawing in a thermoshaker at 25°C with thorough vortexing were applied. After 6 cycles (when the suspension appeared homogenous) 3 aliqouts of the liposome suspension were prepared in new siliconized Eppendorf tubes (100 µl/sample) and kept on room temperature to avoid phase transitions for immediate coupling of peptides. Optional extrusion step before aliquoting: A LipoFast handheld extruder (Avestin, Canada) was assembled according to the manufacturers manual with a polycarbonate membrane of a pore diameter of 400 nm. The liposome preparation was passed 21 times through the filter and harvested on the opposite syringe.

Table 4.2: Lipids used to prepare liposomes.

Lipid	Name	Vendor(#Order)	Stem solution
PI3P	1,2-dioleoyl-sn-glycero-3-phospho-(1'-myo-inositol-3'-phosphate)	Avanti (850150)	0.2 mg/ml in $CHCl_3/MEOH/H_2O$ (65:35:8)
PI4P	L-α-phosphatidylinositol-4-phosphate	Avanti (840045)	0.2 mg/ml in $CHCl_3/MEOH/H_2O$ (65:35:8)
PC	1,2-dioleoyl-sn-glycero-3-phosphocholine	Avanti (850375)	25 mg/ml in $CHCl_3$
PS	L-α-phosphatidylserine	Avanti (840032)	10 mg/ml in $CHCl_3$
PE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine	Avanti (850725)	10 mg/ml in $CHCl_3$
C	cholesterol	Avanti (700000)	25 mg/ml in $CHCl_3$
PE-MCC	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N -[4-(p-maleimidomethyl)cyclohexane-carboxamide]	Avanti (780201)	0.5 mg/ml in $CHCl_3$
PE-CHO	Di-O-hexadecyl-rac-glyceraldehyde was synthesized as described	(*)	10.0 mg/ml in $CHCl_3$

* (Bourel-Bonnet et al., 2005)

4.4.4 Coupling of peptides to liposomes via cysteine

Three aliquots of chosen His-MBP-TEV-receptor tail (50 µl in 20 mM HepesKOH pH 7.2, 125 mM K-Acetate, 2.5 mM Mg-Acetate) were defrosted. Details about the receptor domains used for recruitment assays are listed in Table 4.3. For a coupling control gel, an aliquot of 2.5 µl was removed and 2.5 µl 2x-Laemmli sample buffer was added. 2.5 µl dialyzed TEV-protease (1 mg/ml in 50 mM TrisCl pH 7.0, 150 mM NaCl, 20% glycerol, 1 mM TCEP) was added to the remaining 47.5 µl receptor tail. TCEP was adjusted to a final concentration of 1 mM. Tev-digestion was allowed to take place over night at 25°C. An aliquot of 2.5 µl was removed and 2.5 µl 2x-Laemmli sample buffer was added. Remaining 47.5 µl were kept on ice, until the liposome preparation was ready. Three Cys-controls were prepared as follows: 47.5 µl of cysteine (1 mg /ml) in coupling buffer and 1 mM TCEP was filled in each of three new 1.5 ml siliconized Eppendorf tubes. The 47.5 µl digested protein or 47.5 µl cysteine was added to each 100 µl-liposome suspension (6 samples : 3x Crb or other receptor and 3x Cys). Proteins were allowed to couple to liposomes for 60 min at room temperature. To saturate remaining maleimido groups, beta-mercaptoethanol was added to a final concentration of 1 mM to the liposome suspension, and incubated at RT for 15 min. Liposomes were pelleted in an Eppendorf table-top centrifuge (RT, 5 min, 20000xg). The supernatant (appr. 150 µl) was removed and an aliquot of 7.5 µl was kept and 7.5 µl 2x-Laemmli sample buffer was added. The pellet was washed once with 150 µl recruitment buffer (20 mM HepesKOH pH 7.2, 125 mM K-Acetate, 2.5 mM Mg-Acetate) and centrifugation (RT, 5 min, 20000xg). The liposome-pellet was resuspended in 200 µl recruitment buffer. An aliquot of 10 µl was removed and 10 µl 2x-Laemmli sample buffer was added. A SDS-PAGE was performed with collected samples (2.5 µl pre-TEV digest, 2.5 µl post-TEV digest, 7.5 µl post-coupling supernatant and 10 µl resuspended pellet) and stained with coomassie to assess the success of Tev-digestion and coupling reaction.

4.4.5 Recruitment of cytosolic proteins

The in vitro recruitment approach offers multiple parameters to assess like the concentration of cytosolic proteins, of anchor, phosphoinositide, recruitment time, addition of ATP, an ATP-regenerative system, of GTP, GTPγS (guanosine 5'-O-[gamma-thio]triphosphate) or the addition of drugs like latrunculin B. 500 µl of mouse brain cytosol (protein conc. 4.2 mg/ml) was supplemented with latrunculin B to yield a concentration of 28 µM. GTPγS was added to a concentration of 0.21 mM. The recruitment was started by addition of 200 µl liposome suspension (final protein concentration of 3 mg/ml) and incubation for 20 minutes at 37°C.

Table 4.3: Peptide constructs used for recruitment assays.

Name	UNIPROT	Sequence after cleavage by *Tev*
APP	A4_HUMAN	GSCGKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN
CRB	CRUM2_HUMAN	GSCGRKRRQSEGTYSPSQQEVAGARLEMDSVLKVPPEERLI
gE	GE_VZVO	GSIEGRCTKRMRVKAYRVDKSPYNQSMYYAGLPVDDFEDSESTDTEEEFGNAIGGSHGGSSYTVYIDKTR
NLG	NLGN1_MOUSE	GSCGYYKKDKRRHDVHRRCSPQRTTTNDLTHAPEEEIMSLQMKHTDLDHECESIHPHEVVLRTACPPDYTLAMRRSPDDIP LMTPNTITMIPNTIPGIQPLHTFNTFTGGQNNTLPHPHPHPHSHSTTRV
SORT	SORT_HUMAN	GSCGKKYVCGGRFLVHRYSVLQQHAEANGVDGVDALDTASHTNKSGYHDDSDEDLLE
TFR	TFR1_HUMAN	GSMMDQARSAFSNLFGGEPLSYTRFSLARQVDGDNSHVEMKLAVDEEENADNNTKANVTKPKRCSGSIC
VSVG	VGLG_VSIVG	GSCGYLYIKLKHTKKRQIYTDIEMNRLGR

His-MBP-Tag	MGSSHHHHHHSSGLVPRGSHMASGKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDG PDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELK AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKG ETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGA VALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRITKGENLYFQ

4.4.6 Postrecruitment purification of liposomes

The recruitment reaction was stopped by mixing with 1.4 ml of ice-cold 60% sucrose in recruitment buffer in a 13 ml ultracentrifugation tube. It was overlaid with 30% sucrose in recruitment buffer to 1 cm below tube's top. This was overlaid with recruitment buffer to 2 mm below tube's top. Opposite centrifuge tube were balanced with recruitment buffer if necessary. The sucrose step gradients were centrifugated for 14h, 200.000xg at 4°C. The liposome fraction on top of the gradient was transferred into a 4.5 ml ultracentrifugation tube, diluted with recruitment buffer and filled up to to 2 mm below tube's top. Opposite centrifuge tube were balanced with recruitment buffer if necessary. Liposomes were pelleted by centrifugation for 1h, 370.000xg at 4°C. The supernatant was completely removed and tube walls were dried with a dust-free wipe avoiding touching the pellet. 30 μl of 2x Laemmli sample buffer was added and the pellet was scratched with the pipette's tip. The tubes were placed into an ultrasonic bath and sonicated for 10 minutes at 65°C. The samples were transferred into 0.5 ml Eppendorf tubes and spun down. The sample volumes was adjusted to 30 μl with 2x Laemmli sample buffer and SDS-PAGEs followed by in-gel digestion was performed.

4.5 LC-MS and bioinformatic postacquisition workflows

4.5.1 LC-MS acquisition

Liquid chromatography of peptides before mass spectrometry was performed on an UltiMate 3000 HPLC-System (Thermo Fisher Scientific, MA USA) equipped with a 15 cm analytical reversed-phase C18-column, inner diameter 75 mm ; particle size 2 mm, 100 Å (Dionex, CA USA). The gradient was 5% B to 45% B linear in 80 minutes (A, 0.1% formic acid; B, acetonitrile, 0.1% formic acid). The detailed operation method is shown in listing I.

Mass spectrometry was performed on a LTQ-Orbitrap XL system (Thermo Fisher Scientific, MA USA) in a Top8-procedure. MS1 was hereby accomplished in the orbitrap with a resolution of 60000 (at 400 Th) triggering parallel LTQ acquisition (resolution of 30000) of MS2 of the 8 most intensive MS1 features. Dynamic exclusion was enabled with a repeat count of 2, a repeat duration of 5 seconds, an exclusion duration of 15 seconds with an exclusion list size of 500. Lock mass was optionally set to a polysiloxane adduct of 445.12 (Keller *et al.*, 2008). Collision activated dissociation was executed with an activation-Q of 0.25 and a normalized collision energy of 35 for 30 milliseconds. Details of the operation method are shown in listings II and III.

4.5.2 Processing of mass spectrometric spectra by MaxQuant

All .raw-files from XCalibur were loaded into MaxQuant (V1.2.2.5) (Cox and Mann, 2008) and an experimental table was created in a way that all slices of one lane were assigned to the same experiment (listing VIII). Carbamidomethylation of cysteine was considered as a fixed modification and oxidation of methionine, acetylation of the N-terminus, and deamidation of asparagine or glutamine as variable modifications. Other parameters are multiplicity of 1, trypsin as enzyme, maximum number of modifications per peptide of 3, maximal missed cleavages of 2, maximum charge of 4 with individual peptide mass tolerances allowed. A suitable FASTA data base was assigned (UNIPROT-TrEMBL 2012-01). Parameters in the protein identification tab were as follows: peptide and site FDR of 0.01, maximum peptide posterior error probability of 1, all minimum peptides of 1, filter labeled amino acids disabled, second peptides, re-quantify, label-free quantification and match between runs (2 minutes) enabled. Details of chosen MaxQuant parameters are shown in listing XI.

4.5.3 Parameters of label-free mass spectrometry

Quantification using peptide numbers, exponentialy modified protein abundance index, spectral counts, normalized spectral abundance factors MaxQuant-Intensities, Label-free quantification intensities, extraced ion intensity protein abundance index and intensity-based absolute quantification values are collected or calculated as follows and demonstrated on the prepared and measured accuracy standards (UPS2).

Peptide numbers (PN) were extracted from the respective columns of the MaxQuant proteinGroupfile output.

Exponentialy modified protein abundance indices (emPAI) were calculated for each protein (p) from the peptide numbers (PN) and the number of observable peptides (PO, see below: iBAQ) as follows:

$$emPAI_p = 10^{\left(\frac{PN_p}{PO_p}\right)}$$

Spectral counts (SC) were extracted from the MS/MS-count columns of the MaxQuant proteinGroupfile output.

Normalized spectral abundance factors (NSAF) of single proteins (p) were calculated from the spectral counts (SC), the length of the respective protein (L) and the number (n) of proteins in the data base as follows:

$$NSAF_p = \frac{\left(\frac{SC_p}{L_p}\right)}{\sum_{i=1}^{n}\left(\frac{SC_i}{L_i}\right)}$$

MaxQuant intensities (MQ) were extracted from the repective columns in the proteinGroups.txt file.

Label-free quantification intensities (LFQ) were extracted from the respective columns of the MaxQuant proteinGroupfile output.

Extracted peptide abundance intensities (xPAI) were calculated using the 'Precursor Ions Area Detector' module of the Proteome Discoverer 1.2 Software (Thermo Fisher Scientific, MA USA) with a mass range window of 2 ppm.

Intensity-based absolute quantification intensities (iBAQ) were calculated from the MQ-intensities in the proteinGroups.txt file by division through the number of observable peptides and logarithmizing. Numbers of theoretically observable peptides (PO) were calculated using a self-made python script (Listing 4.5) after retention time limits have been estimated by a yeast proteome analysis and the use of another python script (Listing 4.4 and 4.6). Parameters that restrict observability are extreme hydrophobicity or hydrophilicity, charge states and too low or high number of amino acids and can be passed to the python programs via a setting.ini – file (Listing 4.7). Since implementation into MaxQuant (Version 1.2.2.5) iBAQ values were extracted from the respective columns in the proteinGroups.txt file. This streamlined the workflow, but restricted the so far adjustable parameters to the preset values of 6 to 35 amino acids in peptide length.

All values were paired in a SQL-database and further analyzed in R.

4.5.4 Differential proteomics: G-tests on spectral counts

Spectral counts were used as a simple and easy accessible read-out value for the screening experiments. G-tests were used to judge for confidence in protein abundance changes as

follows. The sum of spectral counts of all proteins in a sample was calculated and likewise the sum of spectral counts of the respective control sample. Both were added to yield the grand total and the ratio of the sum of spectral counts of the sample and the grand total yielded the fraction of these on the grand total. Likewise, the fraction of the sum of spectral counts of the control was calculated. The sum of spectral counts of a specific protein in both the sample and control was calculated and multiplication with the sample's sum spectral count fraction yielded the expected number of spectral counts in the sample under the null hypothesis that there was no difference of abundance for that protein in the sample and control and that the difference in spectral counts reflected just the difference in the overall sum spectral counts of the sample and the control. Likewise, the expected spectral counts of that protein in the control was calculated. The G-value for that protein was calculated using the actual (a) and expected (e) spectral counts of sample (s) and control (c) as follows:

$$G = 2 \left(a_s \ln\frac{a_s}{e_s} + a_c \ln\frac{a_c}{e_c} \right)$$

G-value thresholds for p=0.01 (G=6.64) and p=0.05 (G=3.84) were estimated by approximation of this test statistics with a chi-square function (degree of freedom = 1). All calculations were done using R.

4.5.5 Differential proteomics: Permutation-based t-tests on LFQ intensities

The proteinGroups.txt-file was loaded into Perseus (Hubner *et al.*, 2010) with all LFQ-intensities as 'expression'. All proteins matching any of the three categorical annotations were deleted, LFQ-intensities were logarithmized, experiments were grouped to sample or control and proteins that were not detected in all three replicates of at least one group were deleted. Missing values were imputed by normal distribution with parameters that were adapted using histogram plots. A two sided t-test was performed with a false discovery rate in between 0.1 and 1%, and a slope value of 0.2 to 1, the control was hereby assigned to 'group 2'. The protein table and the threshold curve table were exported for further analysis in R (a brief example is given in listing 4.11).

4.5.6 Gene ontology assignment and enrichment analysis

Proteins in the proteinGroupfile.txt-output of MaxQuant were sorted for their sum intensities in a decreasing order and bins of 50 proteins were created. Uniprot identifiers were pasted into the gene products-field of the term enrichment tool of AmiGO (http://amigo.geneontology.org/cgi-bin/amigo/term_enrichment). Like for microarray analysis, no background set was selected. The database filter was set to UniProt KB and electronically inferred (IEA) annotations were allowed. The Bonferroni-corrected p-value threshold was set to 0.05 and the minimal number of gene products to 2. This procedure was applied for all intensity bins, the tables of GO terms and associated p-values were exported, unified in a SQL database and further analyzed in R (Listing 4.12).

Listing 4.1 : Operating method for the Dionex Ultimate 3000 system

```
20120529_BSAbCAS_500fmol_#3
Dionex Chromatography MS Link
Program for Dionex Chromatography MS Link
Sampler.TempCtrl = On
Sampler.Temperature.Nominal = 8.0 [°C]
Sampler.Temperature.LowerLimit =4.0 [°C]
Sampler.Temperature.UpperLimit =45.0 [°C]
Sampler.ReadyTempDelta = None
ColumnOven.TempCtrl = On
ColumnOven.Temperature.Nominal =40.0 [°C]
ColumnOven.Temperature.LowerLimit =20.0 [°C]
ColumnOven.Temperature.UpperLimit =50.0 [°C]
EquilibrationTime = 0.0 [min]
ColumnOven.ReadyTempDelta = 5.0 [°C]
LoadingPump.Pressure.LowerLimit =0 [bar]
LoadingPump.Pressure.UpperLimit =500 [bar]
LoadingPump.MaximumFlowRampDown =14 [µl/min²]
LoadingPump.MaximumFlowRampUp =14 [µl/min²]
LoadingPump.%A.Equate = "%A"
LoadingPump.%B.Equate = "%B"
%C.Equate = "%C"
NC_Pump.Pressure.LowerLimit = 0 [bar]
NC_Pump.Pressure.UpperLimit = 800 [bar]
NC_Pump.MaximumFlowRampDown = 0.199 [µl/min²]
NC_Pump.MaximumFlowRampUp = 0.199 [µl/min²]
NC_Pump.%A.Equate = "%A"
NC_Pump.%B.Equate = "%B"
DrawSpeed = 200 [nl/s]
DrawDelay = 5000 [ms]
DispSpeed = 2000 [nl/s]
DispenseDelay = 2000 [ms]
WasteSpeed = 4000 [nl/s]
WashSpeed = 4000 [nl/s]
LoopWashFactor = 2.000
SampleHeight = 1.500 [mm]
PunctureDepth = 7.000 [mm]
WashVolume = 500.000 [µl]
RinseBetweenReinjections = Yes
LowDispersionMode = Off
InjectMode = ulPickUp
FirstTransportVial = G1
LastTransportVial = G1
TransportVialCapacity = 99999
TransLiquidHeight = 5.000 [mm]
TransVialPunctureDepth = 8.000 [mm]
FlushVolume = 7.000 [µl]
LoadingPump_Pressure.Step = 0.01 [s]
LoadingPump_Pressure.Average =Off
NC_Pump_Pressure.Step = 0.01 [s]
NC_Pump_Pressure.Average = Off
ValveRight = 1_2
0.000 Wait UV.Ready and LoadingPump.Ready and
NC_Pump.Ready and ColumnOven.Ready and Sampler.Ready and PumpModule.Ready
;Chromeleon sets this property to signal to Xcalibur that it is ready to
start a run.
ReadyToRun = 1
;Xcalibur sets this property to start the run or injection.
Wait StartRun
LoadingPump.Flow = 6.000 [µl/min]
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 5.0 [%]
Wait UV.Ready and LoadingPump.Ready and
NC_Pump.Ready and ColumnOven.Ready and Sampler.Ready and PumpModule.Ready
Inject
LoadingPump_Pressure.AcqOn
NC_Pump_Pressure.AcqOn
;Chromeleon sets this property to signal the injection to Xcalibur.
InjectResponse = 1
```

Listing 4.1 - continued: Operating method for the Dionex Ultimate 3000 system

```
;Depending on your system configuration it might be necessary to manually
insert
;a "Relay" command below in order to send the start signal to the MS.
;Typical syntaxes:
;Pump_Relay_1.Closed Duration =2.00
;UM3PUMP_Relay1.On Duration = 2.00
;Pump_Relay_1.Closed Duration =2.00
;UM3PUMP_Relay1.On Duration = 2.00
LoadingPump.Flow = 6.000 [µl/min]
ValveRight = 1_2
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 5.0 [%]
0.200 Relay_1_MS_Start.Duration = 5.00
10.000 NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 5.0 [%]
ValveRight = 10_1
12.000 LoadingPump.Flow = 6.000 [µl/min]
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
15.000 LoadingPump.Flow = 30.000 [µl/min]
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
90.000 NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 45.0 [%]
95.000 NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 45.0 [%]
96.000 NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 95.0 [%]
100.000 LoadingPump.Flow = 30.000 [µl/min]
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
103.000 LoadingPump.Flow = 6.000 [µl/min]
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
106.000 NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 95.0 [%]
107.000 NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 5.0 [%]
110.000 ValveRight = 1_2
130.000 LoadingPump_Pressure.AcqOff
NC_Pump_Pressure.AcqOff
LoadingPump.Flow = 6.000 [µl/min]
LoadingPump.%B = 0.0 [%]
%C = 0.0 [%]
NC_Pump.Flow = 0.200 [µl/min]
NC_Pump.%B = 5.0 [%]
InjectResponse = 0
End
```

Listing 4.2 : Abstract of the LTQ Orbitrap XL tuning method

```
20120529_BSAbCAS_500fmol_#3
Segment: 1
Tune File Values
Source Type: NSI
Capillary Temp (C): 200.00
APCI Vaporizer Temp (C): 0.00
Sheath Gas Flow (): 0.00
Aux Gas Flow (): 0.00
Sweep Gas Flow (): 0.00
Injection Waveforms: Off
Ion Trap Zoom AGC Target: 3000.00
Ion Trap Full AGC Target: 30000.00
Ion Trap SIM AGC Target: 10000.00
Ion Trap MSn AGC Target: 10000.00
FTMS Injection Waveforms: On
FTMS Full AGC Target: 500000.00
FTMS SIM AGC Target: 50000.00
FTMS MSn AGC Target: 200000.00
Reagent Ion Source Polarity: Negative
Reagent Ion Source Temp (C): 160.00
Reagent Ion Source Emission
Current (uA):
50.00
Reagent Ion Source Electron
Energy (V):
-70.00
Reagent Ion Source CI Pressure
(psi):
11.00
Reagent Vial 1 Ion Time: 100.00
Reagent Vial 1 AGC Target: 100000.00
Reagent Vial 2 Ion Time: 50.00
Reagent Vial 2 AGC Target: 100000.00
Supplemental Activation Energy: 15.00
POSITIVE POLARITY
Source Voltage (kV): 2.10
Source Current (uA): 100.00
Capillary Voltage (V): 6.00
Tube Lens (V): 80.00
Skimmer Offset (V): 0.00
Multipole RF Amplifier (Vp-p): 400.00
Multipole 00 Offset (V): -4.00
Lens 0 Voltage (V): -4.50
Multipole 0 Offset (V): -5.75
Lens 1 Voltage (V): -10.00
Gate Lens Offset (V): -54.00
Multipole 1 Offset (V): -6.50
Front Lens (V): -6.75
Ion Trap Zoom Micro Scans: 1
Ion Trap Zoom Max Ion Time (ms): 50.00
Ion Trap Full Micro Scans: 1
Ion Trap Full Max Ion Time (ms): 25.00
Ion Trap SIM Micro Scans: 1
Ion Trap SIM Max Ion Time (ms): 50.00
Ion Trap MSn Micro Scans: 1
Ion Trap MSn Max Ion Time (ms): 100.00
FTMS Full Micro Scans: 1
FTMS Full Max Ion Time (ms): 500.00
FTMS SIM Micro Scans: 1
FTMS SIM Max Ion Time (ms): 50.00
FTMS MSn Micro Scans: 1
FTMS MSn Max Ion Time (ms): 100.00
Reagent Ion Lens 1 (V): -20.00
Reagent Ion Gate Lens (V): -120.00
Reagent Ion Lens 2 (V): -15.00
Segment: 1
Reagent Ion Lens 3 (V): -15.00
Reagent Ion Back Lens Offset (V): -6.50
Reagent Ion Back Multipole Offset
(V):
-7.00
```

Listing 4.2 - continued: Abstract of the LTQ Orbitrap XL tuning method

```
NEGATIVE POLARITY
Source Voltage (kV): 1.50
Source Current (uA): 100.00
Capillary Voltage (V): -35.00
Tube Lens (V): -100.00
Skimmer Offset (V): 0.00
Multipole RF Amplifier (Vp-p): 400.00
Multipole 00 Offset (V): 4.00
Lens 0 Voltage (V): 4.20
Multipole 0 Offset (V): 4.50
Lens 1 Voltage (V): 15.00
Gate Lens Offset (V): 35.00
Multipole 1 Offset (V): 8.00
Front Lens (V): 5.25
Ion Trap Zoom Micro Scans: 1
Ion Trap Zoom Max Ion Time (ms): 50.00
Ion Trap Full Micro Scans: 1
Ion Trap Full Max Ion Time (ms): 25.00
Ion Trap SIM Micro Scans: 1
Ion Trap SIM Max Ion Time (ms): 50.00
Ion Trap MSn Micro Scans: 1
Ion Trap MSn Max Ion Time (ms): 50.00
FTMS Full Micro Scans: 1
FTMS Full Max Ion Time (ms): 500.00
FTMS SIM Micro Scans: 1
FTMS SIM Max Ion Time (ms): 50.00
FTMS MSn Micro Scans: 1
FTMS MSn Max Ion Time (ms): 100.00
Reagent Ion Lens 1 (V): 140.00
Reagent Ion Gate Lens (V): 21.75
Reagent Ion Lens 2 (V): 88.25
Reagent Ion Lens 3 (V): 31.75
Reagent Ion Back Lens Offset (V): 5.40
Reagent Ion Back Multipole Offset
(V):
19.00
Additional FT Tune File Values
FT Tune Item 1: 0.000000
FT Tune Item 2: 0.000000
FT Tune Item 3: 0.000000
FT Tune Item 4: 0.000000
FT Tune Item 5: 0.000000
FT Tune Item 6: 0.000000
FT Tune Item 7: 0.000000
FT Tune Item 8: 0.000000
FT Tune Item 9: 0.000000
FT Tune Item 10: 0.000000
Calibration File Values
Multiple RF Frequency: 2545.400000
Main RF Frequency: 1190.000000
QMSlope0: 32.633474
QMSlope1: 32.653024
QMSlope2: 32.375685
QMSlope3: 0.000000
QMSlope4: 0.000000
QMInt0: -28.495468
QMInt1: 0.000000
QMInt2: -25.488200
QMInt3: 0.000000
QMInt4: 0.000000
End Section Slope: 0.000000
End Section Int: 12.000000
PQD CE Factor: 11.033762
IsoW Slope: 0.000010
IsoW Int: 0.310542
Reagent MP Slope: 6.014851
Reagent MP Int: -4.431932
Tickle Amp. Slope0: 0.000058
Tickle Amp. Int0: 0.003728
```

Listing 4.2 - continued: Abstract of the LTQ Orbitrap XL tuning method

```
Tickle Amp. Slope1: 0.002000
Tickle Amp. Int1: 0.400000
Tickle Amp. Slope2: 0.002000
Tickle Amp. Int2: 0.400000
Tickle Amp. Slope3: 0.002000
Tickle Amp. Int3: 0.400000
Multiplier 1 Normal Gain (pos): -1160.000000
Multiplier 1 High Gain (pos): -1315.000000
Multiplier 2 Normal Gain (pos): -1175.000000
Multiplier 2 High Gain (pos): -1335.000000
Multiplier 1 Normal Gain (neg): -1310.000000
Multiplier 1 High Gain (neg): -1485.000000
Multiplier 2 Normal Gain (neg): -1295.000000
Multiplier 2 High Gain (neg): -1470.000000
Normal Res. Eject Slope: 0.013551
Normal Res. Eject Intercept: 4.064961
Zoom Res. Eject Slope: 0.004181
Zoom Res. Eject Intercept: 2.089748
Turbo Res. Eject Slope: 0.069200
Turbo Res. Eject Intercept: 35.000000
AGC Res. Eject Slope: 0.069200
AGC Res. Eject Intercept: 17.300000
UltraZoom Res. Eject Slope: 0.001276
UltraZoom Res. Eject Intercept: 0.323719
Normal Mass Slope: 28.609101
Normal Mass Intercept: -6.297141
Zoom Mass Slope: 27.043745
Zoom Mass Intercept: -38.640341
Turbo Mass Slope: 28.282197
Turbo Mass Intercept: 128.677267
AGC Mass Slope: 28.282197
AGC Mass Intercept: 128.677267
UltraZoom Mass Slope: 27.080305
UltraZoom Mass Intercept: -29.121168
Vernier Fine Mass Slope: 428.431667
Vernier Fine Mass Intercept: 0.000000
Vernier Coarse Mass Slope: 0.000000
Vernier Coarse Mass Intercept: 0.000000
Cap. Device Min (V): -139.476020
Cap. Device Max (V): 139.129266
Tube Lens Device Min (V): 260.533478
Tube Lens Device Max (V): -259.376421
Skimmer Device Min (V): -139.832412
Skimmer Device Max (V): 139.770320
Multipole 00 Device Min (V): -139.168632
Multipole 00 Device Max (V): 139.095879
Lens 0 Device Min (V): -140.040678
Lens 0 Device Max (V): 139.381980
Gate Lens Device Min (V): -136.299647
Gate Lens Device Max (V): 135.915741
Split Gate Device Min (V): 0.197153
Split Gate Device Max (V): 0.147953
Multipole 0 Device Min (V): -139.391930
Multipole 0 Device Max (V): 139.222408
Lens 1 Device Min (V): -139.472669
Lens 1 Device Max (V): 139.354340
Multipole 1 Device Min (V): -138.695366
Multipole 1 Device Max (V): 138.496867
Front Lens Device Min (V): -139.762120
Front Lens Device Max (V): 139.666672
Front Section Device Min (V): -142.297965
Front Section Device Max (V): 141.819813
Center Section Device Min (V): -141.865740
Center Section Device Max (V): 141.573812
Back Section Device Min (V): -142.760199
Back Section Device Max (V): 142.557238
Back Lens Device Min (V): -142.688682
Back Lens Device Max (V): 142.365528
Reagent Lens 1 Device Min (V): -143.565954
```

