Dissecting Gene Regulatory Networks Using Targeted Quantitative Proteomics

THÈSE N^{O} 5482 (2012)

PRÉSENTÉE LE 9 NOVEMBRE 2012
À LA FACULTÉ DES SCIENCES DE LA VIE
UNITÉ DU PROF. DEPLANCKE
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Jovan SIMICEVIC

acceptée sur proposition du jury:

Prof. F. G. van der Goot Grunberg, présidente du jury
Prof. B. Deplancke, directeur de thèse
Prof. R. Aebersold, rapporteur
Prof. A. Bairoch, rapporteur
Prof. F. Naef, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2012

ACKNOWLEDGMENTS

During the years spent at EPFL working towards earning a doctoral degree in Biotechnology & Bioengineering I was gifted to meet an ample number of rather extraordinary people. I would like to begin with the ones that have impacted my life in the most profound manner. First, I, I would like to thank my Professor, Bart Deplancke, to whom I am greatly indebted for accepting me in his group. He possesses the rare combination of leadership and scientific innovative thinking from which I have and will continue to learn, in science as well as in other aspects of life. He has indeed greatly impacted my life in many ways, and I cannot find enough words to express my gratitude for his support all along my scientific journey as a member of his laboratory. I would also like to thank Frédérique Lisacek, for her collaborative effort and for her support in bioinformatics, and my mentor Professor Patrick Fraering. Furthermore, I am grateful to thank Marc Moniatte for his guidance and precious practical advices in mass spectrometry as well as in general aspects of proteomics, and to Adrian Schmid for teaching me several aspects of mass spectrometry, selected reaction monitoring in particular, and for his determination and dedication to his work. Without them, without their true interest, dynamic discussion and perseverance in achieving goals, the development of our selected reaction monitoring-based assay would have simply not been possible. I would like to thank my collaborators in Geneva, Markus Mueller and Erik Ahrné in particular, who on top of being a good bioinformatician is also a good friend of mine.

I would like to thank SystemsX.ch, for having funded my PhD project and for having granted additional funding for my last year. In particular, I express my gratitude to the members of its Scientific Advisory Board, as well as to those who have been involved in assessing my project. Furthermore, I am grateful to Professor Gisou Van der Goot, president of the jury of my oral thesis examination, as well as Professors Felix Naef, Ruedi Aebersold, and Amos Bairoch, for having accepted to review this manuscript and for taking the journey to Lausanne to participate at the jury of my PhD defense. During my years of study at EPFL, I have been gifted to have exceptional mentors in my laboratory, Korneel Hens and Sunil Kumar Raghav, as well as a large number of close collaborators: Carine Delattre-Gubelmann, Celine Riera, Alina Isakova, Sebastian Waszak, Jean-Daniel Feuz, Benjamin Zoller, Irina Krier. My deepest gratitude goes to all of my colleagues and those with whom I have exchanged opinions and ideas here at EPFL as well as in other locations worldwide. They have all one way or another brought new and inspiring perspectives to my work and contributed in the development of my research project. Furthermore, I am thankful to Paola Gilardoni for critically reviewing this manuscript. Moreover, I sincerely acknowledge the jury members of my candidacy exam, Professor Johann Auwerx and Manfredo Quadroni, for having read my first year proposal. Today's life science technology development is becoming more and more the result of a cooperative effort, and this interdisciplinary work is not an exception. I am greatly indebted to Marc Moniatte for

having allow me to utilize his proteomics core facility and putting his instruments at my disposal. As well, I appreciate the efforts of Romain Hamelin and Diego Chiappe in helping me with the shotgun analyses of the samples, and many thanks to Florence Armand and Markus Mueller for their advices in bioinformatics. I would like to thank Professor Yury Tsybin for his kindness and for having welcomed me in his laboratory, and Luca Fornelli for his help with sample preparation, as well as to Matthias Gstaiger for providing the SH-tag peptide. As well, I would like to thank Alexander Schmidt for his precious advices and for having read my first research article. I am thankful to Brigitta Saoul from Promega for their collaboration on the *in vitro* protein expression systems. I express my sincere appreciation to my friends as well as colleagues Jean-Daniel Feuz, Andrej Bicanski, Martial Sankar, Tyler Thacher, Adam Klaptocz, Mitko Dimitrov, Tristan Bolmont, David Hacker, Evan Williams, Raffaele Teperino, Suzanna Kallionen, Ben MacMillan, and particularly Luca Fornelli. All of you made my stay at EPFL a wonderful experience; I will cherish the time spent in your company during the last four years. I would also like to thank all the members of the LSBG family as well as to the visiting students, for their day to day help and kindness. I extend my utmost gratitude to the laboratory's secretary Marie-France for her administrative support.

It would not be here today presenting my accomplishment if I had not had a truly amazing family. I seriously do not know where I would have ended up in life without their unconditional love and continuous support. I consider myself extremely lucky to be able to call Zdenka and Dr. Branislav Simicevic, Mom and Dad. As well, I would like to thank Alexandra for having, on top of being the best sister a brother can possibly have, a true friend. Moreover, I would like to thank all of my relatives, particularly my uncle Gradimir, and those ones who have passed onto a better life. Last but certainly not least, I want to thank Tamara, my wife, who has brought happiness, joy, and endless love into my life. Thank you my love, for putting up with me on a daily basis and encourage me to become not only a better scientist, but more importantly a better person. Thank you for having been with me in difficult times, thank you for caring and showing never-ending love, in good times, as well as in times of adversity. I would also like to express my most profound gratitude to my wife's family, in particular to Velimir, Mira and Zoran.

Ultimately, I would like to thank all of my friends and people all over the globe that have supported and inspired me. Thank you Martin and Katarina Janicek, Michele Zinnai, Sandro Oppliger, Thomas Milia, Andrea Bornaghi, Milan Antonovic, Dino Paraskevopoulos, Acdrick Mustin, John Ferioli, Damian Barton, Vanessa and Joel Tinker, the Lohn-Tinker family, Michael Hofer, Austen Chongpison, Michael Colby, Brandon Felkins, Dustin Duran, Matthew Fribourg, Matthew Essenburg, Kevin Wang, Dev Sriranganadane, Giulio Cattarossi, Mauro Mileni, Felix Zelder, Edoado Menozzi, Sara Butterfield, Nada Mitic, Vlada Randjelovic, Scott Hamilton, Noah Zeiser, Sean Bowman, Dave Anning, Steve George, and Tatjana Sajic. I would like to thank to Francesco Conti in particular for having encouraged me to study life sciences. And please, accept my sincere apologies, if you cannot find your name within the list. There are simply too many of you. I would also like to extend my gratitude and affection to collaborators of the University of Lausanne, San Diego State University, University of California San Diego, University of Geneva, EPFL, as well residents of my many home-towns: Lugano, Belgrade, Požega, San Diego … Pacific Beach residents in particular and all of those who I have one way or another encountered … perhaps on the beach, surfing, or simply because you were there. My conviction has always been to achieve high education, and spread my knowledge and wisdom were most needed. It is to all of you that I dedicate this thesis.

“Of all the frictional resistances, the one that most retards human movement is ignorance, what Buddha called 'the greatest evil in the world.' The friction which results from ignorance can be reduced only by the spread of knowledge and the unification of the heterogeneous elements of humanity. No effort could be better spent.”