Listing 4.2 - continued: Abstract of the LTQ Orbitrap XL tuning method

```
Reagent Lens 1 Device Max (V): 143.427374
Reagent Gate Lens Min (V): -132.970381
Reagent Gate Lens Max (V): 132.933349
Reagent Lens 2 Device Min (V): -142.554596
Reagent Lens 2 Device Max (V): 142.067153
Reagent Lens 3 Device Min (V): -143.410874
Reagent Lens 3 Device Max (V): 142.707489
Reagent Electron Lens Device Min
(V):
-0.190768
Reagent Electron Lens Device Max
(V):
149.758668
```

Listing 4.3 : Operating method for the LTQ Orbitrap XL

```
Creator: LTQ Orbitrap XL
Last modified: 5/31/2011 by LTQ Orbitrap XL
MS Run Time (min): 130.00
Sequence override of method parameters not enabled.
Divert Valve: not used during run
Contact Closure: not used during run
Syringe Pump: not used during run
MS Detector Settings:
Real-time modifications to method enabled
Stepped collision energy not enabled
Additional Microscans:
MS2 0 0
MS3 0 0
MS4 0 0
MS5 0 0
MS6 0 0
MS7 0 0
MS8 0 0
MS9 0 0
MS10 0 0
Segment 1 Information
Duration (min): 130.00
Number of Scan Events: 9
Tune Method: 110120_nanoESI
Scan Event Details:
1: FTMS + p norm !corona !pi res=60000 o(200.0-2000.0)
CV = 0.0V
2: ITMS + c norm !corona !pi Dep MS/MS Most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
3: ITMS + c norm !corona !pi Dep MS/MS 2nd most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
```

Listing 4.3 - continued: Operating method for the LTQ Orbitrap XL

```
4: ITMS + c norm !corona !pi Dep MS/MS 3rd most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
5: ITMS + c norm !corona !pi Dep MS/MS 4th most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
6: ITMS + c norm !corona !pi Dep MS/MS 5th most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
7: ITMS + c norm !corona !pi Dep MS/MS 6th most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
8: ITMS + c norm !corona !pi Dep MS/MS 7th most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
9: ITMS + c norm !corona !pi Dep MS/MS 8th most intense ion from (1)
Activation Type: CID
Min. Signal Required: 500.0
Isolation Width: 3.00
Normalized Coll. Energy: 35.0
Default Charge State: 2
Activation Q: 0.250
Activation Time: 30.000
CV = 0.0V
Lock Masses:
Pos List Name: 445
Source: API Source
Mass List: 445.120030
Neg List Name: N/A
Source: API Source
Mass List: (none)
Data Dependent Settings:
Use separate polarity settings disabled
Parent Mass List: (none)
Reject Mass List: (none)
Neutral Loss Mass List: (none)
Product Mass List: (none)
Neutral loss in top: 3
Product in top: 3
Most intense if no parent masses found not enabled
```

Listing 4.3 - continued: Operating method for the LTQ Orbitrap XL

```
Add/subtract mass not enabled
FT master scan preview mode enabled
Charge state screening enabled
Charge state dependent ETD time enabled
Monoisotopic precursor selection enabled
Non-peptide monoisotopic recognition not enabled
Charge state rejection enabled
Unassigned charge states : rejected
Charge state 1 : rejected
Charge state 2 : not rejected
Charge state 3 : not rejected
Charge states 4+ : not rejected
Chromatography mode is disabled
Global Data Dependent Settings:
Use global parent and reject mass lists not enabled
Exclude parent mass from data dependent selection not enabled
Exclusion mass width relative to mass
Exclusion mass width relative to low (ppm): 10.000
Exclusion mass width relative to high (ppm): 10.000
Parent mass width by mass
Parent mass width low: 0.50000
Parent mass width high: 0.50000
Reject mass width relative to mass
Reject mass width relative to low (ppm): 10.000
Reject mass width relative to high (ppm): 10.000
Zoom/UltraZoom scan mass width by mass
Zoom/UltraZoom scan mass width low: 5.00
Zoom/UltraZoom scan mass width high: 5.00
FT SIM scan mass width low: 5.00
FT SIM scan mass width high: 5.00
Neutral Loss candidates processed by decreasing intensity
Neutral Loss mass width by mass
Neutral Loss mass width low: 0.50000
Neutral Loss mass width high: 0.50000
Product candidates processed by decreasing intensity
Product mass width by mass
Product mass width low: 0.50000
Product mass width high: 0.50000
MS mass range: 0.00-1000000.00
MSn mass range by mass
MSn mass range: 0.00-1000000.00
Use m/z values as masses not enabled
Analog UV data dep. not enabled
Dynamic exclusion enabled
Repeat Count: 2
Repeat Duration: 5.00
Exclusion List Size: 500
Exclusion Duration: 15.00
Exclusion mass width relative to mass
Exclusion mass width relative to low (ppm): 10.000
Exclusion mass width relative to high (ppm): 10.000
Expiration: disabled
Isotopic data dependence not enabled
Mass Tags data dependence not enabled
Custom Data Dependent Settings:
Not enabled
```

Listing 4.4 : Python-script to facilitate ascertaining of retention time limits.

```python
import csv
from pyteomics import fasta, parser, mass, achrom, electrochem, auxiliary

def run():
    ini = iniFile("RCLimits.ini")
    writer=csv.writer(open(ini.getSetting('OutPut','RCLimits.csv'), 'wb'), delimiter='\t',
    quotechar='|', quoting=csv.QUOTE_MINIMAL)
    reader=csv.reader(open(ini.getSetting('FileName','peptides.csv'), 'rb'), delimiter='\t',
    quotechar='|')
    firstRow = True
    seqColumn = 0
    for row in reader:
        if firstRow:
            firstRow = False
            header = []
            for entryNum in range(len(row)):
                header.append(row[entryNum])
                if row[entryNum]=='Sequence':
                    seqColumn = entryNum
                    header.append('RT_RP')
                    header.append('RT_normal')
            writer.writerow(header)
        else:
            try:
                rowToWrite = []
                for j in range(seqColumn+1):
                    rowToWrite.append(row[j])
                parsed = parser.parse( row[seqColumn], show_unmodified_termini=True)
                RT_RP = achrom.calculate_RT(parsed, achrom.RCs_zubarev)
                RT_normal = achrom.calculate_RT(parsed, achrom.RCs_yoshida_lc)
                rowToWrite.append(str(RT_RP))
                rowToWrite.append(str(RT_normal))
                for i in range(len(row)-1-seqColumn):
                    rowToWrite.append(row[seqColumn+1+i])
                writer.writerow(rowToWrite)
            except Exception:
                print "Errored Sequence"
class iniFile(object):
    def __init__(self, fname):
        self.fname = fname
        self.settings = {}
        self.read()
    def printOut(self):
        for setting in self.settings:
            print setting + " : " + self.settings[setting]
    def read(self):
        try:
            reader = csv.reader(open(self.fname, 'rb'), delimiter='=', quotechar='|')
            for row in reader:
                self.settings[row[0]]=row[1]
        except Exception:
            print "An error occured while reading "+self.fname
    def getSetting(self, setting, default=""):
        try:
            return self.settings[setting]
        except Exception:
            return default
    def getSettingInt(self, setting, default=0):
        try:
            return int(self.settings[setting])
        except Exception:
            return default
    def getSettingFloat(self, setting, default=0.0):
        try:
            return float(self.settings[setting])
        except Exception:
            return default
    def __del__(self):
        del self.fname
        del self.settings
if (__name__=="__main__"):
    run()
```

Listing 4.5: Python-script to estimate numbers of observable tryptic peptides.

```python
import csv

from pyteomics import fasta, parser, mass, achrom, electrochem, auxiliary

import re
from collections import deque
from itertools import chain, product

def run():
    ini = iniFile('settings.ini')

    debug = True

    writeName = ini.getSetting('OutputPeptidesFileName', 'peptides.csv')
    writeNameProt = ini.getSetting('OutputProteinsFileName', 'proteins.csv')

    writingPeptides = True
    if writeName == "false":
        writingPeptides = False

    writingProteins = True
    if writeNameProt == "false":
        writingProteins = False

    errorWriter = csv.writer(open(ini.getSetting('ErrorLog','error.csv'), 'wb'),
    delimiter='\t', quotechar='|', quoting=csv.QUOTE_MINIMAL)

    csv.field_size_limit(1000000000)

    if writingPeptides:
        writer = csv.writer(open(writeName, 'wb'), delimiter='\t', quotechar='|',
        quoting=csv.QUOTE_MINIMAL)
    if writingProteins:
        protWriter = csv.writer(open(writeNameProt, 'wb'), delimiter='\t', quotechar='|',
        quoting=csv.QUOTE_MINIMAL)

    if writingPeptides:
        writer.writerow(['AccNo', 'PeptideSeq', 'RCrp', 'RCnorm', 'm', 'z', 'mz'])
    if writingProteins:
        protWriter.writerow(['AccNo', 'AccStr', 'GN', 'AA', 'm', 'NoP'])

    try:
        AAFilter = False
        if(ini.getSetting('AA', 'false')=="true"):
            print "AA-Filter enabled"
            AAFilter = True
        else:
            print "AA-Filter disabled"

        RTFilter = False
        if(ini.getSetting('RT_RP', 'false')=="true"):
            print "RT_RP-Filter enabled"
            RTFilter = True
        else:
            print "RT_RP-Filter disabled"

        RTNormFilter = False
        if(ini.getSetting('RT_Norm', 'false')=="true"):
            print "RT_normal-Filter enabled"
            RTNormFilter = True
        else:
            print "RT_normal-Filter disabled"

        ZFilter = False
        if(ini.getSetting('Z', 'false')=="true"):
            print "Z-Filter enabled"
            ZFilter = True
        else:
            print "Z-Filter disabled"
```

Listing 4.5 – continued : Python-script to estimate numbers of observable tryptic peptides.

```python
pHValue = ini.getSettingFloat('pH')

proteinCount = 0
erroredPeptideCount = 0

for description, sequence in fasta.read(ini.getSetting('FileName', 'example.fasta')):
    proteinCount += 1
    new_peptides = cleave(sequence, parser.expasy_rules['trypsin'])
    valid_peptides = []

    AccNo = re.split('\|', description)[1]
    AccStr = re.split('_', re.split('\|', description)[2])[0]

    GN=""
    if 'GN=' in re.split('\|', description)[2]:
        for arg in re.split(' ', re.split('\|', description)[2]):
            if arg.startswith('GN='):
                GN=arg[3:]

    ProtMass = ""
    NoP = 0
    erroredPeptideCountProt = 0

    for p in new_peptides:
        validPeptide = True
        if AAFilter:
            if len(str(p)) < ini.getSettingInt('AA-min'):
                validPeptide = False
            elif len(str(p)) > ini.getSettingInt('AA-max'):
                validPeptide = False

        if validPeptide:
            valid_peptides.append(p)

    new_peptides_length = len(new_peptides)
    del new_peptides

    peptides = [{'sequence':i} for i in valid_peptides]

    del valid_peptides

    for peptide in peptides:
        try:
            peptide['parsed_sequence'] = parser.parse( peptide['sequence'],
            show_unmodified_termini=True)
            peptide['valid'] = True
        except Exception:
            errorWriter.writerow([AccNo, peptide['sequence']])
            peptide['valid'] = False

        if peptide['valid']:
            peptide['length'] = parser.length(peptide['parsed_sequence'])

    for peptide in peptides:
        if peptide['valid']:
            peptide['charge'] = int(round(electrochem.charge(
            peptide['parsed_sequence'], pH=pHValue)))
            peptide['mass'] = mass.calculate_mass(peptide['parsed_sequence'])
            peptide['m/z'] = mass.calculate_mass(
            peptide['parsed_sequence'], charge=peptide['charge'])

    for peptide in peptides:
        if peptide['valid']:
            peptide['RT_RP'] = achrom.calculate_RT(peptide['parsed_sequence'],
            achrom.RCs_zubarev)
            peptide['RT_normal'] = achrom.calculate_RT(peptide['parsed_sequence'],
            achrom.RCs_yoshida_lc)

    counter = 0
```

Listing 4.5 – continued : Python-script to estimate numbers of observable tryptic peptides.

```python
        while counter < len(peptides):
            peptide = peptides[counter]
            removed = False

            if peptide['valid']:

                if RTFilter:
                    if peptide['RT_RP'] < ini.getSettingInt('RT_RP-min'):
                        peptides.remove(peptide)
                        removed = True
                    elif peptide['RT_RP'] > ini.getSettingInt('RT_RP-max'):
                        peptides.remove(peptide)
                        removed = True

                if RTNormFilter and not removed:
                    if peptide['RT_normal'] < ini.getSettingInt('RT_Norm-min'):
                        peptides.remove(peptide)
                        removed = True
                    elif peptide['RT_normal'] > ini.getSettingInt('RT_Norm-max'):
                        peptides.remove(peptide)
                        removed = True

                if ZFilter and not removed:
                    if peptide['charge'] < ini.getSettingInt('Z-min'):
                        peptides.remove(peptide)
                        removed = True
                    elif peptide['charge'] > ini.getSettingInt('Z-max'):
                        peptides.remove(peptide)
                        removed = True

            if not removed:
                counter += 1

        for peptide in peptides:
            try:
                rp = peptide['RT_RP']
                if writingPeptides:
                    writer.writerow([AccNo, peptide['sequence'], peptide['RT_RP'],
                    peptide['RT_normal'], peptide['mass'], peptide['charge'],
                    peptide['m/z']])
            except Exception:
                erroredPeptideCount += 1
                erroredPeptideCountProt += 1

        NoP = len(peptides)-erroredPeptideCountProt

        try:
            parsedProtein = parser.parse( sequence, show_unmodified_termini=True)
            ProtMass = str(mass.calculate_mass(parsedProtein))
        except Exception:
            ProtMass = ""

        if writingProteins:
            protWriter.writerow([AccNo, AccStr, GN, str(len(sequence)), ProtMass,
            str(NoP)])

        print str(proteinCount) + " proteins parsed "
        del peptides
        del AccNo
except Exception:
    print "An error occured while reading "+ini.getSetting('FileName', 'example.fasta')

del errorWriter
if writingPeptides:
    del writer
if writingProteins:
    del protWriter

print 'Done! '+str(erroredPeptideCount)+' peptides have failed to parse'
```

Listing 4.5 – continued : Python-script to estimate numbers of observable tryptic peptides.

```python
raw_input('')

    del ini

def cleave(sequence, rule, missed_cleavages=0):
    peptides = []
    cleavage_sites = deque([0], maxlen=missed_cleavages+2)
    for i in chain(map(lambda x: x.end(), re.finditer(rule, sequence)),
                   [None]):
        cleavage_sites.append(i)
        for j in range(0, len(cleavage_sites)-1):
            peptides.append(sequence[cleavage_sites[j]:cleavage_sites[-1]])
    if '' in peptides:
        peptides.remove('')
    return peptides

class iniFile(object):
    def __init__(self, fname):
        self.fname = fname
        self.settings = {}

        self.read()

    def printOut(self):
        for setting in self.settings:
            print setting + " : " + self.settings[setting]

    def read(self):

        try:
            reader = csv.reader(open(self.fname, 'rb'), delimiter='=', quotechar='|')

            for row in reader:
                self.settings[row[0]]=row[1]

        except Exception:
            print "An error occured while reading "+self.fname

    def getSetting(self, setting, default=""):
        try:
            return self.settings[setting]
        except Exception:
            return default

    def getSettingInt(self, setting, default=0):
        try:
            return int(self.settings[setting])
        except Exception:
            return default

    def getSettingFloat(self, setting, default=0.0):
        try:
            return float(self.settings[setting])
        except Exception:
            return default

    def __del__(self):
        del self.fname
        del self.settings

if(__name__=="__main__"):
        run()
```

Listing 4.6 : Parameter-file for listing 4.4 (RCLimits.ini).

```
FileName=evidence.txt
OutPut=evidence_out.txt
```

Listing 4.7 : Parameter-file for listing 4.5 (Settings.ini).

```
AA=true
AA-min=6
AA-max=60
RT_RP=false
RT_RP-min=5
RT_RP-max=35
Z=true
Z-min=2
Z-max=3
RT_Norm=false
RT_Norm-min=0
RT_Norm-max=0
Gradient_Length=40
pH=3.0
FileName=human.fasta
OutputPeptidesFileName=peptides.csv
OutputProteinsFileName=proteins.csv
Path=C:\Users\Admin\Python
ErrorLog=error.csv
```

Listing 4.8 : Abstract of an experimental design-file for MaxQuant (experimentalDesign.txt).

```
Name       Fraction    Experiment
3APP1-01               3APP1
3APP1-02               3APP1
...                    ...
3APP1-18               3APP1
3APP2-01               3APP2
3APP2-02               3APP2
...                    ...
3APP2-18               3APP2
3APP3-01               3APP3
3APP3-02               3APP3
...                    ...
3APP3-18               3APP3
3Cys1-01               3Cys1
3Cys1-02               3Cys1
...                    ...
```

Listing 4.9 : MaxQuant (Version 1.2.2.5) parameter settings

```
NumberOfThreads = 11           # For an Intel 6-core processor and Hyperthreading activated
VariableModifications = Oxidation (M) ; Deamidation (NQ)
Multiplicity = 1               # No variable modifications through introduced labels
Enzym = Trypsin                # No Lys-C pre-digestion
FirstSearch = 20 ppm
MainSearch = 6 ppm
MaxNumberOfModificationsPerPeptide = 3
MaxMissedCleavages = 2
MaxCharge = 4
IndividualPeptideMassTolerances = TRUE
Type = Standard                # Normal precursor selection
FTMS MS/MS-tolerance = 20 ppm
FTMS TopPeaksPer100Da = 10
FTMS De-isotoping = TRUE
FTMS HigherChargeStates = TRUE
FTMS WaterLoss = TRUE
FTMS AmmoniaLoss = TRUE
FTMS DependendIons = TRUE
ITMS MS/MS-tolerance = 0.5 Da
ITMS TopPeaksPer100Da = 10
ITMS De-isotoping = FALSE
ITMS HigherChargeStates = TRUE
ITMS WaterLoss = TRUE
ITMS AmmoniaLoss = TRUE
ITMS DependendIons = TRUE
FASTA-files = uniprot_sprot_2011_02_HUMAN.fasta # or 2012_01_MOUSE.fasta
IncludeContaminants = TRUE
I=L = TRUE
FixedModifications = Carbamidomethyl (C)
PeptideFDR = 0.01
SiteFDR = 0.01
MaxPeptidePEP = 1
MinPeptides = 1
MinRazorPlusUniquePeptides = 0
MinUniquePeptides = 0
ProteinFDR = 0.01
ApplySiteFDRSeparately = TRUE
MinPeptideLength = 6
MinScore = 0
FilterLabeledAminoAcids = FALSE
SecondPeptides = TRUE
ExperimentalDesignFile = …\experimentalDesign.txt
UseOnlyUnmodifiedPeptides = FALSE
UseRazorAndUniquePeptides = TRUE
DiscardUnmodifiedCounterpartPeptides = FALSE
MinRatioCount = 2
SiteQuantificationMode = UseLeastModifiedPeptides
SiteQuantificationUseForOccupancies = NormalizedRatios
Re-quantify = TRUE
MatchBetweenRuns = TRUE
TimeWindow = 2 min
Label-freeQuantification = TRUE
LFQMinRatioCount = 2
iBAQ = TRUE
LogFit = TRUE
```

Listing 4.10 : Perseus workflow

```
Generic upload{}() => {matrix1}

Filter category{matrix1}(Column=Only identified by site, Find=+, Mode=Remove matching rows) =>
{matrix2}

Filter category{matrix2}(Column=Reverse, Find=+, Mode=Remove matching rows) => {matrix3}

Filter category{matrix3}(Column=Contaminant, Find=+, Mode=Remove matching rows) => {matrix4}

Log{matrix4}(Base=2, Columns=LFQ Intensity 345APP1; LFQ Intensity 345APP2; LFQ Intensity
345APP3; LFQ Intensity 345Cys1; LFQ Intensity 345Cys2; LFQ Intensity 345Cys3; LFQ Intensity
35APP1; LFQ Intensity 35APP2; LFQ Intensity 35APP3; LFQ Intensity 35Cys1; LFQ Intensity
35Cys2; LFQ Intensity 35Cys3; LFQ Intensity 3APP1; LFQ Intensity 3APP2; LFQ Intensity 3APP3;
LFQ Intensity 3Cys1; LFQ Intensity 3Cys2; LFQ Intensity 3Cys3; LFQ Intensity 45APP1; LFQ
Intensity 45APP2; LFQ Intensity 45APP3; LFQ Intensity 45Cys1; LFQ Intensity 45Cys2; LFQ
Intensity 45Cys3; LFQ Intensity 4APP1; LFQ Intensity 4APP2; LFQ Intensity 4APP3; LFQ Intensity
4Cys1; LFQ Intensity 4Cys2; LFQ Intensity 4Cys3) => {matrix5}

Create groups{matrix5}(LFQ Intensity 345APP1=345APP, LFQ Intensity 345APP2=345APP, LFQ
Intensity 345APP3=345APP, LFQ Intensity 345Cys1=345Cys, LFQ Intensity 345Cys2=345Cys, LFQ
Intensity 345Cys3=345Cys, LFQ Intensity 35APP1=35APP, LFQ Intensity 35APP2=35APP, LFQ
Intensity 35APP3=35APP, LFQ Intensity 35Cys1=35Cys, LFQ Intensity 35Cys2=35Cys, LFQ Intensity
35Cys3=35Cys, LFQ Intensity 3APP1=3APP, LFQ Intensity 3APP2=3APP, LFQ Intensity 3APP3=3APP,
LFQ Intensity 3Cys1=3Cys, LFQ Intensity 3Cys2=3Cys, LFQ Intensity 3Cys3=3Cys, LFQ Intensity
45APP1=45APP, LFQ Intensity 45APP2=45APP, LFQ Intensity 45APP3=45APP, LFQ Intensity
45Cys1=45Cys, LFQ Intensity 45Cys2=45Cys, LFQ Intensity 45Cys3=45Cys, LFQ Intensity
4APP1=4APP, LFQ Intensity 4APP2=4APP, LFQ Intensity 4APP3=4APP, LFQ Intensity 4Cys1=4Cys, LFQ
Intensity 4Cys2=4Cys, LFQ Intensity 4Cys3=4Cys) => {matrix6}

Filter valid values{matrix6}(Matrix access=Rows, Min. number of valid values=3, Mode=In at
least one group) => {matrix7}

Replace missing values by normal distribution{matrix7}(Width=0.3, Down shift=1.8) => {matrix8}

Two samples{matrix8}(Group 1=4APP, Group 2=3APP, Test=T-test, Side=both, Use for
truncation=Permutation-based FDR, Threshold value=0.05, S0=1, -Log10=True, Number of
randomizations=250) => {matrix9}

Scatterplot{matrix9}()
```

Listing 4.11 : R-script to generate a volcano-plot.

```
setwd("K:\\LABOR\\20120509 CrbsvsCys PI3P")

ExperimentTable <- 'PI3P_CrbsvsCys-postE.txt'
ThresholdTable <- 'PI3P_CrbsvsCys-thresholds.txt'
TitleText <- 'Crb2 vs. Cys (PI3P)'
FDR <- 0.01
So <- 0.2

Experiment <- read.table(ExperimentTable, quote = "\"", header = TRUE,
           sep = "\t", stringsAsFactors = FALSE, comment.char = "")


Threshold <- read.table(ThresholdTable, quote = "\"", header = TRUE,
           sep = "\t", stringsAsFactors = FALSE, comment.char = "")

xx <- (Experiment$tdiff)                                 # t-test difference

yy <- (Experiment$pval)                                  # t-test -log10(p-value)


par(oma=c(0,0,0,0), mgp = c(3.5,1,0), mar = c(5, 4, 4, 2), mai = c(1.5,1.75,1,1))

plot(  xx,yy,
       main= TitleText,
       xlab = expression(log[2](Crb2/Cys)),
       ylab = expression(-log[10](p-value)),
       xlim = c(-max(abs(xx))*1.2,max(abs(xx)*1.2)),
       ylim = c(0,max(abs(yy)*1.2)),
       pch = 20,col='lightgrey',cex.axis = 1.5, cex.lab=2)


abline(v=0, col='black', lty=1, lwd=2)

text(  x=par()[['usr']][1],
       y=par()[['usr']][3] + (par()[['usr']][4] - par()[['usr']][3])/20,
       labels=c(paste("FDR = ", FDR, "\nSO = ", So)),
       pos=4)

signSignificant <- Experiment$tsign == "+"              # t-test significant

points(xx[signSignificant], yy[signSignificant],col="lightgrey",bg="lightgrey",pch=16)

lines(x=Threshold$diff, y=Threshold$pval, col='black', lwd=2, lty=2)

signAP <- Experiment$AdapterComplex != ""

points(xx[signAP], yy[signAP],col="black",bg="black",pch=16)

text(  xx[signAP],
       yy[signAP],
       labels=Experiment$AccStr[signAP],
       pos=2,cex=0.8)
```

Listing 4.12 : R-script for GO-enrichment analysis by hypergeometric distribution.

```
# for improved regular expression operations, here: strapply()
library(gsubfn)

# Set Working Directory
setwd("K:\\My Paper\\2013 - MCP")

# Parse MaxQuant-Table
Experiment<- read.table( "proteinGroups - CORRECTED.txt", quote = "\"", header = TRUE,
                         sep ="\t", stringsAsFactors = FALSE, comment.char = "")

# Filter for 'only identified by site','reverse','contaminant'
Experiment<-Experiment[Experiment$Reverse != "+",]
Experiment<-Experiment[Experiment$Only.identified.by.site != "+",]
Experiment<-Experiment[Experiment$Contaminant != "+",]
exclude.columns<-c("Only.identified.by.site", "Reverse","Contaminant")
'%!in%' <- function(x,y)!('%in%'(x,y)) # Create a not-in Operator
Experiment<-Experiment[,colnames(Experiment)%!in%exclude.columns]

# Filter for columns needed
include.columns<-c("Only.identified.by.site",
                   "Reverse",
                   "Contaminant",
                   "Gene.Names",
                   "Protein.Names",
                   "Protein.Descriptions",
                   "LFQ.Intensity.B10D",
                   "LFQ.Intensity.B10H",
                   "LFQ.Intensity.B10L",
                   "LFQ.Intensity.B2C",
                   "LFQ.Intensity.B2G",
                   "LFQ.Intensity.B2K",
                   "LFQ.Intensity.C10B",
                   "LFQ.Intensity.C10F",
                   "LFQ.Intensity.C10J",
                   "LFQ.Intensity.C2A",
                   "LFQ.Intensity.C2E",
                   "LFQ.Intensity.C2I",
                   "LFQ.Intensity.T10F",
                   "LFQ.Intensity.T10G",
                   "LFQ.Intensity.T10H",
                   "LFQ.Intensity.T10I",
                   "LFQ.Intensity.T10J",
                   "LFQ.Intensity.T2A",
                   "LFQ.Intensity.T2B",
                   "LFQ.Intensity.T2C",
                   "LFQ.Intensity.T2D",
                   "LFQ.Intensity.T2E")
Experiment<-Experiment[,colnames(Experiment)%in%include.columns]

# log2-transformation of LFQ-values
expression.values<-c(
  "LFQ.Intensity.B10D",
  "LFQ.Intensity.B10H",
  "LFQ.Intensity.B10L",
  "LFQ.Intensity.B2C",
  "LFQ.Intensity.B2G",
  "LFQ.Intensity.B2K",
  "LFQ.Intensity.C10B",
  "LFQ.Intensity.C10F",
  "LFQ.Intensity.C10J",
  "LFQ.Intensity.C2A",
  "LFQ.Intensity.C2E",
  "LFQ.Intensity.C2I",
  "LFQ.Intensity.T10F",
  "LFQ.Intensity.T10G",
  "LFQ.Intensity.T10H",
  "LFQ.Intensity.T10I",
  "LFQ.Intensity.T10J",
  "LFQ.Intensity.T2A",
```