Nikola Tesla

TABLE OF CONTENTS

ABSTRACT

Gene regulatory networks control gene expression levels, and therefore play an essential role in mammalian development and function. Regulation of gene expression is the result of a complex interplay between DNA regulatory elements and their binding partners, known as transcription factors (TFs). Due to their vital role in development, intercellular signalling, cell cycle and disease development, elucidating the mechanisms by which TFs regulate gene expression is of crucial importance in the vast majority of biological processes. In particular, understanding how each TF contributes to the expression output of its respective target gene in space and time will help to elucidate how gene regulatory networks (GRNs) behave under different physiological or pathological conditions. Although extensive work has been accomplished in characterizing the key TFs involved in many biological processes, almost no quantitative information is currently available in the literature. To get a deep insight into the complex mechanisms underlying the regulation of gene expression, we need to acquire quantitative information, since TF abundance within the cell can be linked to their transcriptional capabilities. Such information would be of utmost importance to build accurate *in silico* quantitative DNA binding models that could predict and explain the particular properties of gene regulatory mechanisms. The quantification of TFs is a difficult task due their natural low abundance in cells, and their reliable detection is therefore very much dependent on the overall sensitivity of current technologies. In recent years, a new MS-based technology termed selected reaction monitoring (SRM) has gained popularity due to the targeted nature of its approach that allows the detection and quantification of proteins in complex samples with an exceptional sensitivity and specificity. I will show in this thesis, this approach is particularly well suited for targeting low abundant proteins such as TFs, which are otherwise difficult to identify with conventional *shotgun* proteomics experiments.

Consequently, the main focus of my thesis research project entailed the development of an SRM-based platform aimed at quantifying TFs in absolute amounts based on *in vitro* protein expression during the terminal stage of adipogenesis, using the pre-adipocyte 3T3-L1 cell line. Interestingly, our initial efforts led to the creation of an atlas of TF-specific peptide data, which could be readily used for the design of quantitative assays. In the first phase, abundance measurements in terms of copies per cell were derived at precise differentiation time-points for two major adipogenic players, PPARγ and RXRα. In the second phase, we expanded the number of adipogenic TFs that can be monitored in one assay, allowing for the quantification of up to 10 TFs in one single, integrated SRM run. Such upscale increases the practical usefulness of the methodology while reducing the associated costs, and ultimately allows for non-negligible time-savings. The availability of absolute protein copy number data permitted us ultimately to examine the relationship between the number of genome-wide DNA binding events and TF molecules. We derived a quantitative DNA binding model that allowed the prediction of the number of PPARγ ChIP-seq binding events given its nuclear abundance, chromatin state, and DNA binding energetics. As such, we were able to explain the paradoxical observation of a significant increase in PPARγ binding sites despite a saturation in the number of PPARγ molecules. We thus demonstrate how TF abundance data can be modeled in conjunction with large-scale DNA occupancy and chromatin state data to further our understanding of gene regulatory mechanisms mediating cellular differentiation.

We are now starting to build on our pioneering work to quantify in absolute terms key players of the entire, core adipogenic GRN, as such aiming to provide a quantitative explanation of the regulatory mechanisms at play during the terminal phase of adipocyte differentiation. Moreover, to increase the explicative power of our methodology and to alleviate the throughput limitation that comes with obtaining absolute protein measurements, we decided to perform copy-number estimates for a larger set of adipogenic TFs utilizing a modified version of our original approach. At the cost of a modest loss of accuracy, we are now aiming to develop a sensitive and robust methodology that will allow the quantification of entire GRNs at low cost and in a time-effective manner. This is consistent with the overall goal in life sciences or clinical research to improve our ability to accurately and reproducibly quantify entire pathways or biological networks to improve our systems understanding of biological processes.

Keywords:

Mass spectrometry, proteomics, selected reaction monitoring, multiple reaction monitoring, multiplexing, absolute quantification, quantitative proteomics, protein copy number, systems biology, transcription factor, regulation of gene expression, gene regulatory networks, adipogenesis, terminal differentiation, quantitative modeling, binding events.