Listing 4.12 - continued : R-script for GO-enrichment analysis.

```
  "LFQ.Intensity.T2B",
  "LFQ.Intensity.T2C",
  "LFQ.Intensity.T2D",
  "LFQ.Intensity.T2E")
expression.groups<-c("B10",
                     "B2",
                     "C10",
                     "C2",
                     "T10",
                     "T2")
expression.columns<-paste("log2.",expression.values,sep = "")
Exp.log2<-sapply(Experiment[,colnames(Experiment)%in%expression.values],log2)
colnames(Exp.log2)=expression.columns<-paste("log2.",expression.values,sep = "")

# Convert -Inv into NA
index<-Exp.log2[,]<1
Exp.log2[index] <- NA

# Create Mean-LFQ columns
Exp.log2<-data.frame(Exp.log2,
                     B10=rowMeans(Exp.log2[,1:3],na.rm = TRUE),
                     B2=rowMeans(Exp.log2[,4:6],na.rm = TRUE),
                     C10=rowMeans(Exp.log2[,7:9],na.rm = TRUE),
                     C2=rowMeans(Exp.log2[,10:12],na.rm = TRUE),
                     T10=rowMeans(Exp.log2[,13:17],na.rm = TRUE),
                     T2=rowMeans(Exp.log2[,18:22],na.rm = TRUE))
Experiment<-cbind(Experiment,Exp.log2)

# Create AccStr column
fun.AccStr<-function(x)
  {
    SProt<-unlist(strapply(x,"sp\\|.{6,9}\\|(.{2,6})_"))
    Trembl<-unlist(strapply(x,"\\A>(tr\\|.{6,9}\\|)"))
    if(is.null(SProt)){return(Trembl)}else{return(SProt[1])}
  }
Experiment$AccStr<-(lapply(Experiment$Protein.Descriptions,FUN=fun.AccStr))

# Create AccNo column
fun.AccNo<-function(x)
{
  SProt<-unlist(strapply(x,"sp\\|(.{6,9})\\|"))
  Trembl<-unlist(strapply(x,"\\A>tr\\|(.{6,9})\\|"))
  if(is.null(SProt)){return(Trembl)}else{return(SProt[1])}
}
Experiment$AccNo<-(lapply(Experiment$Protein.Descriptions,FUN=fun.AccNo))

# Map specific gene ontology entries from Uniprot by AccNo
Cytosol<- read.delim("GO0005829.tab",stringsAsFactors = FALSE) # cytosol (CC)
PM<- read.delim("GO0005886.tab",stringsAsFactors = FALSE)      # plasma membrane (CC)
OM<- read.delim("GO0031090.tab",stringsAsFactors = FALSE)      # organelle membrane (CC)
Experiment$Cytosol<-Experiment$AccNo%in%Cytosol$Entry
Experiment$PM<-Experiment$AccNo%in%PM$Entry
Experiment$OM<-Experiment$AccNo%in%OM$Entry

# Create & Sort single Mean-LFQ-AccStr-AccNo-GO subtables
B10<-Experiment[!is.na(Experiment$B10),colnames(Experiment)%in%c("B10","Cytosol","PM","OM")]
B10<-B10[order(B10$B10,decreasing=TRUE),]
B2<-Experiment[!is.na(Experiment$B2),colnames(Experiment)%in%c("B2","Cytosol","PM","OM")]
B2<-B2[order(B2$B2,decreasing=TRUE),]
C10<-Experiment[!is.na(Experiment$C10),colnames(Experiment)%in%c("C10","Cytosol","PM","OM")]
C10<-C10[order(C10$C10,decreasing=TRUE),]
C2<-Experiment[!is.na(Experiment$C2),colnames(Experiment)%in%c("C2","Cytosol","PM","OM")]
C2<-C2[order(C2$C2,decreasing=TRUE),]
T10<-Experiment[!is.na(Experiment$T10),colnames(Experiment)%in%c("T10","Cytosol","PM","OM")]
T10<-T10[order(T10$T10,decreasing=TRUE),]
T2<-Experiment[!is.na(Experiment$T2),colnames(Experiment)%in%c("T2","Cytosol","PM","OM")]
T2<-T2[order(T2$T2,decreasing=TRUE),]
```

Listing 4.12 - continued : R-script for GO-enrichment analysis.

```
# Create bin tables with numbers of GO-assignments
binsize<-100
B10.bins<-data.frame()
DF<-B10
L<-length(DF[,1])%/%binsize

for(i in c(1:L)){
  B10.bins[i,1]<-sum(DF[((i*binsize)-binsize):(i*binsize),2])
  B10.bins[i,2]<-sum(DF[((i*binsize)-binsize):(i*binsize),3])
  B10.bins[i,3]<-sum(DF[((i*binsize)-binsize):(i*binsize),4])
}
B2.bins<-data.frame()
DF<-B2
L<-length(DF[,1])%/%binsize
for(i in c(1:L)){
  B2.bins[i,1]<-sum(DF[((i*binsize)-binsize):(i*binsize),2])
  B2.bins[i,2]<-sum(DF[((i*binsize)-binsize):(i*binsize),3])
  B2.bins[i,3]<-sum(DF[((i*binsize)-binsize):(i*binsize),4])
}
C10.bins<-data.frame()
DF<-C10
L<-length(DF[,1])%/%binsize
for(i in c(1:L)){
  C10.bins[i,1]<-sum(DF[((i*binsize)-binsize):(i*binsize),2])
  C10.bins[i,2]<-sum(DF[((i*binsize)-binsize):(i*binsize),3])
  C10.bins[i,3]<-sum(DF[((i*binsize)-binsize):(i*binsize),4])
}
C2.bins<-data.frame()
DF<-C2
L<-length(DF[,1])%/%binsize
for(i in c(1:L)){
  C2.bins[i,1]<-sum(DF[((i*binsize)-binsize):(i*binsize),2])
  C2.bins[i,2]<-sum(DF[((i*binsize)-binsize):(i*binsize),3])
  C2.bins[i,3]<-sum(DF[((i*binsize)-binsize):(i*binsize),4])
}
T10.bins<-data.frame()
DF<-T10
L<-length(DF[,1])%/%binsize
for(i in c(1:L)){
  T10.bins[i,1]<-sum(DF[((i*binsize)-binsize):(i*binsize),2])
  T10.bins[i,2]<-sum(DF[((i*binsize)-binsize):(i*binsize),3])
  T10.bins[i,3]<-sum(DF[((i*binsize)-binsize):(i*binsize),4])
}
T2.bins<-data.frame()
DF<-T2
L<-length(DF[,1])%/%binsize
for(i in c(1:L)){
  T2.bins[i,1]<-sum(DF[((i*binsize)-binsize):(i*binsize),2])
  T2.bins[i,2]<-sum(DF[((i*binsize)-binsize):(i*binsize),3])
  T2.bins[i,3]<-sum(DF[((i*binsize)-binsize):(i*binsize),4])
}

# calculate p-values from Hpergeometric Distribution
total.HUMAN <- 20253 # Uniprot> organism:homo sapiens AND reviewed:yes
total.Cytosol <- length(Cytosol$Entry)
total.PM <- length(PM$Entry)
total.OM <- length(OM$Entry)
fun.hyper<-function(x,binsize,GO,human){
  (dhyper(x,GO,human-GO,binsize))}
fun.invers<-function(x){(-1)*x}
# B10
B10.bins$V4<-(lapply(B10.bins[,1],FUN=fun.hyper,binsize,total.Cytosol,total.HUMAN))
B10.bins$V5<-(lapply(B10.bins[,2],FUN=fun.hyper,binsize,total.PM,total.HUMAN))
B10.bins$V6<-(lapply(B10.bins[,3],FUN=fun.hyper,binsize,total.OM,total.HUMAN))
# convert p-values to: -1 * log(p-values)
B10.bins$V4<-(lapply(B10.bins[,4],FUN=log10))
B10.bins$V5<-(lapply(B10.bins[,5],FUN=log10))
B10.bins$V6<-(lapply(B10.bins[,6],FUN=log10))
B10.bins[,4]<-as.matrix(sapply(B10.bins[,4],fun.invers))
B10.bins[,5]<-as.matrix(sapply(B10.bins[,5],fun.invers))
```

Listing 4.12 - continued : R-script for GO-enrichment analysis.

```
B10.bins[,6]<-as.matrix(sapply(B10.bins[,6],fun.invers))
# provide meaningful column names
colnames(B10.bins)<-c("NoCytosol","NoPM","NoOM","Cytosol","PM","OM")
# B2
B2.bins$V4<-(lapply(B2.bins[,1],FUN=fun.hyper,binsize,total.Cytosol,total.HUMAN))
B2.bins$V5<-(lapply(B2.bins[,2],FUN=fun.hyper,binsize,total.PM,total.HUMAN))
B2.bins$V6<-(lapply(B2.bins[,3],FUN=fun.hyper,binsize,total.OM,total.HUMAN))

# convert p-values to: -1 * log(p-values)
B2.bins$V4<-(lapply(B2.bins[,4],FUN=log10))
B2.bins$V5<-(lapply(B2.bins[,5],FUN=log10))
B2.bins$V6<-(lapply(B2.bins[,6],FUN=log10))
B2.bins[,4]<-as.matrix(sapply(B2.bins[,4],fun.invers))
B2.bins[,5]<-as.matrix(sapply(B2.bins[,5],fun.invers))
B2.bins[,6]<-as.matrix(sapply(B2.bins[,6],fun.invers))
# provide meaningful column names
colnames(B2.bins)<-c("NoCytosol","NoPM","NoOM","Cytosol","PM","OM")
# C10
C10.bins$V4<-(lapply(C10.bins[,1],FUN=fun.hyper,binsize,total.Cytosol,total.HUMAN))
C10.bins$V5<-(lapply(C10.bins[,2],FUN=fun.hyper,binsize,total.PM,total.HUMAN))
C10.bins$V6<-(lapply(C10.bins[,3],FUN=fun.hyper,binsize,total.OM,total.HUMAN))
# convert p-values to: -1 * log(p-values)
C10.bins$V4<-(lapply(C10.bins[,4],FUN=log10))
C10.bins$V5<-(lapply(C10.bins[,5],FUN=log10))
C10.bins$V6<-(lapply(C10.bins[,6],FUN=log10))
C10.bins[,4]<-as.matrix(sapply(C10.bins[,4],fun.invers))
C10.bins[,5]<-as.matrix(sapply(C10.bins[,5],fun.invers))
C10.bins[,6]<-as.matrix(sapply(C10.bins[,6],fun.invers))
# provide meaningful column names
colnames(C10.bins)<-c("NoCytosol","NoPM","NoOM","Cytosol","PM","OM")
# C2
C2.bins$V4<-(lapply(C2.bins[,1],FUN=fun.hyper,binsize,total.Cytosol,total.HUMAN))
C2.bins$V5<-(lapply(C2.bins[,2],FUN=fun.hyper,binsize,total.PM,total.HUMAN))
C2.bins$V6<-(lapply(C2.bins[,3],FUN=fun.hyper,binsize,total.OM,total.HUMAN))
# convert p-values to: -1 * log(p-values)
C2.bins$V4<-(lapply(C2.bins[,4],FUN=log10))
C2.bins$V5<-(lapply(C2.bins[,5],FUN=log10))
C2.bins$V6<-(lapply(C2.bins[,6],FUN=log10))
C2.bins[,4]<-as.matrix(sapply(C2.bins[,4],fun.invers))
C2.bins[,5]<-as.matrix(sapply(C2.bins[,5],fun.invers))
C2.bins[,6]<-as.matrix(sapply(C2.bins[,6],fun.invers))
# provide meaningful column names
colnames(C2.bins)<-c("NoCytosol","NoPM","NoOM","Cytosol","PM","OM")
# T10
T10.bins$V4<-(lapply(T10.bins[,1],FUN=fun.hyper,binsize,total.Cytosol,total.HUMAN))
T10.bins$V5<-(lapply(T10.bins[,2],FUN=fun.hyper,binsize,total.PM,total.HUMAN))
T10.bins$V6<-(lapply(T10.bins[,3],FUN=fun.hyper,binsize,total.OM,total.HUMAN))
# convert p-values to: -1 * log(p-values)
T10.bins$V4<-(lapply(T10.bins[,4],FUN=log10))
T10.bins$V5<-(lapply(T10.bins[,5],FUN=log10))
T10.bins$V6<-(lapply(T10.bins[,6],FUN=log10))
T10.bins[,4]<-as.matrix(sapply(T10.bins[,4],fun.invers))
T10.bins[,5]<-as.matrix(sapply(T10.bins[,5],fun.invers))
T10.bins[,6]<-as.matrix(sapply(T10.bins[,6],fun.invers))
# provide meaningful column names
colnames(T10.bins)<-c("NoCytosol","NoPM","NoOM","Cytosol","PM","OM")
# T2
T2.bins$V4<-(lapply(T2.bins[,1],FUN=fun.hyper,binsize,total.Cytosol,total.HUMAN))
T2.bins$V5<-(lapply(T2.bins[,2],FUN=fun.hyper,binsize,total.PM,total.HUMAN))
T2.bins$V6<-(lapply(T2.bins[,3],FUN=fun.hyper,binsize,total.OM,total.HUMAN))
# convert p-values to: -1 * log(p-values)
T2.bins$V4<-(lapply(T2.bins[,4],FUN=log10))
T2.bins$V5<-(lapply(T2.bins[,5],FUN=log10))
T2.bins$V6<-(lapply(T2.bins[,6],FUN=log10))
T2.bins[,4]<-as.matrix(sapply(T2.bins[,4],fun.invers))
T2.bins[,5]<-as.matrix(sapply(T2.bins[,5],fun.invers))
T2.bins[,6]<-as.matrix(sapply(T2.bins[,6],fun.invers))
# provide meaningful column names
colnames(T2.bins)<-c("NoCytosol","NoPM","NoOM","Cytosol","PM","OM")
```

Listing 4.12 - continued : R-script for GO-enrichment analysis.

```
# Define Plotframe
nf <- layout(matrix(c(1),1,1, byrow=TRUE), respect=TRUE)

# Define Histogramplot-Function
fun.plot<-function(DF,SAMPLE){
  par(oma=c(0,0,0,0), mgp = c(3.5,1,0), mar = c(5, 4, 4, 2), mai = c(1,1.5,0.5,0.25))
  mp <- barplot(DF,
        beside = TRUE,
        axisnames = FALSE,
        ylim=c(0,25),
        axes=FALSE,
        legend=c("Plasma membrane","Organelle membrane","Cytosol"),
        args.legend=list(x=1.4*length(DF),y=23,box.lwd = 0,box.col = "white",bg = "white"),
        col=c(grey(0.25),grey(0.60),grey(1))
  )
  # create a baseline for column groups of three (in case a bar is zero)
  minmp <- mp[2,]-1.5
  maxmp <- mp[2,]+1.5
  segments(minmp,0,maxmp,0)
  # Add axes texts and y-axis
  mtext(paste("Bins of intensity-ranked proteins (",SAMPLE,")",sep=""),side=1,cex=1.5,padj=2)
  mtext("p-value",side=2,cex=1.5,padj=-4)
  axis(
    2,
    at=c(0,5,10,15,20,25),
    labels=c(expression(-10^0),
             expression(-10^5),
             expression(-10^10),
             expression(-10^15),
             expression(-10^20),
             expression(-10^25)),
    tick=TRUE,
    lty="solid",
    lwd=2,
    lwd.ticks=2,
    col=NULL,
    col.ticks=NULL,
    cex.axis=1.0,
    las=1)
}

# arrange data for barplots
B10.plot<-as.vector(t(as.matrix(B10.bins[,c(5,6,4)])))
B2.plot<-as.vector(t(as.matrix(B2.bins[,c(5,6,4)])))
C10.plot<-as.vector(t(as.matrix(C10.bins[,c(5,6,4)])))
C2.plot<-as.vector(t(as.matrix(C2.bins[,c(5,6,4)])))
T10.plot<-as.vector(t(as.matrix(T10.bins[,c(5,6,4)])))
T2.plot<-as.vector(t(as.matrix(T2.bins[,c(5,6,4)])))

# Do the plots
fun.plot(B10.plot,"B10")
fun.plot(B2.plot,"B2")
fun.plot(T10.plot,"T10")
fun.plot(T2.plot,"T2")
fun.plot(C10.plot,"C10")
fun.plot(C2.plot,"C2")

# Two dimensional plot (pp-plot)

# B10 vs T10
min.length<-min(length(T10.plot),length(B10.plot))
length(T10.plot)<-min.length
length(B10.plot)<-min.length
B10T10.plot<-cbind(T10.plot,B10.plot)
# B2 vs T2
min.length<-min(length(T2.plot),length(B2.plot))
length(T2.plot)<-min.length
length(B2.plot)<-min.length
B2T2.plot<-cbind(T2.plot,B2.plot)
```

Listing 4.12 - continued : R-script for GO-enrichment analysis.

```
# Define pp-plot-function
fun.pp<-function(DF,x,y){
  par(oma=c(0,0,0,0), mgp = c(3.5,1,0), mar = c(5, 4, 4, 2), mai = c(1.5,1.5,0.5,0.25))
  plot(DF,
       xlab="",
       ylab="",
       xlim=c(0,25),
       ylim=c(0,25),
       type="p",
       pch=22,
       bg=c(grey(0.25),grey(0.60),grey(1)),
       cex=2,
       axes=FALSE
       )
  segments(0,0,25.5,0,lty = 2)
  segments(0,0,0,25.5,lty = 2)
  segments(25.5,0,25.5,12.5,lty = 2)
  segments(12.5,12.5,25.5,12.5,lty = 2)
  segments(12.5,12.5,12.5,25.5,lty = 2)
  segments(0,25.5,12.5,25.5,lty = 2)
  segments(0,10,10,10,lty = 2)
  segments(10,0,10,10,lty = 2)
  segments(10,10,12.5,12.5,lty = 2)
  mtext(expression(-log[10](p-value)),side=1,cex=2,padj=2.0,adj=0.5)
  mtext(x,side=1,cex=2,padj=2.7,adj=1)
  mtext(expression(-log[10](p-value)),side=2,cex=2,padj=-1.5,adj=0.5)
  mtext(y,side=2,cex=2,padj=-2.5,adj=1)
  legend(13,25,
         ncol=1,
         x.intersp=0.6,
         y.intersp=0.4,
         cex=1.5,
         pt.cex=2,
         pch = 22,
         pt.bg=c(grey(0.25),grey(0.60),grey(1)),
         c("Plasma membrane","Organelle membrane","Cytosol"),
         box.lwd = 0,box.col = "white",bg = "white")
  axis(
    2,
    at=c(0,5,10,15,20,25),
    labels=c(0,5,10,15,20,25),
    tick=TRUE,
    lty="solid",
    lwd=2,
    lwd.ticks=2,
    col=NULL,
    col.ticks=NULL,
    cex.axis=1.5,
    #yaxp = c(20,40,4),
    las=1)
  axis(
    1,
    at=c(0,5,10,15,20,25),
    labels=c(0,5,10,15,20,25),
    tick=TRUE,
    lty="solid",
    lwd=2,
    lwd.ticks=2,
    col=NULL,
    col.ticks=NULL,
    cex.axis=1.5,
    padj=0,
    pos=-1,
    #yaxp = c(20,40,4),
    las=1)
}

# Do the plots
fun.pp(B10T10.plot,"T10","B10")
fun.pp(B2T2.plot,"T2","B2")
```

REFERENCES

Alconada, A., Bauer, U., and Hoflack, B. (1996). A tyrosine-based motif and a casein kinase II phosphorylation site regulate the intracellular trafficking of the varicella-zoster virus glycoprotein I, a protein localized in the trans-Golgi network. EMBO J *15*, 6096–6110.

Altelaar, A.F.M., Frese, C.K., Preisinger, C., Hennrich, M.L., Schram, A.W., Timmers, H.T.M., Heck, A.J.R., and Mohammed, S. (2013). Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics *88*, 14–26.

Ando, K., Iijima, K.I., Elliott, J.I., Kirino, Y., and Suzuki, T. (2001). Phosphorylation-dependent regulation of the interaction of amyloid precursor protein with Fe65 affects the production of beta-amyloid. J Biol Chem *276*, 40353–40361.

Balklava, Z., Guscott, B., **Niehage, C.**, Currinn, H., Hoflack, B., and Wassmer, T. (2013a). Identification of the intracellular interactome of the Amyloid Precursor Protein establishes its role as a regulator of mTOR. Submitted to PLOS Biology.

Balklava, Z., **Niehage, C.**, Currinn, H., Guscott, B., Hoflack, B., and Wassmer, T. (2013b). The amyloid precursor protein is a regulator of the cellular phosphoinositide metabolism via a direct interaction with the PIKfyve complex. Submitted to Nat Cell Biol.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem *389*, 1017–1031.

Baust, T., Czupalla, C., Krause, E., Bourel-Bonnet, L., and Hoflack, B. (2006). Proteomic analysis of adaptor protein 1A coats selectively assembled on liposomes. Proc Natl Acad Sci U S A *103*, 3159–3164.

Baust, T., Anitei, M., Czupalla, C., Parshyna, I., Bourel, L., Thiele, C., Krause, E., and Hoflack, B. (2008). Protein networks supporting AP-3 function in targeting lysosomal membrane proteins. Mol Biol Cell *19*, 1942–1951.

Behnia, R., and Munro, S. (2005). Organelle identity and the signposts for membrane traffic. Nature *438*, 597–604.

Benedikt, J., Hecimovic, A., Ellerweg, D., and von Keudell, A. (2012). Quadrupole mass spectrometry of reactive plasmas. Journal of Physics D: Applied Physics *45*.

Bhalla, A., Vetanovetz, C.P., Morel, E., Chamoun, Z., Di Paolo, G., and Small, S.A. (2012). The location and trafficking routes of the neuronal retromer and its role in amyloid precursor protein transport. Neurobiol Dis *47*, 126–134.

Blanksby, S.J., and Mitchell, T.W. (2010). Advances in mass spectrometry for lipidomics. Annu Rev Anal Chem (Palo Alto Calif) *3*, 433–465.

Blaum, K. (2006). High-accuracy mass spectrometry with stored ions. Physics Reports *425*, 1–78.

Bonifacino, J.S. (2004). The GGA proteins: adaptors on the move. Nat Rev Mol Cell Biol *5*, 23–32.

Bonifacino, J.S., and Rojas, R. (2006). Retrograde transport from endosomes to the trans-Golgi network. Nat Rev Mol Cell Biol *7*, 568–579.

Bonifacino, J.S., and Traub, L.M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem *72*, 395–447.

Bourel-Bonnet, L., Pécheur, E.-I., Grandjean, C., Blanpain, A., Baust, T., Melnyk, O., Hoflack, B., and Gras-Masse, H. (2005). Anchorage of synthetic peptides onto liposomes via hydrazone and alpha-oxo hydrazone bonds. preliminary functional investigations. Bioconjug Chem *16*, 450–457.

Brosch, M., and Choudhary, J. (2010). Scoring and validation of tandem MS peptide identification methods. Methods Mol Biol *604*, 43–53.

Buerger, C., DeVries, B., and Stambolic, V. (2006). Localization of Rheb to the endomembrane is critical for its signaling function. Biochem Biophys Res Commun *344*, 869–880.

Burgos, P.V., Mardones, G.A., Rojas, A.L., daSilva, L.L.P., Prabhu, Y., Hurley, J.H., and Bonifacino, J.S. (2010). Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. Dev Cell *18*, 425–436.

Byekova, Y.A., Powell, R.R., Welter, B.H., and Temesvari, L.A. (2010). Localization of phosphatidylinositol (3,4,5)-trisphosphate to phagosomes in entamoeba histolytica achieved using glutathione S-transferase- and green fluorescent protein-tagged lipid biosensors. Infect Immun *78*, 125–137.

Caporaso, G.L., Takei, K., Gandy, S.E., Matteoli, M., Mundigl, O., Greengard, P., and De Camilli, P. (1994). Morphologic and biochemical analysis of the intracellular trafficking of the Alzheimer beta/A4 amyloid precursor protein. J Neurosci *14*, 3122–3138.

Catimel, B., Schieber, C., Condron, M., Patsiouras, H., Connolly, L., Catimel, J., Nice, E.C., Burgess, A.W., and Holmes, A.B. (2008). The PI(3,5)P2 and PI(4,5)P2 interactomes. J Proteome Res *7*, 5295–5313.

Catimel, B., Yin, M.-X., Schieber, C., Condron, M., Patsiouras, H., Catimel, J., Robinson, D.E.J.E., Wong, L.S.-M., Nice, E.C., Holmes, A.B., et al. (2009). PI(3,4,5)P3 Interactome. J Proteome Res *8*, 3712–3726.

Chakrabarti, A., and Mukhopadhyay, D. (2012). Novel adaptors of amyloid precursor protein intracellular domain and their functional implications. Genomics Proteomics Bioinformatics *10*, 208–216.

Chelius, D., and Bondarenko, P.V. (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res *1*, 317–323.

Clague, M.J., and Urbé, S. (2001). The interface of receptor trafficking and signalling. J Cell Sci *114*, 3075–3081.

Cottrell, J.S. (1994). Protein identification by peptide mass fingerprinting. Pept Res *7*, 115–124.

Cox, J., and Mann, M. (2012). 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. BMC Bioinformatics *13 Suppl 16*, S12.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367–1372.

Crottet, P., Meyer, D.M., Rohrer, J., and Spiess, M. (2002). ARF1.GTP, tyrosine-based signals, and phosphatidylinositol 4,5-bisphosphate constitute a minimal machinery to recruit the AP-1 clathrin adaptor to membranes. Mol Biol Cell *13*, 3672–3682.

Crump, C.M., Xiang, Y., Thomas, L., Gu, F., Austin, C., Tooze, S.A., and Thomas, G. (2001). PACS-1 binding to adaptors is required for acidic cluster motif-mediated protein traffic. EMBO J *20*, 2191–2201.

Cullen, P.J., and Korswagen, H.C. (2012). Sorting nexins provide diversity for retromer-dependent trafficking events. Nat Cell Biol *14*, 29–37.

Dass, C. (2007). Fundamentals of Contemporary Mass Spectrometry (John Wiley & Sons, Hoboken NJ USA).

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D.F., Burkhard, P.R., and Sanchez, J.-C. (2008). Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem *80*, 2921–2931.

De Matteis, M.A., Di Campli, A., and Godi, A. (2005). The role of the phosphoinositides at the Golgi complex. Biochim Biophys Acta *1744*, 396–405.

Di Paolo, G., and De Camilli, P. (2006). Phosphoinositides in cell regulation and membrane dynamics. Nature *443*, 651–657.

Donaldson, J.G., and Jackson, C.L. (2011). ARF family G proteins and their regulators: roles in membrane transport, development and disease. Nat Rev Mol Cell Biol *12*, 362–375.

Draghici, S. (2013). Statistics and Data Analysis for Microarrays Using R and Bioconductor (CRC Press, Boca Raton FL USA).

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

Favoreel, H.W. (2006). The why's of Y-based motifs in alphaherpesvirus envelope proteins. Virus Res *117*, 202–208.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64–71.

Fernández, L.E.M. (2007). Introduction to ion trap mass spectrometry: Application to the structural characterization of plant oligosaccharides. Carbohydrate Polymers *68*, 797–807.

Fjorback, A.W., and Andersen, O.M. (2012). SorLA is a molecular link for retromer-dependent sorting of the Amyloid precursor protein. Commun Integr Biol *5*, 616–619.

Goebel-Stengel, M., Stengel, A., Taché, Y., and Reeve, Jr, J.R. (2011). The importance of using the optimal plasticware and glassware in studies involving peptides. Anal Biochem *414*, 38–46.

Gorshkov, A.V., Tarasova, I.A., Evreinov, V.V., Savitski, M.M., Nielsen, M.L., Zubarev, R.A., and Gorshkov, M.V. (2006). Liquid chromatography at critical conditions: comprehensive approach to sequence-dependent retention time prediction. Anal Chem *78*, 7770–7777.

Govero, J., Hall, S., and Heineman, T.C. (2007). Intracellular localization of varicella-zoster virus ORF39 protein and its functional relationship to glycoprotein K. Virology *358*, 291–302.

Griffin, N.M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J.A., and Schnitzer, J.E. (2010). Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. Nat Biotechnol *28*, 83–89.

Gronthos, S., Mankani, M., Brahim, J., Robey, P.G., and Shi, S. (2000). Postnatal human dental pulp stem cells (DPSCs) in vitro and in vivo. Proc Natl Acad Sci U S A *97*, 13625–13630.

Gulino, A., Di Marcotullio, L., and Screpanti, I. (2010). The multiple functions of Numb. Exp Cell Res *316*, 900–906.

Gustafsen, C., Glerup, S., Pallesen, L.T., Olsen, D., Andersen, O.M., Nykjær, A., Madsen, P., and Petersen, C.M. (2013). Sortilin and SorLA display distinct roles in processing and trafficking of amyloid precursor protein. J Neurosci *33*, 64–71.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol *17*, 994–999.

Haass, C., Koo, E.H., Mellon, A., Hung, A.Y., and Selkoe, D.J. (1992). Targeting of cell-surface beta-amyloid precursor protein to lysosomes: alternative processing into amyloid-bearing fragments. Nature *357*, 500–503.

Hakulinen, J., Sankkila, L., Sugiyama, N., Lehti, K., Keski-Oja, J. (2008). Secretion of active membrane type 1 matrix metalloproteinase (MMP-14) into extracellular space in microvesicular exosomes. J. Cell. Biochem. 105, 1211–1218.

Hanson, P.I., Shim, S., and Merrill, S.A. (2009). Cell biology of the ESCRT machinery. Curr Opin Cell Biol *21*, 568–574.

Hata, Y., Slaughter, C.A., and Südhof, T.C. (1993). Synaptic vesicle fusion complex contains unc-18 homologue bound to syntaxin. Nature *366*, 347–351.

Henne, W.M., Buchkovich, N.J., and Emr, S.D. (2011). The ESCRT pathway. Dev Cell *21*, 77–91.

Hirst, J., Barlow, L.D., Francisco, G.C., Sahlender, D.A., Seaman, M.N.J., Dacks, J.B., and Robinson, M.S. (2011). The fifth adaptor protein complex. PLoS Biol *9*, e1001170.

Hirst, J., Irving, C., and Borner, G.H.H. (2013). Adaptor protein complexes AP-4 and AP-5: new players in endosomal trafficking and progressive spastic paraplegia. Traffic *14*, 153–164.

Hoffmann, E. de (2007). Mass Spectrometry: Principles and Applications (John Wiley & Sons, Hoboken NJ USA).

Hsueh, Y.P., Yang, F.C., Kharazia, V., Naisbitt, S., Cohen, A.R., Weinberg, R.J., and Sheng, M. (1998). Direct interaction of CASK/LIN-2 and syndecan heparan sulfate proteoglycan and their overlapping distribution in neuronal synapses. J Cell Biol *142*, 139–151.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol *189*, 739–754.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics *4*, 1265–1272.

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M.J., and Frishman, D. (2008). Protein abundance profiling of the Escherichia coli cytosol. BMC Genomics *9*, 102.

Jin, N., Chow, C.Y., Liu, L., Zolov, S.N., Bronson, R., Davisson, M., Petersen, J.L., Zhang, Y., Park, S., Duex, J.E., et al. (2008). VAC14 nucleates a protein complex essential for the acute interconversion of PI3P and PI(3,5)P(2) in yeast and mouse. EMBO J *27*, 3221–3234.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299–2301.

Karbanová, J., Soukup, T., Suchánek, J., and Mokrý, J. (2010). Osteogenic differentiation of human dental pulp-derived stem cells under various ex-vivo culture conditions. Acta Medica (Hradec Kralove) *53*, 79–84.

Karbanová, J., Soukup, T., Suchánek, J., Pytlík, R., Corbeil, D., and Mokrý, J. (2011). Characterization of dental pulp stem cells from impacted third molars cultured in low serum-containing medium. Cells Tissues Organs *193*, 344–365.

Kawashima, N. (2012). Characterisation of dental pulp stem cells: a new horizon for tissue regeneration? Arch Oral Biol *57*, 1439–1458.

Kim, M.-S., Kandasamy, K., Chaerkady, R., and Pandey, A. (2010). Assessment of resolution parameters for CID-based shotgun proteomic experiments on the LTQ-Orbitrap mass spectrometer. J Am Soc Mass Spectrom *21*, 1606–1611.

King, G.D., and Scott Turner, R. (2004). Adaptor protein interactions: modulators of amyloid precursor protein metabolism and Alzheimer's disease risk? Exp Neurol *185*, 208–219.

Kirisits, A., Pils, D., and Krainer, M. (2007). Epidermal growth factor receptor degradation: an alternative view of oncogenic pathways. Int J Biochem Cell Biol *39*, 2173–2182.

Krauss, M., and Haucke, V. (2007). Phosphoinositides: regulators of membrane traffic and protein function. FEBS Lett *581*, 2105–2111.

Kudlicki, A. (2012). The optimal exponent base for emPAI is 6.5. PLoS One *7*, e32339.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lee, J., Retamal, C., Cuitiño, L., Caruano-Yzermans, A., Shin, J.-E., van Kerkhof, P., Marzolo, M.-P., and Bu, G. (2008). Adaptor protein sorting nexin 17 regulates amyloid precursor protein trafficking and processing in the early endosomes. J Biol Chem *283*, 11501–11508.

Leth-Larsen, R., Lund, R.R., and Ditzel, H.J. (2010). Plasma membrane proteomics and its application in clinical cancer biomarker discovery. Mol Cell Proteomics *9*, 1369–1382.

Lin, T.C.Y., and Lee, O.K.S. (2008). Stem cells: a primer. Chin J Physiol *51*, 197–207.

Liu, H., Sadygov, R.G., and Yates, 3rd, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem *76*, 4193–4201.

Liu, K., Zhang, J., Wang, J., Zhao, L., Peng, X., Jia, W., Ying, W., Zhu, Y., Xie, H., He, F., et al. (2009). Relationship between sample loading amount and peptide identification and its effects on quantitative proteomics. Anal Chem *81*, 1307–1314.

Lorenz, K., Rupf, T., Salvetter, J., and Bader, A. (2009). Enrichment of human beta 1 bri/alpha 6 bri/CD71 dim keratinocytes after culture in defined media. Cells Tissues Organs *189*, 382–390.

Makarov (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem *72*, 1156–1162.

McGough, I.J., and Cullen, P.J. (2013). Clathrin is not required for SNX-BAR-retromer-mediated carrier formation. J Cell Sci *126*, 45–52.

Mindaye, S.T., Ra, M., Lo Surdo, J., Bauer, S.R., and Alterman, M.A. (2013). Improved proteomic profiling of the cell surface of culture-expanded human bone marrow multipotent stromal cells. J Proteomics *78*, 1–14.

Mo, C., Lee, J., Sommer, M.H., and Arvin, A.M. (2003). Varicella-zoster virus infection facilitates VZV glycoprotein E trafficking to the membrane surface of melanoma cells. J Med Virol *70 Suppl 1*, S56–S58.

Muresan, Z., and Muresan, V. (2005). Coordinated transport of phosphorylated amyloid-beta precursor protein and c-Jun NH2-terminal kinase-interacting protein-1. J Cell Biol *171*, 615–625.

Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. (2012). System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Mol Cell Proteomics *11*, M111.013722.

Nalivaeva, N.N., and Turner, A.J. (2013). The amyloid precursor protein: a biochemical enigma in brain development, function and disease. FEBS Lett *587*, 2046–2054.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics *4*, 1419–1440.

Niehage, C., Stange, C., Anitei, M., and Hoflack, B. (2013). Liposome-based assays to study membrane-associated protein networks. Meth Enzymol *534*.

Niehage, C., Steenblock, C., Pursche, T., Bornhäuser, M., Corbeil, D., Hoflack, B. (2011). The cell surface proteome of human mesenchymal stromal cells. PLoS ONE 6, e20399.

Ning, K., Fermin, D., and Nesvizhskii, A.I. (2012). Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res *11*, 2261–2271.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics *4*, 1487–1502.

Ong, S.-E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol *1*, 252–262.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics *1*, 376–386.

Oppelt, A., Lobert, V.H., Haglund, K., Mackey, A.M., Rameh, L.E., Liestøl, K., Schink, K.O., Pedersen, N.M., Wenzel, E.M., Haugsten, E.M., et al. (2013). Production of phosphatidylinositol 5-phosphate via PIKfyve and MTMR3 regulates cell migration. EMBO Rep *14*, 57–64.

Paleotti, O., Macia, E., Luton, F., Klein, S., Partisani, M., Chardin, P., Kirchhausen, T., and Franco, M. (2005). The small G-protein Arf6GTP recruits the AP-2 adaptor complex to membranes. J Biol Chem *280*, 21661–21666.

Pan, C., Kumar, C., Bohl, S., Klingmueller, U., and Mann, M. (2009). Comparative proteomic

phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. Mol Cell Proteomics *8*, 443–450.

Park, S.K., Liao, L., Kim, J.Y., and Yates, 3rd, J.R. (2009). A computational approach to correct arginine-to-proline conversion in quantitative proteomics. Nat Methods *6*, 184–185.

Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods *9*, 555–566.

Pocha, S.M., Wassmer, T., **Niehage, C.**, Hoflack, B., and Knust, E. (2011). Retromer controls epithelial cell polarity by trafficking the apical determinant Crumbs. Curr Biol *21*, 1111–1117.

Podwojski, K., Eisenacher, M., Kohl, M., Turewicz, M., Meyer, H.E., Rahnenführer, J., and Stephan, C. (2010). Peek a peak: a glance at statistics for quantitative label-free proteomics. Expert Rev Proteomics *7*, 249–261.

Popoff, V., Mardones, G.A., Tenza, D., Rojas, R., Lamaze, C., Bonifacino, J.S., Raposo, G., and Johannes, L. (2007). The retromer complex and clathrin define an early endosomal retrograde exit site. J Cell Sci *120*, 2022–2031.

Prewitz, M.C., Seib, F.P., von Bonin, M., Friedrichs, J., Stißel, A., **Niehage, C.**, Müller, K., Anastassiadis, K., Waskow, C., Hoflack, B., Bornhäuser, M., Werner, C. (2013). Tightly anchored tissue-mimetic matrices as instructive stem cell microenvironments. Nat. Methods 10, 788–794.

Priller, C., Bauer, T., Mitteregger, G., Krebs, B., Kretzschmar, H.A., and Herms, J. (2006). Synapse formation and function is modulated by the amyloid precursor protein. J Neurosci *26*, 7212–7221.

Rajendran, L., and Annaert, W. (2012). Membrane trafficking pathways in Alzheimer's disease. Traffic *13*, 759–770.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231–1245.

Robinson, M.S., and Bonifacino, J.S. (2001). Adaptor-related proteins. Curr Opin Cell Biol *13*, 444–453.

Rogelj, B., Mitchell, J.C., Miller, C.C.J., and McLoughlin, D.M. (2006). The X11/Mint family of adaptor proteins. Brain Res Rev *52*, 305–315.

Roncarati, R., Sestan, N., Scheinfeld, M.H., Berechid, B.E., Lopez, P.A., Meucci, O., McGlade, J.C., Rakic, P., and D'Adamio, L. (2002). The gamma-secretase-generated intracellular domain of beta-amyloid precursor protein binds Numb and inhibits Notch signaling. Proc Natl Acad Sci U S A *99*, 7102–7107.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154–1169.

Roxas, B.A.P., and Li, Q. (2008). Significance analysis of microarray for relative quantitation of LC/MS data in proteomics. BMC Bioinformatics *9*, 187.

Rucevic, M., Hixson, D., and Josic, D. (2011). Mammalian plasma membrane proteins as potential biomarkers and drug targets. Electrophoresis *32*, 1549–1564.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwartz, J.C., Senko, M.W. (2002). A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. J Am Soc Mass Spectrom *13*, 659–669.

Seaman, M.N.J. (2005). Recycle your receptors with retromer. Trends Cell Biol *15*, 68–75.

Segev, N. (2011). GTPases in intracellular trafficking: an overview. Semin Cell Dev Biol *22*, 1–2.

Shilo, B.-Z., and Schejter, E.D. (2011). Regulation of developmental intercellular signalling by intracellular trafficking. EMBO J *30*, 3516–3526.

Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol *10*, 513–525.

Suri, G., Spiess, M., and Crottet, P. (2008). Recruitment of coat proteins to peptidoliposomes. Methods Mol Biol *457*, 227–239.

Suzuki, T., and Nakaya, T. (2008). Regulation of amyloid beta-protein precursor by phosphorylation and protein interactions. J Biol Chem *283*, 29633–29637.

Suzuki, T., Ando, K., Iijima, K., Oguchi, S., and Takeda, S. (2000). Phosphorylation of Amyloid Precursor Protein (APP) Family Proteins. Methods Mol Med *32*, 271–282.

Tabuchi, K., Biederer, T., Butz, S., and Sudhof, T.C. (2002). CASK participates in alternative tripartite complexes in which Mint 1 competes for binding with caskin 1, a novel CASK-binding protein. J Neurosci *22*, 4264–4273.

Tanaka K., Waki H., Ido Y., Akita S., Yoshida Y., Y.T. (1988). Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. *2*, 151.

Tepass, U. (2012). The apical polarity protein network in Drosophila epithelial cells: regulation of polarity, junctions, morphogenesis, cell growth, and survival. Annu Rev Cell Dev Biol *28*, 655–685.

Tortorella, L.L., Schapiro, F.B., and Maxfield, F.R. (2007). Role of an acidic cluster/dileucine motif in cation-independent mannose 6-phosphate receptor traffic. Traffic *8*, 402–413.

Treiber, C., Quadir, M.A., Voigt, P., Radowski, M., Xu, S., Munter, L.-M., Bayer, T.A., Schaefer, M., Haag, R., and Multhaup, G. (2009). Cellular copper import by nanocarrier systems, intracellular availability, and effects on amyloid beta peptide secretion. Biochemistry *48*, 4273–4284.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A *98*, 5116–5121.

Van Meer, G., Voelker, D.R., and Feigenson, G.W. (2008). Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol *9*, 112–124.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Vicinanza, M., D'Angelo, G., Di Campli, A., and De Matteis, M.A. (2008). Function and dysfunction of the PI system in membrane trafficking. EMBO J *27*, 2457–2470.

Villas-Bôas, S.G., Mas, S., Akesson, M., Smedsgaard, J., and Nielsen, J. (2005). Mass spectrometry in metabolome analysis. Mass Spectrom Rev *24*, 613–646.

Vuckovic, D., Dagley, L.F., Purcell, A.W., and Emili, A. (2013). Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges. Proteomics *13*, 404–423.

Wenk, M.R., and De Camilli, P. (2003). Assembly of endocytosis-associated proteins on liposomes. Methods Enzymol *372*, 248–260.

Yang, M., Virassamy, B., Vijayaraj, S.L., Lim, Y., Saadipour, K., Wang, Y.-J., Han, Y.-C., Zhong, J.-H., Morales, C.R., and Zhou, X.-F. (2013). The intracellular domain of sortilin interacts with amyloid precursor protein and regulates its lysosomal and lipid raft trafficking. PLoS One *8*, e63049.

Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L., and Samatova, N.F. (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. J Proteome Res *5*, 2909–2918.

Zhang, H., Li, X.-J., Martin, D.B., and Aebersold, R. (2003b). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol *21*, 660–666.

Zhang, W., Zhou, G., Zhao, Y., White, M.A., and Zhao, Y. (2003a). Affinity enrichment of plasma membrane for proteomics analysis. Electrophoresis *24*, 2855–2863.

Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L. (2009). Effect of dynamic exclusion duration on spectral count based quantitative proteomics. Anal Chem *81*, 6317–6326.

Zhu, Y., Drake, M.T., and Kornfeld, S. (1999). ADP-ribosylation factor 1 dependent clathrin-coat assembly on synthetic liposomes. Proc Natl Acad Sci U S A *96*, 5013–5018.

Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res *5*, 2339–2347.

6. SUPPLEMENT

Table S.1: CD-marker proteins identified.

CD	AccNo	B	T	Name
CD9	P21926	++++	+++	CD9 antigen
CD10	P08473	++++	+++	Neprilysin
CD13	P15144	++++	++++	Aminopeptidase N
CD14	P08571	++	-	Monocyte differentiation antigen CD14
CD29	P05556	++++	++++	Integrin beta-1
CD39	P49961	++	-	Ectonucleoside triphosphate diphosphohydrolase 1
CD40	P25942	+++	-	Tumor necrosis factor receptor superfamily member 5
CD44	P16070	++++	++++	CD44 antigen
CD46	P15529	+++	-	Membrane cofactor protein
CD47	Q08722	++++	++	Leukocyte surface antigen CD47
CD49a	P56199	++++	-	Integrin alpha-1
CD49b	P17301	++++	++++	Integrin alpha-2
CD49c	P26006	++++	+++	Integrin alpha-3
CD49d	P13612	++++	-	Integrin alpha-4
CD49e	P08648	++++	+++	Integrin alpha-5
CD49f	P23229	+++	-	Integrin alpha-6
CD51	P06756	++++	+++	Integrin alpha-V
CD54	P05362	++	-	Intercellular adhesion molecule 1
CD55	P08174	+++	-	Complement decay-accelerating factor
CD56	P13591	++	-	Neural cell adhesion molecule 1
CD58	P19256	++	-	Lymphocyte function-associated antigen 3
CD59	P13987	++++	+++	CD59 glycoprotein
CD61	P05106	+++	-	Integrin beta-3
CD63	P08962	+++	++++	CD63 antigen
CD68	P34810	++	-	Macrosialin
CD71	P02786	++++	++++	Transferrin receptor protein 1
CD73	P21589	++++	++++	5-nucleotidase
CD81	P60033	++++	+++	CD81 antigen
CD82	P27701	++	-	CD82 antigen
CD87	Q03405	++	-	Urokinase plasminogen activator surface receptor
CD88	P21730	++	-	C5a anaphylatoxin chemotactic receptor
CD90	P04216	++++	+++	Thy-1 membrane glycoprotein
CD91	Q07954	++++	++++	Prolow-density lipoprotein receptor-related protein 1
CD92	Q8WWI5	++++	++	Choline transporter-like protein 1
CD95	P25445	+++	-	Tumor necrosis factor receptor superfamily member 6
CD97	P48960	++++	-	CD97 antigen
CD98	P08195	++++	+++	4F2 cell-surface antigen heavy chain
CD99	P14209	+++	-	CD99 antigen
CD105	P17813	++++	-	Endoglin
CD106	P19320	++	-	Vascular cell adhesion protein 1
CD107a	P11279	+++	++++	Lysosome-associated membrane glycoprotein 1
CD107b	P13473	-	+++	Lysosome-associated membrane glycoprotein 2
CD108	O75326	++	++	Semaphorin-7A
CD109	Q6YHK3	++++	++++	CD109 antigen
CD112	Q92692	++++	-	Poliovirus receptor-related protein 2
CD113	Q9NQS3	++++	-	Poliovirus receptor-related protein 3
CD119	P15260	++	-	Interferon gamma receptor 1
CD120a	P19438	++	-	Tumor necrosis factor receptor superfamily member 1A
CD121a	P14778	++	-	Interleukin-1 receptor type 1
CD130	P40189	+++	-	Interleukin-6 receptor subunit beta
CD140a	P16234	++++	++	Alpha-type platelet-derived growth factor receptor
CD140b	P09619	++++	+++	Beta-type platelet-derived growth factor receptor
CD141	P07204	++	-	Thrombomodulin
CD142	P13726	++	-	Tissue factor
CD143	P12821	++++	-	Angiotensin-converting enzyme
CD146	P43121	+++	-	Cell surface glycoprotein MUC18
CD147	P35613	++++	+++	Basigin
CD148	Q12913	++++	-	Receptor-type tyrosine-protein phosphatase eta
CD151	P48509	++++	+++	CD151 antigen
CD155	P15151	++++	-	Poliovirus receptor
CD156b	P78536	+++	-	Disintegrin and metalloproteinase domain-containing protein 17
CD156c	O14672	++++	++	Disintegrin and metalloproteinase domain-containing protein 10
CD157	Q10588	++	-	ADP-ribosyl cyclase 2

Table S.1 - continued: CD-marker proteins identified.

CD	AccNo	B	T	Name
CD166	Q13740	++++	++++	CD166 antigen
CD167b	Q16832	++++	++	Discoidin domain-containing receptor 2
CD172a	P78324	+++	-	Tyrosine-protein phosphatase non-receptor type substrate 1
CD201	Q9UNN8	++++	-	Endothelial protein C receptor
CD203a	P22413	++++	-	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1
CD213a1	P78552	+	-	Interleukin-13 receptor subunit alpha-1
CD217	Q96F46	+++	-	Interleukin-17 receptor A
CD221	P08069	+++	-	Insulin-like growth factor 1 receptor
CD222	P11717	++++	+++	Cation-independent mannose-6-phosphate receptor
CD224	P19440	++	-	Gamma-glutamyltranspeptidase 1
CD230	P04156	++	-	Major prion protein
CD232	O60486	++	-	Plexin-C1
CD236	P04921	+++	-	Glycophorin-C
CD239	P50895	+++	-	Basal cell adhesion molecule
CD248	Q9HCU0	++++	+++	Endosialin
CD261	O00220	++	-	Tumor necrosis factor receptor superfamily member 10A
CD262	O14763	+++	-	Tumor necrosis factor receptor superfamily member 10B
CD264	Q9UBN6	++	-	Tumor necrosis factor receptor superfamily member 10D
CD266	Q9NP84	++	-	Tumor necrosis factor receptor superfamily member 12A
CD273	Q9BQ51	+++	-	Programmed cell death 1 ligand 2
CD274	Q9NZQ7	++	-	Programmed cell death 1 ligand 1
CD276	Q5ZPR3	++++	++	CD276 antigen
CD277	O00481	++	-	Butyrophilin subfamily 3 member A1
CD280	Q9UBG0	++++	+++	C-type mannose receptor 2
CD282	O60603	+	-	Toll-like receptor 2
CD292	P36894	++	-	Bone morphogenetic protein receptor type-1A
CD295	P48357	+++	-	Leptin receptor
CD298	P54709	++++	+++	Sodium/potassium-transporting ATPase subunit beta-3
CD302	Q8IX05	++	-	CD302 antigen
CD304	O14786	++++	+++	Neuropilin-1
CD316	Q969P0	++++	-	Immunoglobulin superfamily member 8
CD318	Q9H5V8	+++	-	CUB domain-containing protein 1
CD325	P19022	+++	-	Cadherin-2
CD331	P11362	+++	-	Basic fibroblast growth factor receptor 1
CD339	P78504	++	-	Protein jagged-1
CD340	P04626	+++	-	Receptor tyrosine-protein kinase erbB-2
CD362	P34741	++	-	Syndecan-2
CDw210b	Q08334	++	-	Interleukin-10 receptor subunit beta

Not identified CD-marker proteins: CD1a, CD1b, CD1c, CD1d, CD1e, CD2, CD3d, CD3e, CD3g, CD4, CD5, CD6, CD7, CD8a, CD8b, CD11a, CD11b, CD11c, CD11d, CD16a, CD16b, CD18, CD19, CD20, CD21, CD22, CD23, CD24, CD25, CD26, CD27, CD28, CD30, CD31, CD32, CD33, CD34, CD35, CD36, CD37, CD38, CD41, CD42a, CD42b, CD42c, CD42d, CD43, CD45, CD48, CD50, CD52, CD53, CD62E, CD62L, CD62P, CD64, CD66a, CD66b, CD66c, CD66d, CD66e, CD66f, CD69, CD70, CD72, CD74, CD79a, CD79b, CD80, CD83, CD84, CD85a, CD85c, CD85d, CD85e, CD85f, CD85g, CD85h, CD85i, CD85j, CD85k, CD86, CD89, CD93, CD94, CD96, CD100, CD101, CD102, CD103, CD104, CD110, CD111, CD114, CD115, CD116, CD117, CD118, CD120b, CD121b, CD122, CD123, CD124, CD125, CD126, CD127, CD129, CD131, CD132, CD133, CD134, CD135, CD136, CD137, CD138, CD144, CD150, CD152, CD153, CD154, CD156a, CD158a, CD158b1, CD158b2, CD158c, CD158d, CD158e, CD158f1, CD158f2, CD158g, CD158h, CD158i, CD158j, CD158k, CD158z, CD159a, CD159c, CD160, CD161, CD162, CD163, CD163b, CD164, CD167a, CD168, CD169, CD170, CD171, CD172b, CD172g, CD177, CD178, CD179a, CD179b, CD180, CD181, CD182, CD183, CD184, CD185, CD186, CD191, CD192, CD193, CD194, CD195, CD196, CD197, CDw198, CDw199, CD200, CD202b, CD203c, CD204, CD205, CD206, CD207, CD208, CD209, CD210, CD212, CD213a2, CD215, CD218a, CD218b, CD220, CD223, CD225, CD226, CD227, CD228, CD229, CD231, CD233, CD234, CD235a, CD235b, CD238, CD240CE, CD240D, CD241, CD242, CD243, CD244, CD246, CD247, CD249, CD252, CD253, CD254, CD256, CD257, CD258, CD263, CD265, CD267, CD268, CD269, CD270, CD271, CD272, CD275, CD278, CD279, CD281, CD283, CD284, CD286, CD288, CD289, CD290, CDw293, CD294, CD296, CD297, CD299, CD300a, CD300b, CD300c, CD300d, CD300e, CD300f, CD300g, CD301, CD303, CD305, CD306, CD307a, CD307b, CD307c, CD307d, CD307e, CD309, CD312, CD314, CD315, CD317, CD319, CD320, CD321, CD322, CD324, CD326, CD327, CD328, CD329, CD332, CD333, CD334, CD335, CD336, CD337, CD338, CD344, CD349, CD350, CD351, CD352, CD353, CD354, CD355, CD357, CD358, CD360, CD361, CD363

Table S.2 : non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
	+		+				Q15125	++	-	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomerase
+			+	+		+	P08253	++	-	72 kDa type IV collagenase
+	+		+				Q04771	+++	-	Activin receptor type-1
	+		+				P51828	++	-	Adenylate cyclase type 7
	+		+				O60503	+++	-	Adenylate cyclase type 9
+	+		+				Q9H6B4	++++	-	Adipocyte adhesion molecule
+			+	+			Q8IUX7	++	-	Adipocyte enhancer-binding protein 1
	+		+				P12235	+	-	ADP/ATP translocase 1
	+		+				P05141	+++	+++	ADP/ATP translocase 2
+			+				Q9BRR6	++	++	ADP-dependent glucokinase
+			+	+			P43652	++	-	Afamin
+		+	+	+	+		P05186	++	-	Alkaline phosphatase, tissue-nonspecific isozyme
+			+				P04217	++	-	Alpha-1B-glycoprotein
+			+	+	+		P02765	++	-	Alpha-2-HS-glycoprotein
+		+		+			P30533	+++	-	Alpha-2-macroglobulin receptor-associated protein
+	+		+				Q86SJ2	++	-	Amphoterin-induced protein 2
+			+	+			Q9UKU9	++	-	Angiopoietin-related protein 2
	+		+				P04920	++	-	Anion exchange protein 2
	+		+				Q9NW15	-	+++	Anoctamin-10
	+		+				Q4KMQ2	++++	-	Anoctamin-6
+	+		+				Q9H6X2	+++	-	Anthrax toxin receptor 1
+	+		+		+		P58335	+++	-	Anthrax toxin receptor 2
	+		+				Q03518	++	-	Antigen peptide transporter 1
	+		+				Q03519	++	-	Antigen peptide transporter 2
+			+	+	+		P04114	++	-	Apolipoprotein B-100
	+		+				Q12797	+++	++++	Aspartyl/asparaginyl beta-hydroxylase
	+		+				O95477	+++	-	ATP-binding cassette sub-family A member 1
	+		+				O94911	++	-	ATP-binding cassette sub-family A member 8
+	+		+				P20594	+++	-	Atrial natriuretic peptide receptor 2
+	+		+				P17342	++	-	Atrial natriuretic peptide receptor 3
	+		+	+			Q9Y5Q5	++	-	Atrial natriuretic peptide-converting enzyme
+	+		+	+	+		O75882	+++	-	Attractin
	+		+				P30411	+++	-	B2 bradykinin receptor
+			+	+	+	+	P98160	-	+++	Basement membrane-specific heparan sulfate proteoglycan core protein
	+		+				P51572	++	+++	B-cell receptor-associated protein 31
	+		+				P07550	++	-	Beta-2 adrenergic receptor
+			+	+	+		P61769	+++	-	Beta-2-microglobulin
	+		+				Q16585	+++	-	Beta-sarcoglycan
+	+		+				Q13873	+++	-	Bone morphogenetic protein receptor type-2
+	+		+				Q7KYR7	+++	-	Butyrophilin subfamily 2 member A1
	+		+				Q9P296	++	-	C5a anaphylatoxin chemotactic receptor C5L2
+	+		+				P55287	++++	-	Cadherin-11
+		+	+	+	+		P55290	++	-	Cadherin-13
+	+		+				P55283	++	-	Cadherin-4
+	+		+				P55285	++	++	Cadherin-6
	+		+				Q96D31	++	-	Calcium release-activated calcium channel protein 1
	+		+				Q12791	+++	-	Calcium-activated potassium channel subunit alpha-1
	+		+				Q9UJS0	-	+++	Calcium-binding mitochondrial carrier protein Aralar2
+			+	+	+	+	P27797	+++	++++	Calreticulin
+			+				O43852	+++	++++	Calumenin
	+		+				O15438	++	-	Canalicular multispecific organic anion transporter 2
+	+		+				O43570	++	-	Carbonic anhydrase 12
+		+	+				P14384	+++	-	Carboxypeptidase M
+				+	+		O75718	-	+++	Cartilage-associated protein
	+		+				P21964	+++	+++	Catechol O-methyltransferase
+			+	+	+		P07858	++	+++	Cathepsin B
+				+	+		P07339	++	+++	Cathepsin D
+				+	+		P43235	++	-	Cathepsin K
+				+	+		Q9UBR2	-	+++	Cathepsin Z
+	+		+				P20645	+++	-	Cation-dependent mannose-6-phosphate receptor
+	+		+				Q8TCZ2	+++	-	CD99 antigen-like protein 2
	+		+				O14735	++	+	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
+	+		+				Q9BY67	++	-	Cell adhesion molecule 1