RÉSUMÉ

Les réseaux de régulation de gènes contrôlent l'expression des gènes, et donc jouer un rôle essentiel dans le développement des mammifères et dans leur fonctionnement. La régulation de l'expression des gènes est le résultat d'une interaction complexe entre les éléments régulateurs de l’ADN et leurs partenaires de liaison, qui sont connues sous le nom de facteurs de transcription (FT). En raison de leur rôle vital dans le développement, dans la signalisation intercellulaire, dans le cycle cellulaire et dans le développement des maladies, l'élucidation des mécanismes par lesquels les FTs régulent l'expression génique est d'une importance cruciale dans la grande majorité des processus biologiques. En particulier, comprendre comment chaque FTs contribue à l’expression de son gène cible respectivement dans l'espace et dans le temps nous aiderait à mieux comprendre comment les réseaux de régulation des gènes (GRNs) se comportent dans différentes conditions physiologiques ou pathologiques. Bien que d'importants travaux aient été réalisés dans la caractérisation des facteurs de transcription impliqués dans de nombreux processus biologiques, presque aucune information quantitative n’est actuellement disponible. Pour obtenir un aperçu des mécanismes complexes qui sous-tendent la régulation de l'expression des gènes, nous avons besoin d'acquérir des informations quantitatives, car l'abondance des FTs dans la cellule est liée à leurs capacités de transcription. Une telle information serait d'une importance capitale pour construire des modèles quantitatifs qui pourraient prédire et expliquer les propriétés particulières des mécanismes de régulation des gènes. La quantification des FTs est une tâche difficile en raison de leur faible abondance naturelle dans les cellules, et leur détection est donc dépendante de la sensibilité globale des technologies actuelles. Ces dernières années, une nouvelle technologie basée sur la spectrométrie de masse appelée *Selected reaction monitoring* (SRM) a gagné en popularité en raison de la nature ciblée de son approche qui permet la détection et la quantification des protéines dans des échantillons complexes avec une sensibilité et une spécificité exceptionnelle. Dans cette thèse, je montre que cette approche est particulièrement bien adaptée pour cibler des protéines de faible abondance comme les FTs, qui sont autrement difficiles à identifier par la voie classique de la protéomique *shotgun*.

Par conséquent, l'objectif principal de mon projet de recherche de doctorat concernait le développement d'une plate-forme SRM visant à quantifier les FTs de façon absolue en se basant sur l'expression de protéines *in vitro* dans la phase terminale de l'adipogenèse, en utilisant la lignée cellulaire de pré-adipocytes 3T3-L1. Il est intéressant de noter que nos efforts initiaux nous ont a conduit à la création d'un atlas de données peptidiques FT-spécifiques, qui pourraient être facilement utilisés pour la conception d’autre tests quantitatifs. Dans la première phase, les mesures d'abondance en termes de copies par cellule ont été dérivées à des temps de différentiation précis pour deux régulateurs principaux de l’adipognèse, PPARγ et RXRα. Dans la deuxième phase, nous avons élargi le nombre de FTs adipogéniques, permettant la quantification jusqu'à 10 FTs dans une seule analyse SRM intégrée. Cet incrément augmente l'utilité pratique de la méthodologie, tout en réduisant les coûts associés, et permet en fin de compte un gain de temps. La disponibilité de données absolues de protéines en termes de nombre de copies par cellule nous a permis finalement d’examiner la relation entre le nombre de liaison FT-ADN dans le génome entier et le nombre de molécules FT. Nous avons établi un modèle de liaison FT-ADN quantitative permettant la prédiction du nombre de liaison de PPARγ (par ChIP-seq) en prenant compte de son abondance nucléaire, l'état de la chromatine, et le bilan énergétique de liaison à l'ADN. En tant que tel, nous étions en mesure d'expliquer l'observation paradoxale d'une augmentation significative dans les sites de liaison PPARγ en dépit d'une saturation du nombre de molécules PPARγ. Nous avons donc montré comment les données sur l’abondance des FTs peut être modélisé, en collaboration avec l’information sur l'occupation de l'ADN à grande échelle et les données sur l’état de la chromatine, afin de faire progresser notre compréhension des mécanismes de régulation des gènes qui contrôlent la différenciation cellulaire.