Table S.2 - continued: non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
	+		+				Q99788	++	-	Chemokine-like receptor 1
	+		+				O00299	++	++++	Chloride intracellular channel protein 1
	+		+				Q9Y696	++	++++	Chloride intracellular channel protein 4
	+		+				Q8IWA5	++++	++	Choline transporter-like protein 2
+	+		+		+		Q6UVK1	++++	++++	Chondroitin sulfate proteoglycan 4
	+		+				Q9NY35	+++	-	Claudin domain-containing protein 1
	+		+				O96005	-	+++	Cleft lip and palate transmembrane protein 1
+			+	+	+		P02452	+++	++++	Collagen alpha-1(I) chain
+				+	+		P20908	++	++	Collagen alpha-1(V) chain
+			+	+	+		P12109	++	++++	Collagen alpha-1(VI) chain
+				+	+		Q02388	-	++	Collagen alpha-1(VII) chain
+			+	+	+		Q99715	-	++	Collagen alpha-1(XII) chain
+				+	+		P39060	++	++	Collagen alpha-1(XVIII) chain
+				+	+		P08123	++	+++	Collagen alpha-2(I) chain
+			+	+	+		P12110	++	++++	Collagen alpha-2(VI) chain
+			+	+	+		P12111	++++	++++	Collagen alpha-3(VI) chain
+				+	+		A6NMZ7	++	-	Collagen alpha-6(VI) chain
+			+	+	+		Q96CG8	++	++	Collagen triple helix repeat-containing protein 1
+	+		+				P78357	++	-	Contactin-associated protein 1
	+		+				Q9UPY5	++	-	Cystine/glutamate transporter
	+		+				Q07065	+++	++++	Cytoskeleton-associated protein 4
	+		+				Q92629	++++	-	Delta-sarcoglycan
+				+			P81605	++	++	Dermcidin
+	+		+				Q02413	++	-	Desmoglein-1
+	+		+				Q96PD2	+++	-	Discoidin, CUB and LCCL domain-containing protein 2
+	+		+	+	+		Q13443	++++	++	Disintegrin and metalloproteinase domain-containing protein 9
+				+			Q8IXB1	++	++	DnaJ homolog subfamily C member 10
+	+		+	+	+	+	Q14118	++++	-	Dystroglycan
+				+			O43854	++	-	EGF-like repeat and discoidin I-like domain-containing protein 3
	+		+				Q9Y6R1	+++	-	Electrogenic sodium bicarbonate cotransporter 1
+				+	+		Q9Y6C2	+	+++	EMILIN-1
+				+			O94919	++	-	Endonuclease domain-containing 1 protein
	+		+				P42892	+++	-	Endothelin-converting enzyme 1
+	+		+				P29317	++++	+++	Ephrin type-A receptor 2
+	+		+				P54764	+++	-	Ephrin type-A receptor 4
+	+		+				P54756	++	-	Ephrin type-A receptor 5
+	+		+				P29323	+++	-	Ephrin type-B receptor 2
+	+		+				P54753	++	-	Ephrin type-B receptor 3
+	+		+				P54760	++++	-	Ephrin type-B receptor 4
+	+		+	+			O15197	+++	-	Ephrin type-B receptor 6
+		+	+				P52803	++	-	Ephrin-A5
+	+		+				P98172	+++	-	Ephrin-B1
+	+		+				P52799	++	-	Ephrin-B2
+	+		+	+	+		P00533	++++	+++	Epidermal growth factor receptor
	+		+				O43556	+++	-	Epsilon-sarcoglycan
	+		+				Q99808	++	-	Equilibrative nucleoside transporter 1
	+		+				O94905	+++	+++	Erlin-2
+				+			Q8NBQ5	++	++	Estradiol 17-beta-dehydrogenase 11
	+		+				P43005	++	-	Excitatory amino acid transporter 3
	+		+				A0FGR8	+++	+++	Extended synaptotagmin-2
	+		+				O95864	+++	+++	Fatty acid desaturase 2
	+		+				Q9UPI3	++	-	Feline leukemia virus subgroup C receptor-related protein 2
+			+	+	+		P02751	+++	++++	Fibronectin
+			+	+	+		P23142	++	-	Fibulin-1
	+		+				P41440	++	-	Folate transporter 1
+	+		+				Q9UP38	++	-	Frizzled-1
+	+		+				Q14332	+++	-	Frizzled-2
+	+		+				O75084	+++	-	Frizzled-7
+			+	+	+		Q08380	++	-	Galectin-3-binding protein
+			+	+			Q92820	++	++	Gamma-glutamyl hydrolase
	+		+				P36269	++	-	Gamma-glutamyltransferase 5
	+		+				P17302	++++	+++	Gap junction alpha-1 protein
+		+	+				P56159	++	-	GDNF family receptor alpha-1

Table S.2 - continued: non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
+			+	+			P06396	-	++++	Gelsolin
+			+	+	+	+	P07093	+	-	Glia-derived nexin
+	+		+	+			P48060	+++	-	Glioma pathogenesis-related protein 1
+				+			Q96SL4	++	-	Glutathione peroxidase 7
+		+	+	+	+	+	P35052	++++	+++	Glypican-1
+		+	+	+	+	+	Q9Y625	++	-	Glypican-6
+	+					+	Q92896	++++	+++	Golgi apparatus protein 1
+	+		+				Q96PE1	+++	-	G-protein coupled receptor 124
+			+	+			O60565	++	+++	Gremlin-1
+		+	+				P54826	++	-	Growth arrest-specific protein 1
+			+	+			Q14393	+++	-	Growth arrest-specific protein 6
+			+	+			P00738	++	-	Haptoglobin
+			+	+			P02790	++	-	Hemopexin
+	+		+				Q30201	++	-	Hereditary hemochromatosis protein
	+		+				P30825	+++	-	High affinity cationic amino acid transporter 1
+	+		+				P30443	++++	-	HLA class I histocompatibility antigen, A-1 alpha chain
+	+		+				P01892	++++	-	HLA class I histocompatibility antigen, A-2 alpha chain
+	+		+				P05534	++	-	HLA class I histocompatibility antigen, A-24 alpha chain
+	+		+				P04439	-	+++	HLA class I histocompatibility antigen, A-3 alpha chain
+	+		+				P30481	++	-	HLA class I histocompatibility antigen, B-44 alpha chain
+	+		+				P30460	++++	-	HLA class I histocompatibility antigen, B-8 alpha chain
+	+		+				P30504	+++	-	HLA class I histocompatibility antigen, Cw-4 alpha chain
+	+		+				P10321	+++	-	HLA class I histocompatibility antigen, Cw-7 alpha chain
+		+	+				Q12891	++	-	Hyaluronidase-2
+	+		+				P55899	+++	-	IgG receptor FcRn large subunit p51
+	+		+				Q8TDY8	+++	-	Immunoglobulin superfamily DCC subclass member 4
	+		+				Q14573	++	-	Inositol 1,4,5-trisphosphate receptor type 3
	+		+				Q9NQX7	++	-	Integral membrane protein 2C
+	+		+				O75578	+++	++	Integrin alpha-10
+	+		+				Q13683	+++	-	Integrin alpha-7
+	+		+				P53708	+++	+++	Integrin alpha-8
+	+		+				P18084	++++	-	Integrin beta-5
+	+		+				Q9UMF0	+	-	Intercellular adhesion molecule 5
	+		+				Q01628	+++	-	Interferon-induced transmembrane protein 3
+	+		+		+		Q9NPH3	++	-	Interleukin-1 receptor accessory protein
+	+		+	+	+		Q9BX67	+++	-	Junctional adhesion molecule C
+	+		+				Q96J84	+++	-	Kin of IRRE-like protein 1
	+		+				Q86UP2	++	+++	Kinectin
+			+	+	+		Q08431	+++	++	Lactadherin
+			+	+		+	P07942	-	+++	Laminin subunit beta-1
+			+	+			P55268	-	++	Laminin subunit beta-2
+			+	+		+	P11047	-	+++	Laminin subunit gamma-1
	+		+				Q01650	++++	-	Large neutral amino acids transporter small subunit 1
+	+		+				O95490	+++	-	Latrophilin-2
+	+		+		+	+	O43155	++	-	Leucine-rich repeat transmembrane protein FLRT2
+	+		+		+	+	Q9NZU0	+++	-	Leucine-rich repeat transmembrane protein FLRT3
	+		+		+		Q9UIQ6	+++	++	Leucyl-cystinyl aminopeptidase
+		+	+				Q13449	++	-	Limbic system-associated membrane protein
+	+		+				Q86UK5	++	-	Limbin
	+		+				O14495	+++	++	Lipid phosphate phosphohydrolase 3
	+		+				P52569	++	-	Low affinity cationic amino acid transporter 2
+	+		+	+	+		P01130	+++	++	Low-density lipoprotein receptor
+	+		+				O75581	++	-	Low-density lipoprotein receptor-related protein 6
	+		+				Q92633	++++	-	Lysophosphatidic acid receptor 1
	+		+				Q14108	++	+++	Lysosome membrane protein 2
+			+	+			P61626	++	-	Lysozyme C
+	+		+	+	+		P09603	++	-	Macrophage colony-stimulating factor 1
+	+		+				Q9H0U3	++	-	Magnesium transporter protein 1
+					+		Q9UM22	+++	-	Mammalian ependymin-related protein 1
	+		+				Q96AM1	+++	-	Mas-related G-protein coupled receptor member F
+	+		+			+	P50281	++++	+++	Matrix metalloproteinase-14
+					+		Q9NR99	++	++	Matrix-remodeling-associated protein 5
	+		+				Q5HYA8	++	-	Meckelin

Table S.2 - continued: non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
+			+	+			P55145	-	++	Mesencephalic astrocyte-derived neurotrophic factor
	+		+				Q8NE01	++	-	Metal transporter CNNM3
	+		+				Q6P4Q7	++	-	Metal transporter CNNM4
+				+	+		P01033	++	-	Metalloproteinase inhibitor 1
	+		+				Q9UHE8	+++	-	Metalloreductase STEAP1
	+		+				Q8NFT2	++	-	Metalloreductase STEAP2
	+		+				Q658P3	+++	-	Metalloreductase STEAP3
+	+		+	+	+		Q29983	++	-	MHC class I polypeptide-related sequence A
	+		+				Q8TCT9	+++	+++	Minor histocompatibility antigen H13
	+		+				Q02978	+++	+++	Mitochondrial 2-oxoglutarate/malate carrier protein
	+		+				Q8IXI2	++	-	Mitochondrial Rho GTPase 1
	+		+				Q8IXI1	++	-	Mitochondrial Rho GTPase 2
	+		+				P53985	++++	++	Monocarboxylate transporter 1
	+		+				O60669	++	-	Monocarboxylate transporter 2
	+		+				O15427	++++	+++	Monocarboxylate transporter 4
	+		+				O15374	++	-	Monocarboxylate transporter 5
	+		+				P36021	+++	-	Monocarboxylate transporter 8
	+		+				P33527	++++	-	Multidrug resistance-associated protein 1
	+		+				O15439	+++	-	Multidrug resistance-associated protein 4
+	+		+				O95297	++++	-	Myelin protein zero-like protein 1
	+		+				Q9NZM1	++++	++++	Myoferlin
+				+			Q96PD5	++	-	N-acetylmuramoyl-L-alanine amidase
	+		+				Q9Y2A7	++	+++	Nck-associated protein 1
+	+		+				Q92859	++++	++	Neogenin
+		+	+				Q9Y2I2	++	-	Netrin-G1
+	+		+				O15394	++	-	Neural cell adhesion molecule 2
+	+		+	+	+		O00533	++	-	Neural cell adhesion molecule L1-like protein
+	+		+				O94856	++++	-	Neurofascin
+	+		+	+			P46531	++	-	Neurogenic locus notch homolog protein 1
+	+		+	+			Q04721	++++	++	Neurogenic locus notch homolog protein 2
+	+		+	+			Q9UM47	+++	-	Neurogenic locus notch homolog protein 3
+	+		+				Q8N2Q7	++	-	Neuroligin-1
+	+		+				Q8NFZ4	++	-	Neuroligin-2
+	+		+				Q92823	++	-	Neuronal cell adhesion molecule
+		+	+				Q7Z3B1	+++	-	Neuronal growth regulator 1
	+		+				Q13491	++	-	Neuronal membrane glycoprotein M6-b
+	+		+				O60462	++	-	Neuropilin-2
+	+		+				Q9Y639	++++	+++	Neuroplastin
+		+	+				Q9P121	+++	-	Neurotrimin
	+		+				P43007	+++	-	Neutral amino acid transporter A
	+		+				Q15758	++++	+++	Neutral amino acid transporter B(0)
+				+	+		P59665	+	-	Neutrophil defensin 1
+	+		+				Q92542	++++	+++	Nicastrin
+				+	+		Q14112	++	-	Nidogen-2
+		+	+				Q9BZM4	++	-	NKG2D ligand 3
+	+		+				Q99650	+++	-	Oncostatin-M-specific receptor subunit beta
	+		+				Q96RD7	+++	-	Pannexin-1
+				+			P26022	+	-	Pentraxin-related protein PTX3
+			+	+	+		Q92626	++	+++	Peroxidasin homolog
+			+	+			Q13162	+++	+++	Peroxiredoxin-4
	+		+				Q00325	++++	++++	Phosphate carrier protein, mitochondrial
	+		+		+		O15162	+++	-	Phospholipid scramblase 1
	+		+				Q9NRY6	-	+++	Phospholipid scramblase 3
	+		+				P20020	++++	++	Plasma membrane calcium-transporting ATPase 1
	+		+				P20020	++++	++	Plasma membrane calcium-transporting ATPase 1
	+		+				P23634	++++	++++	Plasma membrane calcium-transporting ATPase 4
+			+	+	+		P05121	+++	+++	Plasminogen activator inhibitor 1
+	+		+				Q5VY43	+++	-	Platelet endothelial aggregation receptor 1
+	+		+				Q9UIW2	++++	-	Plexin-A1
+	+		+				O75051	+++	-	Plexin-A2
+	+		+				P51805	+++	-	Plexin-A3
+	+		+		+		O43157	++	-	Plexin-B1
+	+		+				O15031	++++	-	Plexin-B2

Table S.2 - continued: non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
+	+		+				Q9Y4D7	++++	-	Plexin-D1
	+		+				Q13563	++	-	Polycystin-2
+			+	+	+		Q9UHG3	++	+++	Prenylcysteine oxidase 1
	+		+				P49768	+++	-	Presenilin-1
+				+	+		P07602	-	+++	Proactivator polypeptide
	+		+				Q14439	+++	-	Probable G-protein coupled receptor 176
	+		+				Q9C0B5	++	-	Probable palmitoyltransferase ZDHHC5
	+		+				Q9Y2G3	+++	-	Probable phospholipid-transporting ATPase IF
	+		+				Q8NB49	++++	-	Probable phospholipid-transporting ATPase IG
	+		+				P98196	++	-	Probable phospholipid-transporting ATPase IH
	+		+				Q9P241	++	-	Probable phospholipid-transporting ATPase VD
+				+	+		Q15113	++	-	Procollagen C-endopeptidase enhancer 1
	+		+				Q9HCJ1	++	-	Progressive ankylosis protein homolog
+			+		+		P12273	++	-	Prolactin-inducible protein
+				+	+		Q32P28	-	+++	Prolyl 3-hydroxylase 1
	+		+				P43116	++	-	Prostaglandin E2 receptor EP2 subtype
	+		+				Q8TCG1	+	-	Protein CIP2A
+				+			O00622	+	+++	Protein CYR61
+		+		+			P07237	+++	++++	Protein disulfide-isomerase
+				+			Q92520	++	++	Protein FAM3C
+	+		+		+		Q9ULI3	++	-	Protein HEG homolog 1
	+		+				Q86UE4	++	+++	Protein LYRIC
	+		+				Q92508	+++	++	Protein PIEZO1
	+		+				Q9C0H2	+++	++	Protein tweety homolog 3
+			+	+	+	+	P41221	++	-	Protein Wnt-5a
+			+	+	+	+	Q9H1J7	++	-	Protein Wnt-5b
+	+		+		+		P25116	+++	-	Proteinase-activated receptor 1
+				+	+	+	P28300	++	-	Protein-lysine 6-oxidase
	+		+				Q04941	+++	-	Proteolipid protein 2
+	+		+				Q6V0I7	+++	-	Protocadherin Fat 4
+	+		+				Q9Y5H3	++	-	Protocadherin gamma-A10
+	+		+				Q9Y5G0	+++	-	Protocadherin gamma-B5
+	+		+				Q9Y5F8	++	-	Protocadherin gamma-B7
+	+		+				Q9UN70	++++	-	Protocadherin gamma-C3
+	+		+				Q96JQ0	++	-	Protocadherin-16
+	+		+				Q9HCL0	+++	-	Protocadherin-18
+	+		+				Q9HC56	++	-	Protocadherin-9
+	+		+				Q68D85	+	-	Putative Ig-like domain-containing protein
	+		+				Q9BSK0	+++	-	Putative MARVEL domain-containing protein 1
+	+		+				P18433	+++	-	Receptor-type tyrosine-protein phosphatase alpha
+	+		+				P10586	+++	-	Receptor-type tyrosine-protein phosphatase F
+	+		+				P23470	++++	-	Receptor-type tyrosine-protein phosphatase gamma
+	+		+				Q15262	+++	-	Receptor-type tyrosine-protein phosphatase kappa
+	+		+				P28827	+++	-	Receptor-type tyrosine-protein phosphatase mu
+	+		+				Q13332	+++	-	Receptor-type tyrosine-protein phosphatase S
+	+		+				Q8IUW5	+++	-	RELT-like protein 1
+	+		+				O75787	++	-	Renin receptor
	+		+				Q9NQC3	++++	++++	Reticulon-4
+		+	+				O95980	+++	-	Reversion-inducing cysteine-rich protein with Kazal motifs
+		+	+				Q6NW40	+++	-	RGM domain family member B
	+		+				Q9HAB3	+++	-	Riboflavin transporter 3
+	+		+				Q9Y6N7	++++	-	Roundabout homolog 1
+	+		+				Q9HCK4	+++	-	Roundabout homolog 2
	+		+				P16615	++++	++++	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
+	+		+	+	+		Q8WVN6	++	-	Secreted and transmembrane protein 1
+			+	+	+	+	Q8N474	++	++	Secreted frizzled-related protein 1
	+		+	+	+		Q12884	++++	+++	Seprase
	+		+				Q9NRX5	++	-	Serine incorporator 1
+				+	+		Q92743	-	++	Serine protease HTRA1
+			+		+		Q15165	++	-	Serum paraoxonase/arylesterase 2
	+		+				Q99720	+++	++	Sigma non-opioid intracellular receptor 1
+	+		+				Q8TCT8	+++	-	Signal peptide peptidase-like 2A
+	+		+				Q8TCT7	++	-	Signal peptide peptidase-like 2B

Table S.2 - continued: non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
	+		+				P31641	+++	-	Sodium- and chloride-dependent taurine transporter
	+		+				Q9Y6M7	++	-	Sodium bicarbonate cotransporter 3
+	+		+				P32418	+++	-	Sodium/calcium exchanger 1
	+		+				P19634	+++	-	Sodium/hydrogen exchanger 1
	+		+				Q92581	++	-	Sodium/hydrogen exchanger 6
	+		+				P53794	++	-	Sodium/myo-inositol cotransporter
	+		+				P05023	++++	++++	Sodium/potassium-transporting ATPase subunit alpha-1
	+		+				P05026	++++	-	Sodium/potassium-transporting ATPase subunit beta-1
	+		+				Q96QD8	+++	+++	Sodium-coupled neutral amino acid transporter 2
	+		+				Q8WUX1	++	-	Sodium-coupled neutral amino acid transporter 5
	+		+				Q9Y289	+++	-	Sodium-dependent multivitamin transporter
	+		+				Q8WUM9	++	-	Sodium-dependent phosphate transporter 1
	+		+				Q08357	++	-	Sodium-dependent phosphate transporter 2
	+		+				P55011	+	-	Solute carrier family 12 member 2
	+		+				Q9UP95	+++	-	Solute carrier family 12 member 4
	+		+				Q9UHW9	++	-	Solute carrier family 12 member 6
	+		+				Q9BXP2	++	-	Solute carrier family 12 member 9
	+		+				P11166	++++	+++	Solute carrier family 2, facilitated glucose transporter member 1
	+		+				O95528	++	-	Solute carrier family 2, facilitated glucose transporter member 10
	+		+				Q96BI1	++	-	Solute carrier family 22 member 18
+	+		+				Q99523	++	-	Sortilin
+			+	+	+		P09486	++	-	SPARC
	+		+				O15121	+++	-	Sphingolipid delta(4)-desaturase DES1
	+		+				O95136	++	-	Sphingosine 1-phosphate receptor 2
	+		+				Q99500	++	-	Sphingosine 1-phosphate receptor 3
+			+	+	+		P48061	+	-	Stromal cell-derived factor 1
	+		+				P50443	++	-	Sulfate transporter
+	+		+	+			O00391	++	-	Sulfhydryl oxidase 1
+	+		+	+			Q6ZRP7	+	-	Sulfhydryl oxidase 2
	+		+				O43760	++	-	Synaptogyrin-2
	+		+				Q16563	++++	-	Synaptophysin-like protein 1
+	+		+	+			P31431	+++	-	Syndecan-4
	+		+				Q12846	+++	-	Syntaxin-4
	+		+				Q9UNK0	++	++	Syntaxin-8
+	+			+			Q8TB96	++	-	T-cell immunomodulatory protein
+			+	+	+		P24821	-	+++	Tenascin
+				+			Q9Y6I9	++	-	Testis-expressed sequence 264 protein
	+		+				O14817	++	-	Tetraspanin-4
	+		+				O75954	+++	-	Tetraspanin-9
+	+		+				P36897	++	-	TGF-beta receptor type-1
+	+		+				P37173	+++	-	TGF-beta receptor type-2
	+		+				Q9UKU6	+++	-	Thyrotropin-releasing hormone-degrading ectoenzyme
+	+		+	+	+	+	Q03167	+++	-	Transforming growth factor beta receptor type 3
+			+	+	+	+	Q15582	++	+++	Transforming growth factor-beta-induced protein ig-h3
	+		+				Q8TD43	++	-	Transient receptor potential cation channel subfamily M member 4
	+		+				Q9Y5S1	+++	++	Transient receptor potential cation channel subfamily V member 2
+	+		+				Q99805	-	++	Transmembrane 9 superfamily member 2
+	+		+				P49755	+++	+++	Transmembrane emp24 domain-containing protein 10
	+		+		+		O14668	++	-	Transmembrane gamma-carboxyglutamic acid protein 1
+	+		+				Q14956	++	-	Transmembrane glycoprotein NMB
	+		+				Q86WV6	++	++	Transmembrane protein 173
	+		+				Q9BQJ4	++	-	Transmembrane protein 47
	+		+				Q96GC9	++	-	Transmembrane protein 49
+	+		+				Q13641	++++	++	Trophoblast glycoprotein
+	+		+	+	+		P30530	+++	++	Tyrosine-protein kinase receptor UFO
+	+		+				Q01973	++	-	Tyrosine-protein kinase transmembrane receptor ROR1
+	+		+				Q13308	++++	+++	Tyrosine-protein kinase-like 7
+				+			Q969H8	-	+++	UPF0556 protein C19orf10
+	+			+			Q6EMK4	++++	+++	Vasorin
	+		+				P63027	+++	-	Vesicle-associated membrane protein 2
	+		+				Q15836	-	+++	Vesicle-associated membrane protein 3
	+		+				O95183	++	-	Vesicle-associated membrane protein 5
	+		+				Q9P0L0	+++	+++	Vesicle-associated membrane protein-associated protein A

Table S.2 - continued: non CD-marker cell surface proteins identified.

SP	TM	GA	PM	ECS	ECR	ECM	AccNo	B	T	Name
	+		+				P21796	+++	++++	Voltage-dependent anion-selective channel protein 1
+	+		+				P54289	++++	+++	Voltage-dependent calcium channel subunit alpha-2/delta-1
	+		+				Q93050	++	-	V-type proton ATPase 116 kDa subunit a isoform 1
	+		+				Q13488	++	+++	V-type proton ATPase 116 kDa subunit a isoform 3
	+		+				Q92536	+++	-	Y+L amino acid transporter 2
	+		+				Q9Y6M5	+++	-	Zinc transporter 1
+	+		+				Q15043	+++	-	Zinc transporter ZIP14
+	+		+				Q13433	++	-	Zinc transporter ZIP6
+			+	+			P25311	+	-	Zinc-alpha-2-glycoprotein

Identified proteins were accepted as cell surface proteins if they are GPI-anchored (GA) or transmembrane proteins (TM) assigned to the plasma membrane (PM) or signal peptide-containing proteins (SP) that are assigned to either the extracellular space (ECS), region (ECR) or matrix (ECM) see listing S.1. SLIK2 and LRC15 are missing because they are not assigned to either PM, ECS, ECR or ECM.

Listing S.1 : Boolean expression to create a list of and to filter for cell surface proteins

```
organism:homo sapiens AND reviewed:yes AND (( GO:0005886 AND (keyword:GPI-anchor OR
annotation:(type:transmem count:[1 TO *])) ) OR ((GO:0005615 OR GO:0005576 OR GO:0031012) AND
annotation:(type:signal peptide)))
```

4508 matches in Uniprot (Version 2011_01). GO:0005886 - plasma membrane – PM, GO:0005615 - extracellular space – ECS, GO:0005576 - extracellular region – ECR, GO:0031012 - extracellular matrix - ECM

Table S.3: Proteins recruited with gE on PI4P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
P58871	TB182	+++	+	+++	182 kDa tankyrase-1-binding protein
Q9Z1B3	PLCB1	+++	++	++++	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-1
P51432	PLCB3	++	++	++++	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-3
P70290	EM55	++	+	++++	55 kDa erythrocyte membrane protein
O54950	AAKG1	+	++	++	5-AMP-activated protein kinase subunit gamma-1
P70265	F262	++	+	+++	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2
P61211	ARL1	++	+	+++	ADP-ribosylation factor-like protein 1
Q9WUL7	ARL3	++	+	++	ADP-ribosylation factor-like protein 3
O88848	ARL6	+++	+	++++	ADP-ribosylation factor-like protein 6
Q80WT5	AFTIN	++	++	++	Aftiphilin
Q8R5F3	CF130	++	++	++	AI314976 protein
Q8K341	ATAT	++	+	+++	Alpha-tubulin N-acetyltransferase
Q7TQF7	AMPH	++	+	+++	Amphiphysin
B2RW11	B2RW11	++	+	+++	Ankyrin repeat domain 34A
Q8CC13	AP1B1	+++	++	++++	AP-1 complex subunit beta-1
P22892	AP1G1	+++	+	+++	AP-1 complex subunit gamma-1
P35585	AP1M1	+++	+	++	AP-1 complex subunit mu-1
P61967	AP1S1	++	++	+++	AP-1 complex subunit sigma-1A
Q9DB50	AP1S2	++	++	+++	AP-1 complex subunit sigma-2
P17426	AP2A1	+++	+	+++	AP-2 complex subunit alpha-1
P17427	AP2A2	+++	++	++++	AP-2 complex subunit alpha-2
P84091	AP2M1	+++	++	++++	AP-2 complex subunit mu
P62743	AP2S1	+++	++	+++	AP-2 complex subunit sigma
Q3UHJ0	AAK1	++	++	+++	AP2-associated protein kinase 1
Q8VHH5	AGAP3	++	+	+++	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3
Q9Z2A5	ATE1	++	++	+++	Arginyl-tRNA--protein transferase 1
Q9D7A8	ARMC1	++	++	+++	Armadillo repeat-containing protein 1
Q6P542	ABCF1	++	++	+++	ATP-binding cassette sub-family F member 1
Q91VR5	DDX1	+++	++	+++	ATP-dependent RNA helicase DDX1
Q62167	DDX3X	++	+	++	ATP-dependent RNA helicase DDX3X
Q9Z2H5	E41L1	+++	+	++	Band 4.1-like protein 1
P62204	CALM	++	++	++	Calmodulin
Q3UHL1	CAMKV	+++	+	+++	CaM kinase-like vesicle-associated protein
Q9Z0H8	CLIP2	++	+	+++	CAP-Gly domain-containing linker protein 2
Q9DBC3	MTR1	++	+	+++	Cap-specific mRNA (nucleoside-2-O-)-methyltransferase 1
Q8BK63	KC1A	++	++	++++	Casein kinase I isoform alpha
Q9JMK2	KC1E	++	++	++	Casein kinase I isoform epsilon
Q8BHE3	ATCAY	+++	+	++++	Caytaxin
Q5M8N0	CNRP1	++	+	+++	CB1 cannabinoid receptor-interacting protein 1
Q8BI72	CARF	++	++	+++	CDKN2A-interacting protein
Q9JKY0	RCD1	++	++	++++	Cell differentiation protein RCD1 homolog
Q6A068	CDC5L	++	++	+++	Cell division cycle 5-related protein
Q6A065	CE170	++	++	+++	Centrosomal protein of 170 kDa
Q8R1T1	CHMP7	+	+	++	Charged multivesicular body protein 7
Q61548	AP180	+++	++	++++	Clathrin coat assembly protein AP180
Q68FD5	CLH	+++	++	++++	Clathrin, heavy polypeptide (Hc)
Q9CQF3	CPSF5	++	++	+++	Cleavage and polyadenylation specificity factor subunit 5
Q8K1A6	C2D1A	++	+	+++	Coiled-coil and C2 domain-containing protein 1A
Q9EQG9	C43BP	++	++	+++	Collagen type IV alpha-3-binding protein
Q9WUM4	COR1C	++	++	+++	Coronin-1C
P45481	CBP	++	++	+++	CREB-binding protein
Q9D5V5	CUL5	+	+	++	Cullin 5
O35495	CDK14	+	++	++	Cyclin-dependent kinase 14
P97315	CSRP1	++	++	+++	Cysteine and glycine-rich protein 1
P63254	CRIP1	++	++	++	Cysteine-rich protein 1
Q9DCT8	CRIP2	+++	++	+++	Cysteine-rich protein 2
P63034	CYH2	++	+	++	Cytohesin-2
Q8K4R4	PITC1	+++	++	+++	Cytoplasmic phosphatidylinositol transfer protein 1
A2AGT5	CKAP5	+++	++	+++	Cytoskeleton-associated protein 5
Q61656	DDX5	++	++	+++	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
A2RSQ0	DEN5B	++	+	+++	DENN domain-containing protein 5B
Q9D9Z5	DDA1	++	++	+++	DET1- and DDB1-associated protein 1

Table S.3 - continued: Proteins recruited with gE on PI4P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
P32233	DRG1	++	++	+++	Developmentally-regulated GTP-binding protein 1
Q9QXB9	DRG2	++	+	+++	Developmentally-regulated GTP-binding protein 2
Q6P5E8	DGKQ	++	+++	++	Diacylglycerol kinase theta
B7ZMZ7	B7ZMZ7	++	+	+++	Dip2c protein
Q8BWT5	DIP2A	+++	++	++++	Disco-interacting protein 2 homolog A
Q3UH60	DIP2B	+++	++	+++	Disco-interacting protein 2 homolog B
Q811D0	DLG1	++	+	++	Disks large homolog 1
P07903	ERCC1	+	++	++	DNA excision repair protein ERCC-1
P52431	DPOD1	+	+	++	DNA polymerase delta catalytic subunit
P28352	APEX1	++	++	+++	DNA-(apurinic or apyrimidinic site) lyase
Q8C7M3	TRIM9	++	+	+++	E3 ubiquitin-protein ligase TRIM9
Q80TA9	EPG5	++	+	++	Ectopic P granules protein 5 homolog
P62631	EF1A2	+++	+	+++	Elongation factor 1-alpha 2
P42567	EPS15	++	++	++++	Epidermal growth factor receptor substrate 15
Q60902	EP15R	++	++	+++	Epidermal growth factor receptor substrate 15-like 1
Q8C460	ERI3	++	++	++	ERI1 exoribonuclease 3
Q6ZWX6	IF2A	++	++	+++	Eukaryotic translation initiation factor 2 subunit 1
Q99L45	IF2B	++	++	+++	Eukaryotic translation initiation factor 2 subunit 2
Q9Z0N1	IF2G	++	++	+++	Eukaryotic translation initiation factor 2 subunit 3, X-linked
Q6NZJ6	IF4G1	++	+	+++	Eukaryotic translation initiation factor 4 gamma 1
P59325	IF5	+	+	++	Eukaryotic translation initiation factor 5
Q61545	EWS	++	+++	+++	Ewing sarcoma breakpoint region 1
Q8CIB5	FERM2	++	++	+++	Fermitin family homolog 2
Q7TPM6	FSD1	++	+	++	Fibronectin type III and SPRY domain-containing protein 1
Q9ESZ8	GTF2I	++	+	++	General transcription factor II-I
P48318	DCE1	++	+	++	Glutamate decarboxylase 1
P26443	DHE3	+++	++	+++	Glutamate dehydrogenase 1, mitochondrial
Q9WV60	GSK3B	+++	+	+++	Glycogen synthase kinase-3 beta
Q8BS40	GLTD1	++	++	+++	Glycolipid transfer protein domain-containing protein 1
Q8BW56	GTDC1	+	++	+++	Glycosyltransferase-like domain-containing protein 1
O70311	NMT2	++	+	+++	Glycylpeptide N-tetradecanoyltransferase 2
Q60780	GAS7	+++	+++	++++	Growth arrest-specific protein 7
Q9JLM9	GRB14	+	++	++	Growth factor receptor-bound protein 14
O08582	GTPB1	++	++	+++	GTP-binding protein 1
P36916	GNL1	++	+	+++	Guanine nucleotide-binding protein-like 1
Q8C547	HTR5B	++	+	++	HEAT repeat-containing protein 5B
Q9R257	HEBP1	++	++	+++	Heme-binding protein 1
Q921F4	HNRLL	++	++	++++	Heterogeneous nuclear ribonucleoprotein L-like
Q8VEK3	HNRPU	++	+	++	Heterogeneous nuclear ribonucleoprotein U
P63158	HMGB1	+++	+++	+++	High mobility group protein B1
P70288	HDAC2	++	+	+++	Histone deacetylase 2
Q6GSS7	H2A2A	++	++	++	Histone H2A type 2-A
Q9Z0R4	ITSN1	+++	++	+++	Intersectin-1
Q9DBG3	AP2B1	+++	++	+++	Isoform 2 of AP-2 complex subunit beta
Q7TQH0	ATX2L	+	+	+++	Isoform 2 of Ataxin-2-like protein
Q8BKX1	BAIP2	+++	+	+++	Isoform 2 of Brain-specific angiogenesis inhibitor 1-associated protein 2
Q6ZQ08	CNOT1	++	+	+++	Isoform 2 of CCR4-NOT transcription complex subunit 1
Q61211	EIF2D	+	++	++	Isoform 2 of Eukaryotic translation initiation factor 2D
Q99M31	HSP7E	+	++	++	Isoform 2 of Heat shock 70 kDa protein 14
Q8CGP2	H2B1P	++	++	+++	Isoform 2 of Histone H2B type 1-P
Q5DU31	ICEF1	++	+	++	Isoform 2 of Interactor protein for cytohesin exchange factors 1
Q3V3V9	LR16C	++	++	++++	Isoform 2 of Leucine-rich repeat-containing protein 16C
Q5SUF2	LC7L3	++	++	++	Isoform 2 of Luc7-like protein 3
Q9DB27	MCTS1	++	++	+++	Isoform 2 of Malignant T cell-amplified sequence 1
Q9D868	PPIH	+	++	++	Isoform 2 of Peptidyl-prolyl cis-trans isomerase H
Q6ZWR4	2ABB	+	++	+++	Isoform 2 of Serine/threonine-protein phosphatase 2A 55 kDa reg.…
P53996	CNBP	++	++	+++	Isoform 3 of Cellular nucleic acid-binding protein
Q9JM52	MINK1	++	++	++	Isoform 3 of Misshapen-like kinase 1
Q80TZ3	AUXI	++	++	+++	Isoform 3 of Putative tyrosine-protein phosphatase auxilin
Q9JLB0	MPP6	++	+	+++	Isoform Alpha of MAGUK p55 subfamily member 6
P10637	TAU	+++	++++	++++	Isoform Tau-B of Microtubule-associated protein tau
Q9CR30	JOS2	++	++	+++	Josephin-2

Table S.3 - continued: Proteins recruited with gE on PI4P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
Q8K0T4	KATL1	++	++	+++	Katanin p60 ATPase-containing subunit A-like 1
F7A5U9	F7A5U9	++	++	++	Kelch-like protein 3
Q60749	KHDR1	++	++	+++	KH domain-containing, RNA-binding, signal transduction-ass. protein 1
Q9DBS5	KLC4	+	++	++	Kinesin light chain 4
Q9QXL2	KI21A	+++	++	+++	Kinesin-like protein KIF21A
P28740	KIF2A	+++	+	+++	Kinesin-like protein KIF2A
Q6TA13	Q6TA13	++	++	+++	Kinesin-related microtuble-based motor protein
O89112	LANC1	++	+	++	LanC-like protein 1
Q505F5	LRC47	++	++	++++	Leucine-rich repeat-containing protein 47
Q61792	LASP1	++	+	+++	LIM and SH3 domain protein 1
P32067	LA	+++	++	+++	Lupus La protein homolog
Q3UYG8	MACD2	+	++	++	MACRO domain-containing protein 2
Q14BB9	MA6D1	+++	++	+++	MAP6 domain-containing protein 1
A2AJI0	MA7D1	+++	++	+++	MAP7 domain-containing protein 1
A2AG50	MA7D2	+++	++	+++	MAP7 domain-containing protein 2
O54879	HMGB3	++	++	+++	MCG114640
Q9CR80	FA32A	++	++	++	MCG125076
Q9D883	U2AF1	++	++	++	MCG14259, isoform CRA_b
Q80TK0	K1107	+++	+++	+++	MCG19133, isoform CRA_b
G3X9K3	G3X9K3	+++	+	++	MCG8317
Q8K4B0	MTA1	++	++	++	Metastasis-associated protein MTA1
Q9R190	MTA2	++	+	+++	Metastasis-associated protein MTA2
Q9CQG2	MET16	++	++	++	Methyltransferase 10 domain containing
P10637	TAU	+++	++	++++	Microtubule-associated protein
E9PZ43	E9PZ43	++	++	++++	Microtubule-associated protein
G3UZJ2	G3UZJ2	++	++	+++	Microtubule-associated protein (Fragment)
Q9QYR6	MAP1A	+++	++	+++	Microtubule-associated protein 1 A
Q8C052	MAP1S	++	++	+++	Microtubule-associated protein 1S
P20357	MAP2	++++	++	+++	Microtubule-associated protein 2
P27546	MAP4	+++	++	++++	Microtubule-associated protein 4
Q7TSJ2	MAP6	++++	++	+++	Microtubule-associated protein 6
O88735	MAP7	++	++	+++	Microtubule-associated protein 7, isoform CRA_b
Q61166	MARE1	++	+	+++	Microtubule-associated protein RP/EB family member 1
Q6PER3	MARE3	+++	+	+++	Microtubule-associated protein RP/EB family member 3
Q91VR7	MLP3A	++	++	++	Microtubule-associated proteins 1A/1B light chain 3A
Q9CYG7	TOM34	+++	+	+++	Mitochondrial import receptor subunit TOM34
Q8BPM2	M4K5	++	+	++	Mitogen-activated protein kinase kinase kinase kinase 5
Q9WVA3	BUB3	++	+	+++	Mitotic checkpoint protein BUB3
O08539	BIN1	+++	++	+++	Myc box-dependent-interacting protein 1
Q80TA6	MTMRC	+	++	++	Myotubularin-related protein 12
Q99KK2	NEUA	+++	++	+++	N-acylneuraminate cytidylyltransferase
Q8BG30	NELFA	+	++	++	Negative elongation factor A
Q8C4Y3	NELFB	+	++	+++	Negative elongation factor B
Q6ZWR6	SYNE1	++	++	++	Nesprin-1
Q6ZWQ0	SYNE2	+++	++++	+++	Nesprin-2
Q8K4Q0	RPTOR	++	+	+++	Novel protein containing six WD40 domains at C-terminus
Q9JJR9	NRIP3	++	++	+++	Nrip3 protein
P97863	NFIB	++	++	++	Nuclear factor 1 B-type
O08919	NUMBL	++	+	+++	Numb-like protein
Q8CI95	OSB11	++	+	+++	Oxysterol binding protein-like 11
Q3B7Z2	OSBP1	+++	++	+++	Oxysterol-binding protein
Q91XL9	OSBL1	+++	++	+++	Oxysterol-binding protein-related protein 1
Q8R326	PSPC1	++	++	+++	Paraspeckle component 1
Q99JF8	PSIP1	++	++	+++	PC4 and SFRS1-interacting protein
Q9JMD3	PCTL	++	+++	+++	PCTP-like protein
Q9D0W5	PPIL1	+	++	++	Peptidyl-prolyl cis-trans isomerase-like 1
Q2PFD7	PSD3	+++	++	+++	PH and SEC7 domain-containing protein 3
P59644	PI5PA	++	+	++	Phosphatidylinositol 4,5-bisphosphate 5-phosphatase A
P53810	PIPNA	++	++	+++	Phosphatidylinositol transfer protein alpha isoform
Q7M6Y3	PICA	++	++	+++	Phosphatidylinositol-binding clathrin assembly protein
Q3V3Q7	PACS2	++	+	+++	Phosphofurin acidic cluster sorting protein 2
P53811	PIPNB	++	++	+++	Pitpnb protein

Table S.3 - continued: Proteins recruited with gE on PI4P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
Q922V4	PLRG1	++	++	+++	Pleiotropic regulator 1
P11103	PARP1	++	++	+++	Poly (ADP-ribose) polymerase family, member 1
O54724	PTRF	+	++	++	Polymerase I and transcript release factor
Q99KP6	PRP19	+++	++	++++	Pre-mRNA-processing factor 19
Q9D287	SPF27	++	++	+++	Pre-mRNA-splicing factor SPF27
Q3TSG4	ALKB5	++	++	++	Probable alpha-ketoglutarate-dependent dioxygenase ABH5
P54823	DDX6	+++	++	+++	Probable ATP-dependent RNA helicase DDX6
Q8BWU5	OSGEP	++	+++	++	Probable tRNA threonylcarbamoyladenosine biosynthesis protein Osgep
E9PV26	E9PV26	++	+++	++	Protein BC068157
Q6PGH1	BUD31	+	++	++	Protein BUD31 homolog
Q924A2	CIC	++	++	+++	Protein capicua homolog
Q7TNV0	DEK	++	++	+++	Protein DEK
E9PUQ8	E9PUQ8	++	+	++	Protein Dgkd
D3YXJ0	D3YXJ0	++	+	+++	Protein Dgkh
D3YWQ0	D3YWQ0	++	+	+++	Protein Dgki
E9PV14	E9PV14	++	+	+++	Protein Epb4.1l1
Q78T81	F102A	++	++	+++	Protein FAM102A
Q8BJH1	F164A	+++	+++	++++	Protein FAM164A
Q8BHZ0	FA49A	+++	+	+++	Protein FAM49A
Q80VD1	FA98B	++	++	+++	Protein FAM98B
E9PYD1	E9PYD1	++	++	+++	Protein Fam98c
E9PVA8	E9PVA8	+++	+	++	Protein Gcn1l1
P28867	KPCD	++	+	++++	Protein kinase C delta type
Q9D0R8	LSM12	++	++	+++	Protein LSM12 homolog
Q8K2F8	LS14A	++	++	+++	Protein LSM14 homolog A
Q8CGC4	LS14B	++	+	++	Protein LSM14 homolog B
Q9QZS3	NUMB	++	+	++	Protein numb homolog
Q60829	PPR1B	++	++	++	Protein phosphatase 1 regulatory subunit 1B
Q7TSC1	PRC2A	++	++	++	Protein PRRC2A
Q3TLH4	E9QKG5	++	++	+++	Protein PRRC2C
E9PZZ8	E9PZZ8	+++	+	+++	Protein Rap1gap
Q8BK67	RCC2	++	++	++++	Protein RCC2
P31725	S10A9	++	++	++	Protein S100-A9
Q8BX57	PXK	++	++	++++	PX domain-containing protein kinase-like protein
A2ALS7	A2ALS7	++	+	+++	Rap1 GTPase-activating protein (Fragment)
Q5SVL6	RPGP2	++	++	+++	Rap1 GTPase-activating protein 2
Q921Q7	RIN1	++	+	++	Ras and Rab interactor 1
Q60790	RASA3	++	+	+++	Ras GTPase-activating protein 3
Q9JKF1	IQGA1	++	++	+++	Ras GTPase-activating-like protein IQGAP1
Q9QUG9	GRP2	+	+	++	RAS guanyl-releasing protein 2
Q80U57	RIMS3	+	++	++	Regulating synaptic membrane exocytosis protein 3
Q9DC04	RGS3	++	+	++	Regulator of G-protein signaling 3
Q5SSM3	RHG44	+++	+	++++	Rho GTPase-activating protein 44
P70335	ROCK1	++	+	+++	Rho-associated protein kinase 1
P70336	ROCK2	+++	+	+++	Rho-associated protein kinase 2
Q14BI8	Q14BI8	++	++	+++	Rltpr protein
Q6NZN0	RBM26	++	++	+++	RNA-binding protein 26
P56959	FUS	++	++	++	RNA-binding protein FUS
Q9JLC8	SACS	++	+	++	Sacsin
Q9JHW4	SELB	+	++	++	Selenocysteine-specific elongation factor
Q3UQA7	SELH	++	+++	++	Selenoprotein H
P29621	SPA3C	++	++	++	Serine protease inhibitor A3C
P84104	SRSF3	++	++	+++	Serine/arginine-rich splicing factor 3
Q8QZV4	ST32C	++	++	+++	Serine/threonine-protein kinase 32C
P28028	BRAF	++	+	+++	Serine/threonine-protein kinase B-raf
Q9JLM8	DCLK1	+++	++	+++	Serine/threonine-protein kinase DCLK1
Q9CZ91	SRFB1	+	++	+++	Serum response factor-binding protein 1
Q6A026	PDS5A	++	++	+++	Sister chromatid cohesion protein PDS5 homolog A
Q4VA53	PDS5B	++	+++	+++	Sister chromatid cohesion protein PDS5 homolog B
Q9CSN1	SNW1	++	++	+++	SNW domain-containing protein 1
Q6NZD2	Q6NZD2	+++	++	+++	Sorting nexin 1
Q91ZR2	SNX18	++	++	+++	Sorting nexin 18

Table S.3 - continued: Proteins recruited with gE on PI4P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
Q9CWK8	SNX2	+++	+++	++++	Sorting nexin-2
Q3UHD6	SNX27	++	+	+++	Sorting nexin-27
Q8CE50	SNX30	++	++	+++	Sorting nexin-30
Q91YJ2	SNX4	+++	+	+++	Sorting nexin-4
Q9D8U8	SNX5	+++	+++	+++	Sorting nexin-5
Q6P8X1	SNX6	+++	+++	+++	Sorting nexin-6
Q9QYY8	SPAST	++	++	+++	Spastin
Q8VIJ6	SFPQ	++	++	+++	Splicing factor, proline- and glutamine-rich
Q91VZ6	SMAP1	++	++	+++	Stromal membrane-associated protein 1
Q6P5D8	SMHD1	++	++	+++	Structural maintenance of chromosomes flexible hinge domain-con….
Q8BGT7	SPF30	+	++	++	Survival of motor neuron-related-splicing factor 30
Q6PDG5	SMRC2	+	+	+++	SWI/SNF complex subunit SMARCC2
Q8BJL0	SMAL1	++	+	+++	SWI/SNF-related matrix-associated actin-dependent regulator of chroma….
O88935	SYN1	+++	++	+++	Synapsin-1
Q64332	SYN2	+++	+	+++	Synapsin-2
Q8JZP2	SYN3	++	+	+++	Synapsin-3
Q80TB8	VAT1L	+++	++	++++	Synaptic vesicle membrane protein VAT-1 homolog-like
Q8CHC4	SYNJ1	+++	++	++++	Synaptojanin-1
Q8R3T5	STXB6	++	+	+++	Syntaxin binding protein 6 (Amisyn)
P62340	TBPL1	++	+++	+++	TATA box-binding protein-like protein 1
Q6PCN3	TTBK1	++	++	+++	Tau-tubulin kinase 1
Q8BYJ6	TBCD4	++	+	+++	TBC1 domain family member 4
Q5SSZ5	TENS3	++	++	++++	Tensin-3
Q8VD63	TSYL4	++	++	+++	Testis-specific Y-encoded-like protein 4
Q8BYY4	TT39B	++	++	++	Tetratricopeptide repeat protein 39B
P97770	THUM3	+	++	++	THUMP domain-containing protein 3
Q9ES56	TPPC4	++	+	++	Trafficking protein particle complex subunit 4
Q8CGF7	TCRG1	++	+	+++	Transcription elongation regulator 1
Q9ESN6	TRIM2	++	++	+++	Tripartite motif-containing protein 2
Q9R1R2	TRIM3	++	++	+++	Tripartite motif-containing protein 3
Q99LF4	RTCB	+++	++	++++	tRNA-splicing ligase RtcB homolog
Q7TQD2	TPPP	+++	+++	+++	Tubulin polymerization-promoting protein
D3Z230	D3Z230	++	+	++	Type I inositol 3,4-bisphosphate 4-phosphatase
Q91WQ3	SYYC	++	++	++++	Tyrosyl-tRNA synthetase
Q922U1	PRPF3	++	++	+++	U4/U6 small nuclear ribonucleoprotein Prp3
Q8CCF0	PRP31	++	++	++++	U4/U6 small nuclear ribonucleoprotein Prp31
Q9DAW6	PRP4	++	++	+++	U4/U6 small nuclear ribonucleoprotein Prp4
Q3TIX9	SNUT2	++	+++	+++	U4/U6.U5 tri-snRNP-associated protein 2
F8VPZ3	F8VPZ3	++	++	++	Ubiquitin carboxyl-terminal hydrolase
Q99PL6	UBXN6	++	+	+++	UBX domain-containing protein 6
E9QB05	E9QB05	++	++	+++	Uncharacterized protein
E9QMW6	E9QMW6	++	++	++	Uncharacterized protein
Q9D735	CS043	++	++	++	Uncharacterized protein C19orf43 homolog
Q9CQE8	CN166	++	++	++	UPF0568 protein C14orf166 homolog
Q9CWN7	CB029	+	+	++	UPF0760 protein C2orf29 homolog
A2AT37	A2AT37	++	++	+++	UPF2 regulator of nonsense transcripts homolog (Yeast)
Q3ULL6	Q3ULL6	++	++	++	UPF3 regulator of nonsense transcripts homolog B
Q8VDJ3	VIGLN	++	++	++++	Vigilin
Q3UMB9	WASH7	+	++	++	WASH complex subunit 7
Q8C2E7	STRUM	+	++	++	WASH complex subunit strumpellin
Q8BH57	WDR48	++	+	++	WD repeat-containing protein 48
Q3TIV5	ZC3HF	++	++	++	Zinc finger CCCH domain-containing protein 15

Abundance - averaged log$_2$(LFQ)-values of the gE-triplicates, classified as below 18 and between 18 to 23, 23 to 28, 28 to 33 and above 33; Ratio - log$_2$-t-test differences, classified as below 1 and between 1 to 3, 3 to 6, 6 to 9 and above 9; Significance – -log$_{10}$(p-value), classified as below 1 and between 1 to 2, 2 to 4, 4 to 6 and above 6; all classes were accordingly labeled as: -, +, ++, +++, ++++. The t-test threshold value was hereby set to 1.2% and the slope value to 0.2.

Table S.4: Proteins recruited with Crb on PI3P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
Q9Z1B3	PLCB1	+++	+	++	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-1
Q99J27	ACATN	++	+	++	Acetyl-coenzyme A transporter 1
Q91VR5	DDX1	++	+	+++	ATP-dependent RNA helicase DDX1
Q9WV92	E41L3	++	-	++	Band 4.1-like protein 3
Q8CCN5	BCAS3	+++	+	+++	Breast carcinoma-amplified sequence 3 homolog
Q9WTR5	CAD13	+	+	++	Cadherin-13
Q8BGD5	CPT1C	++	+	++	Carnitine O-palmitoyltransferase 1, brain isoform
Q6IRU5	CLCB	++	++	++	Clathrin light chain B
B1AWD9	B1AWD9	++	++	++	Clathrin light polypeptide (Lca)
Q68FD5	CLH	+++	++	+++	Clathrin, heavy polypeptide (Hc)
Q8BXA5	CLP1L	++	+	++	Cleft lip and palate transmembrane protein 1-like protein
O55029	COPB2	++	+	++	Coatomer subunit beta
Q9QZE5	COPG	++	+	++	Coatomer subunit gamma
Q9QXK3	COPG2	+	+	++	Coatomer subunit gamma-2
Q9D024	CC47	+	+	++	Coiled-coil domain-containing protein 47
Q80YA8	CRUM2	++++	+++	++	Crumbs homolog 2
Q811D0	DLG1	++	+	+++	Disks large homolog 1
Q9DC23	DJC10	++	+	++	DnaJ homolog subfamily C member 10
Q3TDQ1	STT3B	++	+	++	Dolichyl-diphosphooligosaccharide—protein glycosyltransferase sub. STT3B
Q8CCJ3	UFL1	++	+	++	E3 UFM1-protein ligase 1
Q8BL66	EEA1	+++	++	+++	Early endosome antigen 1
Q60900	ELAV3	+++	++++	+++	ELAV-like protein 3
Q8VHI3	OFUT2	+	+	++	GDP-fucose protein O-fucosyltransferase 2
O08795	GLU2B	++	+	++	Glucosidase 2 subunit beta (Isoform 2)
Q99JX3	GORS2	++	++	+++	Golgi reassembly-stacking protein 2
Q3UUQ7	PGAP1	++	+	++	GPI inositol-deacylase
P07901	HS90A	++	-	++	Heat shock protein HSP 90-alpha
P11499	HS90B	+++	-	++	Heat shock protein HSP 90-beta
Q9QUJ7	ACSL4	++	+	++	Long-chain-fatty-acid--CoA ligase 4
Q99PU5	ACBG1	++	+	++	Long-chain-fatty-acid--CoA ligase ACSBG1
Q9JLB2	MPP5	++	++	++	MAGUK p55 subfamily member 5
Q80UM7	MOGS	++	+	++	Mannosyl-oligosaccharide glucosidase
G3X922	G3X922	+++	+	++	MCG115602
A2A8R0	A2A8R0	++	+	++	MCG17975
Q9QYR6	MAP1A	++	+	++	Microtubule-associated protein 1 A
P35282	RAB21	+++	+	++	MKIAA0118 protein (Fragment)
Q62074	KPCI	++	++	++	MKIAA4165 protein (Fragment)
Q9Z2L6	MINP1	+	+	+	Multiple inositol polyphosphate phosphatase 1
Q8VBX6	MPDZ	++	++	++	Multiple PDZ domain protein
Q6ZWR6	SYNE1	++	+++	++	Nesprin-1
Q62443	NPTX1	+	+	++	Neuronal pentraxin-1
Q9DBG5	PLIN3	+	+	++	Perilipin-3
Q9Z2G6	SE1L1	++	+	++	Protein sel-1 homolog 1
P53994	RAB2A	+++	+	++	Ras-related protein Rab-2A
P35278	RAB5C	+++	+	++	Ras-related protein Rab-5C
E9Q401	E9Q401	+++	+	++	Ryanodine receptor 2
Q3URD3	SLMAP	++	+	++	Sarcolemmal membrane-associated protein
Q9R118	HTRA1	+	++	+++	Serine protease HTRA1
Q9JLM8	DCLK1	++	+	++	Serine/threonine-protein kinase DCLK1
Q5F2E8	TAOK1	++	+	++	Serine/threonine-protein kinase TAO1
Q6NZD2	Q6NZD2	+++	+	++	Sorting nexin 1
O70493	SNX12	+++	++	+++	Sorting nexin 12
O70492	SNX3	+++	++	+++	Sorting nexin 3
Q91WE1	SNX15	++	++	++	Sorting nexin-15
Q8C080	SNX16	++	++	+++	Sorting nexin-16
Q8BVL3	SNX17	+	+	++	Sorting nexin-17
Q9CWK8	SNX2	+++	-	++	Sorting nexin-2
Q3UHD6	SNX27	+++	+++	++	Sorting nexin-27
Q8CE50	SNX30	+++	+	+++	Sorting nexin-30
Q9D8U8	SNX5	++	+	++	Sorting nexin-5
Q8CFD4	SNX8	++	++	++	Sorting nexin-8
P58404	STRN4	++	+	++	Striatin-4

Table S.4 - continued: Proteins recruited with Crb on PI3P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
P70302	STIM1	++	+	++	Stromal interaction molecule 1
Q9Z1F9	SAE2	++	+	++	SUMO-activating enzyme subunit 2
Q9JIS5	SV2A	++	+	++	Synaptic vesicle glycoprotein 2A
Q99KW9	TIP	++	+	++	T-cell immunomodulatory protein
Q80W04	TMCC2	++	+	++	Transmembrane and coiled-coil domains protein 2
G3X9Y5	G3X9Y5	+	+	++	Ubiquitination factor E4A, UFD2 homolog (S. cerevisiae)
F6R441	F6R441	+	+	++	Uncharacterized protein (Fragment)
Q3UE31	K0930	+	++	++	Uncharacterized protein KIAA0930 homolog
Q922R1	CP070	++	+	++	UPF0183 protein C16orf70 homolog
P40336	VP26A	++	+++	++++	Vacuolar protein sorting-associated protein 26A
Q8C0E2	VP26B	+++	+++	+++	Vacuolar protein sorting-associated protein 26B
Q9QZ88	VPS29	++	+++	+++	Vacuolar protein sorting-associated protein 29
Q9EQH3	VPS35	+++	++++	+++	Vacuolar protein sorting-associated protein 35
P97390	VPS45	++	++	++	Vacuolar protein sorting-associated protein 45
Q60930	VDAC2	+++	+	++	Voltage-dependent anion-selective channel protein 2
Q3UMB9	WASH7	+	+	++	WASH complex subunit 7
Q6PGL7	FAM21	++	+++	++	WASH complex subunit FAM21
Q8R3E3	WIPI1	++	+++	+++	WD repeat domain phosphoinositide-interacting protein 1
Q80W47	WIPI2	+++	++	+++	WD repeat domain phosphoinositide-interacting protein 2
Q9CR39	WIPI3	+++	++	+++	WD repeat domain phosphoinositide-interacting protein 3

Abundance - averaged log2(LFQ)-values of the Crb-triplicates, classified as below 18 and between 18 to 23, 23 to 28, 28 to 33 and above 33; Ratio - log2-t-test differences, classified as below 1 and between 1 to 3, 3 to 6, 6 to 9 and above 9; Significance – -log10(p-value), classified as below 1 and between 1 to 2, 2 to 4, 4 to 6 and above 6; all classes were accordingly labeled as: -, +, ++, +++, ++++. The t-test threshold value was hereby set to 1.2% and the slope value to 0.2.

Table S.5: Sorting nexins recruited on bare PI3P- with respect to PI4P-containing liposomes.

AccNo	AccStr	Abundance	Ratio	Significance	Name
Q6NZD2	Q6NZD2	+++	+++	+++	Sorting nexin-1
Q9CWK8	SNX2	+++	+++	+++	Sorting nexin-2
O70492	SNX3	++	++	++	Sorting nexin-3
Q91YJ2	SNX4	+++	+	++	Sorting nexin-4
Q9D8U8	SNX5	++	++	++	Sorting nexin-5
Q6P8X1	SNX6	+++	+++	+++	Sorting nexin-6
Q9CY18	SNX7	++	+++	++	Sorting nexin-7
Q91VH2	SNX9	++	++	+	Sorting nexin-9
O70493	SNX12	++	++	++	Sorting nexin-12
Q6PHS6	SNX13	++	+	+	Sorting nexin-13
Q91ZR2	SNX18	++	+	+	Sorting nexin-18
Q3ZT31	SNX25	++	++	+	Sorting nexin-25
Q8CE50	SNX30	+++	+++	++	Sorting nexin-30

Abundance - averaged log2(LFQ)-values of the Cys3-triplicates, classified as below 18 and between 18 to 23, 23 to 28, 28 to 33 and above 33; Ratio - log2-t-test differences, classified as below 1 and between 1 to 3, 3 to 6, 6 to 9 and above 9; Significance – -log10(p-value), classified as below 1 and between 1 to 2, 2 to 4, 4 to 6 and above 6; all classes were accordingly labeled as: -, +, ++, +++, ++++. The t-test threshold value was hereby set to 1.2% and the slope value to 0.2.

Table S.6: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q9Z1T6	FYV1			+	+	+	1-phosphatidylinositol-3-phosphate 5-kinase
Q8VDM4	PSMD2					+	26S proteasome non-ATPase regulatory subunit 2
Q9Z2A0	PDPK1					+	3-phosphoinositide-dependent protein kinase 1
P14131	RS16	++				-	40S ribosomal protein S16
Q9JLJ2	AL9A1				+	+	4-trimethylaminobutyraldehyde dehydrogenase
Q9JM14	NT5C					+	5(3)-deoxyribonucleotidase, cytosolic type
Q5EG47	AAPK1	+				+	5-AMP-activated protein kinase catalytic subunit alpha-1
Q6PAM0	AAKB2	+					5-AMP-activated protein kinase subunit beta-2
O08848	RO60			+			60 kDa SS-A/Ro ribonucleoprotein
Q9WUA3	K6PP				+	+	6-phosphofructokinase type C
P12382	K6PL				+	+	6-phosphofructokinase, liver type
P47857	K6PF				+	+	6-phosphofructokinase, muscle type
Q9CQ60	6PGL					+	6-phosphogluconolactonase
P97819	PA2G6	+			+	+	85 kDa calcium-independent phospholipase A2
Q4JIM5	ABL2			+	+	+	Abelson tyrosine-protein kinase 2
Q8VD66	ABHD4	+					Abhydrolase domain-containing protein 4
Q9D2R0	AACS			+	+	+	Acetoacetyl-CoA synthetase
Q8QZT1	THIL			+			Acetyl-CoA acetyltransferase, mitochondrial
P68033	ACTC				+++	-	Actin, alpha cardiac muscle 1
Q99MR0	ACL6B				+	+	Actin-like protein 6B
Q6P4T2	Q6P4T2			++	-		Activating signal cointegrator 1 complex subunit 3-like 1
Q9WTL7	LYPA2			++		++	Acyl-protein thioesterase 2
Q9CR95	NECP1	++				+	Adaptin ear-binding coat-associated protein 1
A2BDX3	MOCS3			+		+	Adenylyltransferase and sulfurtransferase MOCS3
P48962	ADT1	+					ADP/ATP translocase 1
P61205	ARF3			-	+		ADP-ribosylation factor 3
Q6P5E6	GGA2			+	+		ADP-ribosylation factor-binding protein GGA2
P61211	ARL1					+	ADP-ribosylation factor-like protein 1
Q9QZQ1	AFAD			+			Afadin (Protein Af-6)
Q8CG76	ARK72					+	Aflatoxin B1 aldehyde reductase member 2
Q80WT5	AFTIN			+	+		Aftiphilin
Q3THG9	AASD1					+	Alanyl-tRNA editing protein Aarsd1
Q9JII6	AK1A1					+	Alcohol dehydrogenase [NADP+]
P28474	ADHX					+	Alcohol dehydrogenase class-3
P45376	ALDR					+	Aldose reductase
P29699	FETUA			+			Alpha-2-HS-glycoprotein
Q9DBF1	AL7A1					++	Alpha-aminoadipic semialdehyde dehydrogenase
P61164	ACTZ			+			Alpha-centractin
Q8BGW1	FTO			+			Alpha-ketoglutarate-dependent dioxygenase FTO
Q9D3D0	TTPAL			+			Alpha-tocopherol transfer protein-like
Q8K341	ATAT				+		Alpha-tubulin N-acetyltransferase
Q3UH49	Q3UH49	+++	++++	+++	++++	++++	Amyloid beta (A4) protein binding, family A
P98084	APBA2	+++	+++	+++	+++	+++	Amyloid beta A4 precursor protein-binding family A member 2
Q9QXJ1	APBB1	+++	++	++	++	++	Amyloid beta A4 precursor protein-binding family B member 1
Q3TPE9	ANKY2					+	Ankyrin repeat and MYND domain-containing protein 2
P59672	ANS1A			+	+	+	Ankyrin repeat and SAM domain containing 1
Q91ZU1	ASB6			++	+		Ankyrin repeat and SOCS box protein 6
Q8BIZ1	ANS1B				+	+	Ankyrin repeat and sterile alpha motif domain….
B2RW11	B2RW11				+		Ankyrin repeat domain 34A
P97384	ANX11	+			-	-	Annexin A11
P97429	ANXA4	+			-	-	Annexin A4
O88512	AP1G2					+	AP-1 complex subunit gamma-like 2
Q3UHJ0	AAK1				+	+	AP2-associated protein kinase 1
B0V3P2	B0V3P2	++		++	++	++	AP-4 complex subunit beta-1
Q80V94	AP4E1	++		++	++	++	AP-4 complex subunit epsilon-1
Q9JKC7	AP4M1	+		++	++	++	AP-4 complex subunit mu-1
Q9WVL1	AP4S1	++		+	+	++	AP-4 complex subunit sigma-1
Q3TAP4	YK046			++	+	++	AP-5 complex subunit beta-1
Q8BPD5	APOA1	+					Apolipoprotein A-I
P33622	APOC3	++					Apolipoprotein C-III (Apo-CIII) (ApoC-III) (Apolipoprotein C3)
P51910	APOD	+				-	Apolipoprotein D

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q62415	ASPP1			+	+	+	Apoptosis-stimulating of p53 protein 1
Q80WC7	AGFG2				+	+	Arf-GAP domain and FG repeat-containing protein 2
Q8K2K6	AGFG1	++		+		-	Arf-GAP domain and FG repeats-containing protein 1
Q7SIG6	ASAP2			+			Arf-GAP with SH3 domain, ANK repeat and PH ….
P16460	ASSY				+	+	Argininosuccinate synthase
Q8BNU0	ARMC6			+			Armadillo repeat-containing protein 6
Q61024	ASNS			+	+	+	Asparagine synthetase [glutamine-hydrolyzing]
O70305	ATX2				++	+	Ataxin-2 (Spinocerebellar ataxia type 2 protein homolog)
Q91V92	ACLY			+			ATP citrate lyase
O54984	ASNA			+			ATPase Asna1
P61222	ABCE1				+	+	ATP-binding cassette sub-family E member 1
Q91VR5	DDX1					+	ATP-dependent RNA helicase DDX1
Q62167	DDX3X				+	+	ATP-dependent RNA helicase DDX3X
Q9D8Z6	ATGA1	+					Autophagy-related protein 101
Q9CQC6	BZW1			+			Basic leucine zipper and W2 domain-containing protein 1
Q61335	BAP31			-	+	-	B-cell receptor-associated protein 31
O88597	BECN1					+	Beclin-1
Q8BHT6	B3GLT			+			Beta-1,3-glucosyltransferase
Q9CWL8	CTBL1			+			Beta-catenin-like protein 1
Q8R5C5	ACTY			+			Beta-centractin
Q9DBL7	COASY					+	Bifunctional coenzyme A synthase
Q8C3R1	BRAT1					+	BRCA1-associated ATM activator 1
Q6PAJ1	BCR				+	+	Breakpoint cluster region protein
Q8CCN5	BCAS3			+		+	Breast carcinoma-amplified sequence 3 homolog
D3Z6Q9	D3Z6Q9			-		+	Bridging integrator 2
Q8K2Q7	BROX			+			BRO1 domain-containing protein BROX
Q8K1X1	WDR11	++		+	++	++	Bromodomain and WD repeat domain containing 2
O54825	BYST			+			Bystin
Q922D8	C1TC			+	+	+	C-1-tetrahydrofolate synthase, cytoplasmic
Q5HZI2	C2C4C					++	C2 calcium-dependent domain-containing protein 4C
P12658	CALB1			+			Calbindin
B1AUH5	B1AUH5	++		+	+		Calcium/calmodulin-dependent serine protein kinase
Q9CXW3	CYBP			+			Calcyclin-binding protein
Q80VC9	CAMP3			+	+	++	Calmodulin-regulated spectrin-associated protein 3
P14211	CALR			+			Calreticulin
Q9DBC3	MTR1			+		+	Cap-specific mRNA (nucleoside-2-O-)-methyltransferase 1
Q8K354	Q8K354					+	Carbonyl reductase 3
Q00493	CBPE					+	Carboxypeptidase E
Q9DC28	KC1D					++	Casein kinase 1, delta, isoform CRA_b
Q9JMK2	KC1E					++	Casein kinase I isoform epsilon
B9EJ23	B9EJ23	++		+++	+++	+++	Cask protein
Q8BHE3	ATCAY			-	-	+	Caytaxin
Q5M8N0	CNRP1					+	CB1 cannabinoid receptor-interacting protein 1
Q8BH15	CNOTA				++	+	CCR4-NOT transcription complex subunit 10
Q8C5L3	CNOT2				+		CCR4-NOT transcription complex subunit 2
Q8K0V4	CNOT3				+	+	CCR4-NOT transcription complex subunit 3
Q60809	CNOT7				+		CCR4-NOT transcription complex subunit 7
Q9JKC6	CEND	-			+		Cell cycle exit and neuronal differentiation protein 1
Q9JKY0	RCD1				+		Cell differentiation protein RCD1 homolog
Q6ZQK5	ACAP2				+	+	Centaurin beta 2
Q9D8B3	CHM4B			+			Charged multivesicular body protein 4b
P49025	CTRO					+	Citron Rho-interacting kinase (CRIK)
Q61548	AP180					+	Clathrin coat assembly protein AP180
Q9CQF3	CPSF5				+		Cleavage and polyadenylation specificity factor subunit 5
Q8BTV2	CPSF7				++	+	Cleavage and polyadenylation specificity factor subunit 7
Q8BIQ5	CSTF2			+	++	++	Cleavage stimulation factor subunit 2
Q99LI7	CSTF3	++		++	+++	++	Cleavage stimulation factor subunit 3
Q8K1A6	C2D1A	+		+	++	++	Coiled-coil and C2 domain-containing protein 1A
Q9JIG7	CCD22	++		++	+++	++	Coiled-coil domain-containing protein 22
Q5SP85	CC85A	+					Coiled-coil domain-containing protein 85A
Q7TQK5	CCD93			++	++	++	Coiled-coil domain-containing protein 93
Q91W50	CSDE1			+		+	Cold shock domain-containing protein E1

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q9EQG9	C43BP					+	Collagen type IV alpha-3-binding protein
Q8K4M5	COMD1				+	++	COMM domain containing 1
Q8JZY2	COMDA			+	++	++	COMM domain-containing protein 10
Q8BXC6	COMD2			+	++	++	COMM domain-containing protein 2
Q63829	COMD3	+		+	++	++	COMM domain-containing protein 3
Q9CQ02	COMD4	++		++	++	++	COMM domain-containing protein 4
Q8R395	COMD5			++	++	+++	COMM domain-containing protein 5
Q3V4B5	COMD6				+		COMM domain-containing protein 6
Q8BG94	COMD7	+		++	++	++	COMM domain-containing protein 7
Q9CZG3	COMD8			+	+	++	COMM domain-containing protein 8
Q8K2Q0	COMD9	++		++	++	++	COMM domain-containing protein 9
P84086	CPLX2					++	Complexin-2
Q921L5	COG2			+			Conserved oligomeric Golgi complex subunit 2
Q8CI04	COG3			+		+	Conserved oligomeric Golgi complex subunit 3
Q8C0L8	COG5			+	+	+	Conserved oligomeric Golgi complex subunit 5
Q3UM29	COG7			+	+		Conserved oligomeric Golgi complex subunit 7
Q9JJA2	COG8			+			Conserved oligomeric Golgi complex subunit 8
P12960	CNTN1			+			Contactin-1
Q8C166	CPNE1	+			+	+	Copine-1
O89053	COR1A				+	+	Coronin-1A
Q9D2V7	CORO7			+			Coronin-7
P30275	KCRU			++	++	+	Creatine kinase U-type, mitochondrial
P45481	CBP			++	++	+++	CREB-binding protein (EC 2.3.1.48)
Q6P1J1	Q6P1J1					+	Crmp1 protein
P70698	PYRG1			+	+		CTP synthase 1
Q9D4H8	CUL2				+		Cullin-2
Q3TCH7	CUL4A	+			+		Cullin-4A
Q6ZQ38	CAND1				+		Cullin-associated NEDD8-dissociated protein 1
Q6ZQ73	CAND2			+			Cullin-associated NEDD8-dissociated protein 2
O35495	CDK14					+	Cyclin-dependent kinase 14
Q04899	CDK18					+	Cyclin-dependent kinase 18
Q99KY4	GAK	++				++	Cyclin-G-associated kinase
P62897	CYC			+			Cytochrome c, somatic
Q8K0T2	DC2L1			++			Cytoplasmic dynein 2 light intermediate chain 1
Q9CPY7	AMPL				+	+	Cytosol aminopeptidase
Q8R0Y6	AL1L1					+	Cytosolic 10-formyltetrahydrofolate dehydrogenase
Q61753	SERA				+	+	D-3-phosphoglycerate dehydrogenase
Q61655	DD19A					+	DEAD (Asp-Glu-Ala-Asp) box polypeptide 19b
Q61656	DDX5					++	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5
A2AF47	DOC11			+	++	+	Dedicator of cytokinesis protein 11
Q8CIQ7	DOCK3			+	+	++	Dedicator of cytokinesis protein 3
P59764	DOCK4	+		++	++	++	Dedicator of cytokinesis protein 4
Q8R1A4	DOCK7			+	+	+	Dedicator of cytokinesis protein 7
Q8BIK4	DOCK9				+		Dedicator of cytokinesis protein 9
E9Q5T9	E9Q5T9			++	++		Deleted.
Q71KT5	ERG24			+			Delta(14)-sterol reductase
Q8CHT0	AL4A1			+			Delta-1-pyrroline-5-carboxylate dehydrogenase
A2RT67	DEND3			+			DENN domain-containing protein 3
Q6PAL8	DEN5A			+	+	++	DENN domain-containing protein 5A
A2RSQ0	DEN5B					+	DENN domain-containing protein 5B
Q9QXB9	DRG2					+	Developmentally-regulated GTP-binding protein 2
Q9EQF6	DPYL5					+	Dihydropyrimidinase-related protein 5
P50285	FMO1			+			Dimethylaniline monooxygenase [N-oxide-forming] 1
Q99JF5	MVD1			++	+		Diphosphomevalonate decarboxylase
Q3UHC7	DAB2P					+	Disabled homolog 2-interacting protein
Q8BWT5	DIP2A					++	Disco-interacting protein 2 homolog A
Q3UH60	DIP2B					++	Disco-interacting protein 2 homolog B
P70175	DLG3					+	Disks large homolog 3
Q3U1J4	DDB1	+	++	+	++	++	DNA damage-binding protein 1
P54276	MSH6				+		DNA mismatch repair protein Msh6
P54797	T10					+	DNA segment, Chr 16, human D22S680E, expressed
O35134	RPA1			+			DNA-directed RNA polymerase I subunit RPA1
Q923G2	RPAB3			+			DNA-directed RNA polymerases I, II, and III subunit RPABC3

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q9QYI3	DNJC7			+			DnaJ homolog subfamily C member 7
P46978	STT3A			+			Dolichyl-diphosphooligosaccharide-protein glycosyltransf….
O35075	DSCR3	+		++	++	+++	Down syndrome critical region protein 3 homolog
Q62418	DBNL				+		Drebrin-like protein
P47809	MP2K4			+	+	+	Dual specificity mitogen-activated protein kinase kinase 4
Q9WVS7	MP2K5				+		Dual specificity mitogen-activated protein kinase kinase 5
P56387	DYLT3	++			-		Dynein light chain Tctex-type 3
Q91ZU6	DYST			-	+		Dystonin
Q9D2N4	DTNA				+		Dystrobrevin alpha, isoform CRA_a
Q8CHI8	EP400				+	+	E1A-binding protein p400
Q5DTM8	BRE1A			++	++	++	E3 ubiquitin-protein ligase BRE1A
Q3U319	BRE1B			++	++	++	E3 ubiquitin-protein ligase BRE1B
Q6A009	LTN1			+			E3 ubiquitin-protein ligase listerin
P46935	NEDD4					+	E3 ubiquitin-protein ligase NEDD4
Q9JI90	RNF14				+		E3 ubiquitin-protein ligase RNF14
Q69ZI1	SH3R1					+	E3 ubiquitin-protein ligase SH3RF1
Q8CH72	TRI32	++		+	+	++	E3 ubiquitin-protein ligase TRIM32
Q99LC5	ETFA			+		+	Electron transfer flavoprotein subunit alpha, mitochondrial
P58252	EF2					+	Elongation factor 2
Q8C0D5	ETUD1			+		+	Elongation factor Tu GTP-binding domain-containing protein 1
Q7TT37	ELP1					+	Elongator complex protein 1
Q91WG4	ELP2			+		+	Elongator complex protein 2
Q9ER73	ELP4				+		Elongator complex protein 4
Q3UJB9	EDC4				+		Enhancer of mRNA decapping 4, isoform CRA_b
Q8K2D3	EDC3				+		Enhancer of mRNA-decapping protein 3
P42567	EPS15	+			+	+	Epidermal growth factor receptor substrate 15
Q60902	EP15R	+					Epidermal growth factor receptor substrate 15-like 1
Q8VEH5	EPMIP					+	EPM2A-interacting protein 1
Q8CHU3	EPN2	++			-		Epsin-2
Q91VC3	IF4A3	+				+	Eukaryotic initiation factor 4A-III
Q149F3	ERF3B				+		Eukaryotic peptide chain release factor GTP-binding ….
Q91WK2	EIF3H			++		++	Eukaryotic translation initiation factor 3 subunit H
Q99JX4	EIF3M			+			Eukaryotic translation initiation factor 3 subunit M
Q80XI3	IF4G3			+		+	Eukaryotic translation initiation factor 4 gamma 3
Q62448	IF4G2			+	+	++	Eukaryotic translation initiation factor 4, gamma 2
Q9WUK2	IF4H			+			Eukaryotic translation initiation factor 4H
Q61545	EWS				+++	+	Ewing sarcoma breakpoint region 1
Q8R3S6	EXOC1			+			Exocyst complex component 1
Q8R313	EXOC6			+	+	+	Exocyst complex component 6 (SEC15-like protein 1)
Q924C1	XPO5			+	+	+	Exportin-5
Q924Z6	XPO6			+	+		Exportin-6
Q9EPK7	XPO7					+	Exportin-7
Q9CRT8	XPOT			+			Exportin-T
Q61553	FSCN1				+		Fascin
P19096	FAS			+	+		Fatty acid synthase
Q8VDH1	FBX21			+			F-box only protein 21
Q91W61	FXL15			+			F-box/LRR-repeat protein 15
Q9QXW2	FBXW5	+			++	++	F-box/WD repeat-containing protein 5
Q9CRA9	FGOP2				+		FGFR1 oncogene partner 2 homolog
Q8BYN5	FSD1L					+	FSD1-like protein
Q8VEB1	GRK5					+	G protein-coupled receptor kinase 5
Q9R0N0	GALK1					+	Galactokinase
P17183	ENOG			+			Gamma-enolase
Q99JP7	GGT7			+			Gamma-glutamyltransferase 7
P58854	GCP3			+	++	+	Gamma-tubulin complex component 3
Q8K0C9	GMDS			+	+	+	GDP-mannose 4,6 dehydratase
Q9JHK4	PGTA			+		+	Geranylgeranyl transferase type-2 subunit alpha
O55126	NIPS2	+++			+	+	Glioblastoma amplified sequence
E9PYV1	E9PYV1					+	Glucocorticoid receptor
P47856	GFPT1					+	Glucosamine--fructose-6-phosphate aminotransferase ….
Q9Z2Z9	GFPT2				++++	++++	Glucosamine--fructose-6-phosphate aminotransferase ….
Q00612	G6PD1				+	+	Glucose-6-phosphate 1-dehydrogenase X
D3Z7P3	D3Z7P3			+		+	Glutaminase kidney isoform, mitochondrial

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
P15105	GLNA					+	Glutamine synthetase
O70325	GPX41					+	Glutathione peroxidase
Q8QZY2	GLCTK			+			Glycerate kinase
Q64516	GLPK	-				+	Glycerol kinase
Q3ULJ0	GPD1L				+	+	Glycerol-3-phosphate dehydrogenase 1-like protein
Q8C0L9	GPCP1				+	+	Glycerophosphocholine phosphodiesterase GPCPD1
Q9Z1E4	GYS1			++	++	++	Glycogen [starch] synthase, muscle
Q8CI94	PYGB					+	Glycogen phosphorylase, brain form
Q2NL51	GSK3A				+	+	Glycogen synthase kinase-3 alpha
Q9CZD3	SYG					+	Glycyl-tRNA synthetase
Q9DCZ1	GMPR1			++		+	GMP reductase 1
Q3THK7	GUAA			+	+	+	GMP synthase [glutamine-hydrolyzing]
Q8BMP6	GCP60	++				+	Golgi resident protein GCP60
Q9CW79	GOGA1					+	Golgin subfamily A member 1
Q8BUV3	GEPH			+	+	+	Gphn protein
Q9JJQ0	PIGB			+		++	GPI mannosyltransferase 3
Q9D4H2	GCC1			++		++	GRIP and coiled-coil domain-containing protein 1
Q9JLM9	GRB14	++		++	++		Growth factor receptor-bound protein 14
Q9WUR9	KAD4			-		++	GTP:AMP phosphotransferase AK4, mitochondrial
P08556	RASN	+				+	GTPase NRas
Q921J2	RHEB					+	GTP-binding protein Rheb
Q9CQC9	SAR1B	+					GTP-binding protein SAR1b
E9PXE2	E9PXE2			+			Guanine nucleotide exchange factor DBS
P27601	GNA13			+			Guanine nucleotide-binding protein subunit alpha-13
Q9D114	MESH1					+	Guanosine-3,5-bis(diphosphate) 3-pyrophosphohydrolase ….
Q9EP53	TSC1			++	++	++	Hamartin
Q8C547	HTR5B			+	+		HEAT repeat-containing protein 5B
Q6P1G0	HEAT6			+			HEAT repeat-containing protein 6
P48722	HS74L			+			Heat shock 70 kDa protein 4L
O35737	HNRH1					+	Heterogeneous nuclear ribonucleoprotein H1
P70333	HNRH2			++			Heterogeneous nuclear ribonucleoprotein H2
Q9D0E1	HNRPM	+		++	++	++	Heterogeneous nuclear ribonucleoprotein M
Q8VDM6	HNRL1			+	+	+	Heterogeneous nuclear ribonucleoprotein U-like protein 1
Q9CPS6	HINT3	+					Histidine triad nucleotide-binding protein 3
Q61035	SYHC				+	+	Histidyl-tRNA synthetase, cytoplasmic
B2RWS6	EP300			+	++	++	Histone acetyltransferase p300
P43274	H14			-		+	Histone H1.4
Q9WVG6	CARM1				+		Histone-arginine methyltransferase CARM1
Q9Z2G9	HTAI2				-	+	HIV-1 tat interactive protein 2, homolog (Human)
Q9Z2Y3	HOME1				+	+	Homer protein homolog 1
P38060	HMGCL				+	+	Hydroxymethylglutaryl-CoA lyase, mitochondrial
Q5D0E0	Q5D0E0					+	Ikbkb protein
O35344	IMA3	+					Importin subunit alpha-3
P70168	IMB1			+		+	Importin subunit beta-1
Q8K0C1	IPO13				++		Importin-13
Q91YE6	IPO9			+		+	Importin-9
P24547	IMDH2				++	++	Inosine-5-monophosphate dehydrogenase 2
Q8BYN3	ITPK1					+	Inositol-tetrakisphosphate 1-kinase
Q6P4S8	INT1			+			Integrator complex subunit 1
Q7TPD0	INT3			+	+		Integrator complex subunit 3
O55222	ILK				+		Integrin-linked protein kinase
Q8R3Y8	I2BP1					+	Interferon regulatory factor 2-binding protein 1
Q8K3X4	I2BPL		++		+	+	Interferon regulatory factor 2-binding protein-like
Q6NWV3	IF122				++		Intraflagellar transport protein 122 homolog
Q8R0S2	IQEC1	++		+	++	+	IQ motif and SEC7 domain-containing protein 1
P85094	ISC2A			+		+	Isochorismatase domain-containing protein 2A, mitochondrial
Q04690	NF1			+	+	++	Isoform 1 of Neurofibromin
Q80TQ2	CYLD			+	+	+	Isoform 1 of Ubiquitin carboxyl-terminal hydrolase CYLD
Q5SWU9	ACACA			+	+	+	Isoform 2 of Acetyl-CoA carboxylase 1
Q7TQH0	ATX2L			+	+	++	Isoform 2 of Ataxin-2-like protein
Q45VK7	DYHC2			+	+	+	Isoform 2 of Cytoplasmic dynein 2 heavy chain 1
Q99N11	DUS22					++	Isoform 2 of Dual specificity protein phosphatase 22
O08788	DCTN1			+			Isoform 2 of Dynactin subunit 1

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
P58281	OPA1					+	Isoform 2 of Dynamin-like 120 kDa protein, mitochondrial
Q8BHL5	ELMO2					+	Isoform 2 of Engulfment and cell motility protein 2
Q8R123	FAD1				+	+	Isoform 2 of FAD synthase
Q99M31	HSP7E			+		+	Isoform 2 of Heat shock 70 kDa protein 14
P61979	HNRPK					+	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K
Q6PAR0	KLD10	++				+	Isoform 2 of Kelch domain-containing protein 10
Q9DB27	MCTS1					+	Isoform 2 of Malignant T cell-amplified sequence 1
Q8VHN8	SDOS					+	Isoform 2 of Protein syndesmos
Q6ZQ29	TAOK2	+		+	++	++	Isoform 2 of Serine/threonine-protein kinase TAO2
Q8CI15	SPHK1				-	+	Isoform 2 of Sphingosine kinase 1
Q6P2B1	TNPO3				+		Isoform 2 of Transportin-3
Q8CGY8	OGT1			+	+	+	Isoform 2 of UDP-N-acetylglucosamine-peptide N-acety...
P40336	VP26A	++		+	+	+	Isoform 2 of Vacuolar protein sorting-associated protein 26A
Q9JM52	MINK1				+	+	Isoform 2 of Misshapen-like kinase 1
Q9WU78	PDC6I				+	+	Isoform 3 of Programmed cell death 6-interacting protein
Q6PDS3	SARM1			+	+	++	Isoform 3 of Sterile alpha and TIR motif-containing protein 1
Q6ZQ08	CNOT1				+	+	Isoform 4 of CCR4-NOT transcription complex subunit 1
Q80YX1	TENA			+			Isoform 4 of Tenascin
Q80X50	UBP2L				+		Isoform 5 of Ubiquitin-associated protein 2-like
P12023	A4		+++	+++	++	++	Isoform APP695 of Amyloid beta A4 protein
Q99PP7	TRI33			+		+	Isoform Beta of E3 ubiquitin-protein ligase TRIM33
Q5DTN8	JKIP3			++	++		Janus kinase and microtubule-interacting protein 3
Q9CR30	JOS2			-	-	+	Josephin-2
Q02257	PLAK				+		Junction plakoglobin
Q8K0T4	KATL1					+	Katanin p60 ATPase-containing subunit A-like 1
Q8BG40	KTNB1					+	Katanin p80 WD40-containing subunit B1
Q99M74	KRT82	-			+++		Keratin, type II cuticular Hb2
Q60749	KHDR1			+	+	++	KH domain-containing, RNA-binding, signal transduction-ass….
Q9CPU0	LGUL			+			Lactoylglutathione lyase
Q80U56	AVL9	+		-	-	-	Late secretory pathway protein AVL9 homolog
Q6ZPK7	LST2	+					Lateral signaling target protein 2 homolog
Q80TR1	LPHN1				+		Latrophilin-1
Q3UGP9	LRC58					+	Leucine-rich repeat-containing protein 58
Q8BMJ2	SYLC			+	+	+	Leucine--tRNA ligase, cytoplasmic
Q9ESE1	LRBA					++	Lipopolysaccharide-responsive and beige-like anchor protein
P41216	ACSL1			+			Long-chain-fatty-acid--CoA ligase 1
Q99PU5	ACBG1					+	Long-chain-fatty-acid--CoA ligase ACSBG1
Q8C142	ARH	++	++	+	+		Low density lipoprotein receptor adapter protein 1
Q9D358	PPAC					+	Low molecular weight phosphotyrosine protein phosphatase
Q91X52	DCXR			+		+	L-xylulose reductase
Q9WUC3	LY6H				+		Lymphocyte antigen 6H (Ly-6H)
P70699	LYAG			++		+	Lysosomal alpha-glucosidase
Q9WV34	MPP2				+	+	MAGUK p55 subfamily member 2
Q9JLB0	MPP6				+	+	MAGUK p55 subfamily member 6
Q14BB9	MA6D1					+	MAP6 domain-containing protein 1
A2AG50	MA7D2					+	MAP7 domain-containing protein 2
O89029	MATN4			+			Matrilin-4 (MAT-4)
Q8C525	M21D2	+++			+	++	MCG125361, isoform CRA_a
E0CYQ0	E0CYQ0					+	MCG13350, isoform CRA_b
G5E895	G5E895					+	MCG142264, isoform CRA_b
Q9D5R2	WDR20	++		+	+	+	MCG14935, isoform CRA_a
A2A8R0	A2A8R0			+		+	MCG17975
Q6ZQG1	Q6ZQG1			+			MCG21756, isoform CRA_a (Fragment)
Q925J9	MED1				+	+	Mediator of RNA polymerase II transcription subunit 1
A2ABV5	MED14	+		++	+++	++	Mediator of RNA polymerase II transcription subunit 14
Q924H2	MED15				-	+	Mediator of RNA polymerase II transcription subunit 15
Q8VCD5	MED17				++	++	Mediator of RNA polymerase II transcription subunit 17
Q9CZ82	MED18			+			Mediator of RNA polymerase II transcription subunit 18
Q80YQ2	MED23				++	+	Mediator of RNA polymerase II transcription subunit 23
Q8VCB2	MED25			+		+	Mediator of RNA polymerase II transcription subunit 25
P41242	MATK					+	Megakaryocyte-associated tyrosine-protein kinase

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
P70290	EM55				+	+	Membrane protein, palmitoylated
Q3UHE1	PITM3			+			Membrane-associated phosphatidylinositol transfer protein 3
Q99LB6	MAT2B					+	Methionine adenosyltransferase 2 subunit beta
Q9R008	KIME			+	+	+	Mevalonate kinase
Q7TSJ2	MAP6					+	Microtubule-associated protein 6
Q9R1L5	MAST1					+	Microtubule-associated serine/threonine-protein kinase 1
Q59J78	MIMIT				+	-	Mimitin, mitochondrial
Q9CYG7	TOM34				+	+	Mitochondrial import receptor subunit TOM34
Q9D6Y7	MSRA	+				+	Mitochondrial peptide methionine sulfoxide reductase
Q8VCM5	MUL1	++			+		Mitochondrial ubiquitin ligase activator of NFKB 1
P63085	MK01				+		Mitogen-activated protein kinase 1
Q91Y86	MK08				+		Mitogen-activated protein kinase 8
Q99JP0	M4K3				+	+	Mitogen-activated protein kinase kinase kinase 3
P35282	RAB21				+		MKIAA0118 protein (Fragment)
Q99P88	NU155	+		+	+	+	MKIAA0791 protein (Fragment)
O35382	EXOC4			+			MKIAA1699 protein (Fragment)
Q91WG7	DGKG					+	MKIAA4131 protein (Fragment)
Q9D071	MMS19				+	+	MMS19 nucleotide excision repair protein homolog
Q8C570	RAE1L				+	+	mRNA export factor
Q9Z2C9	MTMR7					+	Myotubularin-related protein 7
Q68FH4	GALK2			+		+	N-acetylgalactosamine kinase
Q8VDQ8	SIRT2	+		+	+	+	NAD-dependent deacetylase sirtuin-2
Q6PHQ8	NAA35			+		+	N-alpha-acetyltransferase 35, NatC auxiliary subunit
Q8BG30	NELFA			+	+	+	Negative elongation factor A
Q8C4Y3	NELFB			+	+		Negative elongation factor B
Q6ZWR6	SYNE1	++	++	++	++	+++	Nesprin-1
P13595	NCAM1			+			Neural cell adhesion molecule 1
Q9EPN1	NBEA					++	Neurobeachin
P97300	NPTN			+			Neuroplastin
Q8R151	ZNFX1			+	++	+	NFX1-type zinc finger-containing protein 1
Q8CC86	PNCB				+	+	Nicotinate phosphoribosyltransferase
Q9Z0J4	NOS1			+	+	++	Nitric oxide synthase, brain
Q8K2T1	NMRL1					+	NmrA-like family domain-containing protein 1
Q99K48	NONO	+			+	+	Non-POU domain-containing octamer-binding protein
A2ANC6	A2ANC6	++				+	Novel protein
Q8K4Q0	RPTOR	+		++	++	++	Novel protein containing six WD40 domains at C-terminus
Q8R0G9	NU133	+		+			Nuclear pore complex protein Nup133
Q9Z0W3	NU160			+			Nuclear pore complex protein Nup160
Q80U93	NU214			+		+	Nuclear pore complex protein Nup214
Q05BA5	Q05BA5					+	Nuclear receptor coactivator 3
Q78ZA7	NP1L4				+		Nucleosome assembly protein 1-like 4
O08919	NUMBL	++	+	+	+	+	Numb-like protein
Q6NVF0	Q6NVF0			++	+	+++	Oculocerebrorenal syndrome of Lowe
Q99PG2	OGFR			+			Opioid growth factor receptor
Q8R326	PSPC1	++		+	++	+	Paraspeckle component 1
Q8BFQ8	PDDC1				++	+	Parkinson disease 7 domain-containing protein 1
Q64378	FKBP5				+	+	Peptidyl-prolyl cis-trans isomerase FKBP5
Q9DBG5	PLIN3	+		-	-	-	Perilipin-3
O70589	CSKP	++		+++	+++	+++	Peripheral plasma membrane protein CASK
O08807	PRDX4			+	+	+	Peroxiredoxin-4
P98191	CDS1			+			Phosphatidate cytidylyltransferase 1
Q6PF93	PK3C3			+	+	+	Phosphatidylinositol 3-kinase catalytic subunit type 3
O08908	P85B			+			Phosphatidylinositol 3-kinase regulatory subunit beta
Q8BTI9	PK3CB				+		Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic ….
Q7M6Y3	PICA	++					Phosphatidylinositol-binding clathrin assembly protein
Q8VD65	PI3R4			+	+	+	Phosphoinositide 3-kinase regulatory subunit 4
Q91UZ1	Q91UZ1			-		+	Phospholipase C beta 4
Q9D1G2	PMVK					+	Phosphomevalonate kinase
P59114	PCIF1					+	Phosphorylated CTD-interacting factor 1
P20918	PLMN				++		Plasminogen
Q8BUL6	PKHA1					+	Pleckstrin homology domain-containing family A member 1
P11103	PARP1			+	+	+	Poly (ADP-ribose) polymerase family, member 1
Q8BZ20	PAR12				+		Poly [ADP-ribose] polymerase 12

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q91WF7	FIG4	+		++	++	+	Polyphosphoinositide phosphatase
P54823	DDX6				+	+	Probable ATP-dependent RNA helicase DDX6
Q6DIC0	SMCA2			+			Probable global transcription activator SNF2L2
P12815	PDCD6					+	Programmed cell death protein 6
Q9DB60	PGFS					+	Prostamide/prostaglandin F synthase
Q6PDI5	ECM29			+			Proteasome-associated protein ECM29 homolog
D3YVV3	D3YVV3			-	+		Protein Ankrd29
O88737	BSN			+	++	++	Protein bassoon
Q3V3N7	Q3V3N7	+			+	+	Protein Bbs1
E9PV26	E9PV26					+	Protein BC068157
Q8C7D2	CRBN	++	++	++	++	++	Protein cereblon
Q80U30	CL16A					+	Protein CLEC16A
D3YXJ0	D3YXJ0					+	Protein Dgkh
Q8BHZ0	FA49A	+					Protein FAM49A
Q9D7I8	FA83D	+	+++			++	Protein FAM83D
Q3UVG3	F91A1	++		+	++	++	Protein FAM91A1
Q80VD1	FA98B			+		+	Protein FAM98B
Q8BIL5	HOOK1					+	Protein Hook homolog 1
Q80U49	K0284				+	+	Protein KIAA0284
P23298	KPCL				+		Protein kinase C eta type
Q02956	KPCZ				+		Protein kinase C zeta type
Q8JZS0	LIN7A	+				+	Protein lin-7 homolog A
O88951	LIN7B	++		+	++	+	Protein lin-7 homolog B
O88952	LIN7C					+	Protein lin-7 homolog C
Q8K2F8	LS14A					+	Protein LSM14 homolog A
O55125	NIPS1	++			+	+	Protein NipSnap homolog 1 (NipSnap1)
D3YVV7	D3YVV7					+	Protein Nova2
Q9QZS3	NUMB	++	++	+	++	+	Protein numb homolog
Q3TDD9	PPR21				+	+	Protein phosphatase 1 regulatory subunit 21
Q60676	PPP5					+	Protein phosphatase 5, catalytic subunit
Q8BVQ5	PPME1				+		Protein phosphatase methylesterase 1
Q52KR3	PRUN2			-		++	Protein prune homolog 2
Q9D394	RUFY3			-	+		Protein RUFY3
Q6DID3	SCAF8					+	Protein SCAF8 (RNA-binding motif protein 16) (SR-related ….
Q8C8N2	SCAI			+		+	Protein SCAI
Q9D1M0	SEC13				+		Protein SEC13 homolog
Q9DB90	SMG9			++			Protein SMG9
Q01405	SC23A				+		Protein transport protein Sec23A
Q3UPL0	SC31A	+		+++	+++	+++	Protein transport protein Sec31A
O70274	TP4A2			-	+		Protein tyrosine phosphatase type IVA 2
Q99KD5	UN45A			+		++	Protein unc-45 homolog A
Q80WQ2	VAC14	+		+	++	+	Protein VAC14 homolog
B1ART2	B1ART2					+	Protein Vps13d
Q80ZJ6	ZER1				+		Protein zer-1 homolog
Q3UFS0	ZY11B			+			Protein zyg-11 homolog B
P59913	PCMD1				+		Protein-L-isoaspartate O-methyltransferase domain-….
Q9QVP9	FAK2			+	+	+	PTK2 protein tyrosine kinase 2 beta, isoform CRA_a
Q80SW1	SAHH2	+		+	+	+	Putative adenosylhomocysteinase 2
Q99K01	PDXD1	+					Pyridoxal-dependent decarboxylase domain-….
Q05920	PYC			+			Pyruvate carboxylase, mitochondrial
P52480	KPYM					+	Pyruvate kinase isozymes M1/M2
P47199	QOR					+	Quinone oxidoreductase
A2AWA9	RBGP1	++	++	+	+	++	Rab GTPase-activating protein 1
A6H6A9	RBG1L	+				+	Rab GTPase-activating protein 1-like
Q921Q7	RIN1				+	+	Ras and Rab interactor 1
Q60790	RASA3					+	Ras GTPase-activating protein 3
Q9QUG9	GRP2				+	+	RAS guanyl-releasing protein 2
Q9DD03	RAB13			+++			Ras-related protein Rab-13
Q9D1G1	RAB1B				+		Ras-related protein Rab-1B
Q99KL7	RAB28					+	Ras-related protein Rab-28
P62823	RAB3C				+		Ras-related protein Rab-3C
P35279	RAB6A				+		Ras-related protein Rab-6A
P61294	RAB6B				+		Ras-related protein Rab-6B

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
P27671	RGRF1			+	+	+	Ras-specific guanine nucleotide-releasing factor 1
Q9ESK9	RBCC1					+	RB1-inducible coiled-coil protein 1
Q8VE37	RCC1			+			Rcc1 protein
Q9CSU0	RPR1B				+	++	Regulation of nuclear pre-mRNA domain-containing ….
Q6NXI6	RPRD2			++	++	++	Regulation of nuclear pre-mRNA domain-containing protein 2
Q9EPU0	RENT1			+		+	Regulator of nonsense transcripts 1
Q8VEE4	RFA1			+	+	+	Replication protein A1
Q8BP92	RCN2			+			Reticulocalbin-2
P24549	AL1A1				+	++	Retinal dehydrogenase 1
Q9Z275	RLBP1				+	+	Retinaldehyde-binding protein 1
O54829	RGS7					+	Rgs7 protein
Q60875	ARHG2			+	+	+	Rho guanine nucleotide exchange factor 2
Q8BYP3	RHOF	+					Rho-related GTP-binding protein RhoF
Q8CFV9	RIFK					+	Riboflavin kinase
F6Q8A4	F6Q8A4				+	+	Ribosomal protein S6 kinase
Q8C050	KS6A5					+	Ribosomal protein S6 kinase alpha-5
Q80TE0	RPAP1			+		+	RNA polymerase II-associated protein 1
Q8R4X3	RBM12			+	+	++	RNA-binding protein 12
Q6NZN0	RBM26					+	RNA-binding protein 26
Q6P5B0	RRP12			+	-		RRP12-like protein
Q8BIJ7	RUFY1				+		RUN and FYVE domain-containing protein 1
P60122	RUVB1			++	+		RuvB-like 1
Q9WTM5	RUVB2	+		++	+		RuvB-like 2
Q9JLC8	SACS			++	++	++	Sacsin
Q922D4	PP6R3					+	SAPS domain family, member 3, isoform CRA_c
P36536	SAR1A	+					SAR1 gene homolog A (S. cerevisiae)
Q8CFE4	SCYL2					+	SCY1-like protein 2
Q8BTY8	SCFD2			+			Sec1 family domain-containing protein 2
A2A9B9	A2A9B9					+	SEC14-like 1 (S. cerevisiae)
Q80ZX0	Q80ZX0			+		+	Sec24 related gene family, member B (S. cerevisiae)
Q9DBC0	SELO			++	+	+	Selenoprotein O
Q9QUR8	SEM7A			++			Semaphorin-7A
Q64105	SPRE				+	+	Sepiapterin reductase
Q9QZX7	SRR					+	Serine racemase
Q9Z2W1	STK25					+	Serine/threonine-protein kinase 25
Q8QZV4	ST32C					+	Serine/threonine-protein kinase 32C
Q7TSE6	ST38L			+		+	Serine/threonine-protein kinase 38-like
Q9JI11	STK4			++			Serine/threonine-protein kinase 4
Q9JKK8	ATR			++	++		Serine/threonine-protein kinase ATR
P28028	BRAF				+	+	Serine/threonine-protein kinase B-raf
E9QMP6	E9QMP6					+	Serine/threonine-protein kinase MARK2
Q99JT2	MST4					+	Serine/threonine-protein kinase MST4
Q9JLN9	MTOR	+		+	+	+	Serine/threonine-protein kinase mTOR
Q8BTW9	PAK4	+			+		Serine/threonine-protein kinase PAK 4
E9PU87	E9PU87				+	+	Serine/threonine-protein kinase SIK3
Q8BKX6	SMG1	+		++	++	++	Serine/threonine-protein kinase SMG1
Q3U3Q1	ULK3			++	+	++	Serine/threonine-protein kinase ULK3
P58389	PTPA			+	+		Serine/threonine-protein phosphatase 2A activator
Q62388	ATM			++			Serine-protein kinase ATM
Q62086	PON2	-		+			Serum paraoxonase/arylesterase 2
Q9CWR2	SMYD3					+	SET and MYND domain-containing protein 3
Q9R0P3	ESTD					+	S-formylglutathione hydrolase
P14576	SRP54			-	++		Signal recognition particle 54 kDa protein
P42225	STAT1					+	Signal transducer and activator of transcription 1
P42227	STAT3			+	+	++	Signal transducer and activator of transcription 3
Q9D032	SSBP3			+			Single-stranded DNA-binding protein 3
Q6A026	PDS5A			+	+	++	Sister chromatid cohesion protein PDS5 homolog A
P63163	RSMN					+	Small nuclear ribonucleoprotein-associated protein N
Q02384	SOS2				+	+	Son of sevenless homolog 2
Q91ZR2	SNX18			-		+	Sorting nexin 18
O70492	SNX3	++				+	Sorting nexin 3
Q9WV80	SNX1			-		++	Sorting nexin-1

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q8BVL3	SNX17	+++		+	+++	+++	Sorting nexin-17
Q9CWK8	SNX2			-	-	+++	Sorting nexin-2
Q3UHD6	SNX27	++				++	Sorting nexin-27
Q8CE50	SNX30				-	+	Sorting nexin-30
Q80ZJ7	SNX32					++	Sorting nexin-32
Q9D8U8	SNX5			-		+++	Sorting nexin-5
Q6P8X1	SNX6			-		++	Sorting nexin-6
Q3UHA3	SPTCS	+		+	+	+	Spatacsin
Q9JIA7	SPHK2					+	Sphingosine kinase 2
Q9Z1N5	DX39B			-		+	Spliceosome RNA helicase Ddx39b
Q64213	SF01				+	++	Splicing factor 1
Q99NB9	SF3B1			+			Splicing factor 3b, subunit 1
Q8VIJ6	SFPQ	++		+	++	+	Splicing factor, proline- and glutamine-rich
Q78PY7	SND1			+	+	+	Staphylococcal nuclease domain-containing protein 1
Q6P5D8	SMHD1			+	+	+	Structural maintenance of chromosomes flexible hinge ....
Q9Z1F9	SAE2					+	SUMO-activating enzyme subunit 2
Q6PDG5	SMRC2			+	+	+	SWI/SNF complex subunit SMARCC2
Q8BJL0	SMAL1			+		+	SWI/SNF-related matrix-associated actin-dependent ....
Q80X82	SYMPK			+	++	+	Symplekin
O88935	SYN1	++			+	+	Synapsin-1
Q64332	SYN2	+			+	+	Synapsin-2
Q8JZP2	SYN3	+			+		Synapsin-3
Q8R570	SNP47					+	Synaptosomal-associated protein 47
Q61234	SNTA1			+			Syntrophin acidic 1
P26039	TLN1	-		+			Talin-1
Q71LX4	TLN2			+			Talin-2
Q9DCJ1	LST8	+			+	+	Target of rapamycin complex subunit LST8
Q3UUG6	TBC24			+		+	TBC1 domain family member 24
Q8BYJ6	TBCD4	+					TBC1 domain family member 4
Q80XQ2	TBCD5	++		+	+	+	TBC1 domain family member 5
Q9D0K0	TBCD7					+	TBC1 domain family member 7
Q9Z1A9	TBCD8					+	TBC1 domain family member 8
Q5SVR0	TBC9B					+	TBC1 domain family member 9B
P80314	TCPB				+	+	T-complex protein 1 subunit beta
P80315	TCPD				+	+	T-complex protein 1 subunit delta
P80316	TCPE				+	+	T-complex protein 1 subunit epsilon
P80313	TCPH					+	T-complex protein 1 subunit eta
Q80VP0	TCPR1			+			Tectonin beta-propeller repeat-containing protein 1
Q9DC40	TELO2			+		+	Telomere length regulation protein TEL2 homolog
Q5SSZ5	TENS3			+		+	Tensin-3
Q8BYY4	TT39B					++	Tetratricopeptide repeat protein 39B
Q810A3	TTC9C				+		Tetratricopeptide repeat protein 9C
Q60803	TRAF3	+		+	+		TNF receptor-associated factor 3
Q5NCF2	TPPC1	+					Trafficking protein particle complex subunit 1
Q3TLI0	TPC10			+	+	+	Trafficking protein particle complex subunit 10
Q8K2L8	TPC12			++	+	++	Trafficking protein particle complex subunit 12
Q9CQP2	TPPC2	+					Trafficking protein particle complex subunit 2
Q9JME7	TPC2L	+		+		+	Trafficking protein particle complex subunit 2-like protein
O55013	TPPC3	+		+	+	+	Trafficking protein particle complex subunit 3
Q9ES56	TPPC4	+		+	+	+	Trafficking protein particle complex subunit 4
Q9D289	TPC6B	+					Trafficking protein particle complex subunit 6B
Q3U0M1	TPPC9			+	+	+	Trafficking protein particle complex subunit 9
Q3TKT4	SMCA4			+	+	+	Transcription activator BRG1
O55201	SPT5H				+		Transcription elongation factor SPT5
Q8CGF7	TCRG1				+		Transcription elongation regulator 1
Q3UR70	TGFA1				+	++	Transforming growth factor-beta receptor-associated 1
Q4VBD2	TAPT1			+			Transmembrane anterior posterior transformation protein 1
Q3UBX0	TM109			++			Transmembrane protein 109
Q99LG2	TNPO2			+	+	+	Transportin-2
Q64737	PUR2			+	+	+	Trifunctional purine biosynthetic protein adenosine-3
Q505D9	TRI67			++	++	++	Tripartite motif-containing protein 67
Q91XI1	DUS3L				+		tRNA-dihydrouridine synthase 3-like
Q99LF4	RTCB				+	++	tRNA-splicing ligase RtcB homolog

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
P32921	SYWC				+	+	Tryptophanyl-tRNA synthetase, cytoplasmic
Q61037	TSC2			+	+	++	Tuberin (Tuberous sclerosis 2 protein homolog)
P68373	TBA1C				+		Tubulin alpha-1C chain
P05214	TBA3			+		+	Tubulin alpha-3 chain
P68368	TBA4A				+		Tubulin alpha-4A chain
Q9JJZ2	TBA8				+		Tubulin alpha-8 chain
Q7TMM9	TBB2A				+		Tubulin beta-2A chain
Q9ERD7	TBB3				+		Tubulin beta-3 chain
P68372	TBB4B				+		Tubulin beta-4B chain
P99024	TBB5				+		Tubulin beta-5 chain
P83887	TBG1	+		++	++	++	Tubulin gamma-1 chain
Q921G8	GCP2			+	++	++	Tubulin, gamma complex associated protein 2, isoform CRA_f
Q8CIV8	TBCE			+	+	+	Tubulin-specific chaperone E
Q62393	TPD52	++		-	-	-	Tumor protein D52 (mD52)
Q9CYZ2	TPD54	+		-	-	-	Tumor protein D52-like 2
Q61187	TS101				+	+	Tumor susceptibility gene 101 protein
D3Z230	D3Z230			+	+	+	Type I inositol 3,4-bisphosphate 4-phosphatase
E9PVM1	E9PVM1				+		Type II inositol 3,4-bisphosphate 4-phosphatase
P41241	CSK			+		+	Tyrosine-protein kinase CSK
Q6PB44	PTN23	+		+	+		Tyrosine-protein phosphatase non-receptor type 23
Q80U87	UBP8				+		Ubiquitin carboxyl-terminal hydrolase
Q99K46	UBP11				+		Ubiquitin carboxyl-terminal hydrolase 11
Q9D9M2	UBP12	++					Ubiquitin carboxyl-terminal hydrolase 12
Q8R5H1	UBP15				+	++	Ubiquitin carboxyl-terminal hydrolase 15
Q5DU02	UBP22			+		+	Ubiquitin carboxyl-terminal hydrolase 22
P35123	UBP4			+		+	Ubiquitin carboxyl-terminal hydrolase 4
P56399	UBP5			+			Ubiquitin carboxyl-terminal hydrolase 5
Q9ES00	UBE4B			+			Ubiquitin conjugation factor E4 B
G3X9Y5	G3X9Y5			+			Ubiquitination factor E4A, UFD2 homolog (S. cerevisiae)
Q9D906	ATG7				+	+	Ubiquitin-like modifier-activating enzyme ATG7
Q9ES34	UBE3B			+			Ubiquitin-protein ligase E3B
Q80U95	UBE3C					+	Ubiquitin-protein ligase E3C
Q8VCH8	UBXN4					+	UBX domain-containing protein 4
Q8R059	GALE					+	UDP-glucose 4-epimerase
O70475	UGDH			+			UDP-glucose 6-dehydrogenase
Q8VC42	MIC1			+	+	+	Uncharacterized protein C18orf8 homolog
Q80TK0	K1107					+	Uncharacterized protein KIAA1107
Q9JK81	MYG1			+		+	UPF0160 protein MYG1, mitochondrial
Q8BWQ6	CP062	++		++	+++	+++	UPF0505 protein C16orf62 homolog
Q9CPT4	CS010			+		+	UPF0556 protein C19orf10 homolog
Q9CWN7	CB029	+			+	+	UPF0760 protein C2orf29 homolog
P13439	UMPS				+	+	Uridine 5-monophosphate synthase
P52623	UCK1					+	Uridine kinase
Q91YL3	UCKL1				+	+	Uridine-cytidine kinase-like 1
Q6PDG8	MON1A			+		+	Vacuolar fusion protein MON1 homolog A
Q91W86	VPS11	+		+	+	+	Vacuolar protein sorting-associated protein 11 homolog
Q8BX70	VP13C	+					Vacuolar protein sorting-associated protein 13C
Q920Q4	VPS16				+	+	Vacuolar protein sorting-associated protein 16 homolog
Q8R307	VPS18			+		+	Vacuolar protein sorting-associated protein 18 homolog
Q8C0E2	VP26B	++			+	+	Vacuolar protein sorting-associated protein 26B
Q9QZ88	VPS29	++		+	++	++	Vacuolar protein sorting-associated protein 29
Q9D2N9	VP33A				+	+	Vacuolar protein sorting-associated protein 33A
P59016	VP33B	+			-		Vacuolar protein sorting-associated protein 33B
Q9EQH3	VPS35	++		+	+	+	Vacuolar protein sorting-associated protein 35
Q8VEJ9	VPS4A				++		Vacuolar protein sorting-associated protein 4A
P46467	VPS4B				++	+	Vacuolar protein sorting-associated protein 4B
Q5SPW0	VPS54			+			Vacuolar protein sorting-associated protein 54
Q0P5W1	VPS8			+	+	+	Vacuolar protein sorting-associated protein 8 homolog
Q9CQ80	VPS25					+	Vacuolar protein-sorting-associated protein 25
Q9Z1Q9	SYVC			+			Valyl-tRNA synthetase
Q8R5L3	VPS39		+		+	+	Vam6/Vps39-like protein
P50544	ACADV	++		+	+		Very long-chain specific acyl-CoA dehydrogenase ….
P46460	NSF				+	+	Vesicle-fusing ATPase

Table S.6 – continued: Proteins recruited with APP on different liposomes.

AccNo	AccStr	Ratio3	Ratio4	Ratio35	Ratio45	Ratio345	Name
Q8VDJ3	VIGLN			+	+	+	Vigilin
P62482	KCAB2	+		+	+	+	Voltage-gated potassium channel subunit beta-2
Q9Z1G4	VPP1			+			V-type proton ATPase 116 kDa subunit a isoform 1
Q3UMB9	WASH7					+	WASH complex subunit 7
Q8R3E3	WIPI1	++		+			WD repeat domain phosphoinositide-interacting protein 1
Q80W47	WIPI2					++	WD repeat domain phosphoinositide-interacting protein 2
Q924H7	WAC			+	++	++	WW domain-containing adapter protein with coiled-coil
Q91WL8	WWOX					+	WW domain-containing oxidoreductase
Q6P1B1	XPP1					+	Xaa-Pro aminopeptidase 1
Q6ZPZ3	ZC3H4					+	Zinc finger CCCH domain-containing protein 4
Q810J8	ZFYV1			+		+	Zinc finger FYVE domain-containing protein 1
Q5DU37	ZFY26	+		+	++	+	Zinc finger FYVE domain-containing protein 26