Nous allons appliquer notre technique pour la quantification en termes absolus des principaux acteurs impliqués dans le réseau des gènes adipogénique. Nous visons à fournir une explication quantitative des mécanismes de régulation en jeu au cours de la phase terminale de la différenciation adipocytaire. Par ailleurs, pour augmenter la puissance explicative de notre méthodologie et pour atténuer la limitation de débit et de l'obtention de mesures absolues, nous avons décidé d'effectuer des estimations par rapport au nombre de copies par cellule pour un ensemble plus vaste de FTs adipogéniques en utilisant une version modifiée de notre approche originale. Au prix d'une perte modeste de précision, nous visons maintenant à élaborer une méthodologie sensible et robuste qui permettra la quantification de GRN entières à faible coût et de manière efficace. Ceci est cohérent avec l'objectif global, en sciences de la vie ou dans la recherche clinique, d’améliorer notre capacité à quantifier de façon précise et reproductible des voies de signalisation ou des réseaux biologiques afin d'améliorer notre compréhension des systèmes biologiques.

Mots-clés:

Spectrométrie de masse, protéomique, selected reaction monitoring, multiple reaction monitoring, multiplexage, quantification absolue, protéomique quantitative, copies des protéines par cellule, biologie des systèmes, facteurs de transcription, régulation de l'expression des gènes, réseaux de régulation des gènes, adipogenèse, différenciation terminale, modélisation quantitative, événements de liaison de l’ADN .

LIST OF ACRONYMS

ChIP: chromatin immuno-precipitation

ChIP-seq: chromatin immuno-precipitation sequencing

CV: Coefficient of variation

GNR: gene regulatory network

LC: liquid chromatography

LLOD: lower limit of detection

NE: nuclear extract

MS: mass spectrometry

ORF: open reading frame

SD: standard deviation

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SME: standard error of the mean

SRM: selected reaction monitoring

SAX: strong anion exchange

SCX: strong cation exchange

TF: transcription factor

I. INTRODUCTION

I.I. DISSECTING GENE REGULATORY MECHANISMS

Understanding how the expression of large sets of genes is orchestrated at the systems-level is a topic of fundamental importance in Systems biology. Its ultimate goal is to integrate different levels of information to obtain a global view of how complex biological systems function. In this regard, the vast majority of biological processes, from homoeostasis maintenance to development, from cell cycle to cell differentiation are tuned by differential gene expression. The latter is under the control of gene regulatory networks (GRNs), which consist of physical and functional interactions between DNA-binding regulatory proteins, transcription factors (TFs), and regulatory elements of their target genes (i.e. promoters, enhancers). The key function of these networks is to coordinate the progression of distinct transcription regulatory states both in space and time, a *sine qua non* condition for cell survival. Unfortunately, most models of GRNs are incomplete and their players far from being completely characterized. In essence, regulation of gene expression, therefore of the proteins that genes encode, can be seen as a multi-layered process involving several steps and a vast number of participants. The final outcome is the result of the combined effects of the multiple events affecting a particular gene, from the early stages of transcription all the way to the modification that may occur on the protein once the transcript has been translated. At the transcriptional level, understanding how each TF contributes to the expression output of its respective target gene in space and time will help us to understand how gene regulatory networks behave.

Although qualitative information explaining TF interactions and behavior is widely available in the literature[1-3], reliable quantitative proteomic measurements are hardly available. This disparity is explained by TF natural low abundance in cells, which make quantitative analyses a substantial challenge considering the current state of proteomics technologies. Reliable quantitative data would be extremely useful for *in silico* modelling for computational biologists. In the last decade, quantitative models have been generated aiming to predict gene expression levels using as inputs selected TFs, whose abundance and hence activity was inferred from mRNA levels. However, the reliability of mRNA-protein activity correlations appears to be inadequate, limiting the predictive power of the respective models[4-6]. Thus, absolute TF levels need to be experimentally measured to determine how much of a binding site will be occupied by a TF in order to assess its regulatory input. Moreover, the pivotal role that TFs play in dictating normal development and proper functioning of the organism animates an increasing interest for this particular family of proteins to become potential pharmaceutical targets in the therapeutics of various diseases, including cancer[7-9]. Transcription-based drugs represent a significant percentage of the drugs currently present in the market [10]. Hence, accurate quantitative TF data is of critical importance to effectively understand the role and behaviour of TFs as regulatory proteins also from a medical perspective. Aberrations in the fine-tuning mechanisms of regulation could be related to disease processes. This in turn could help in the design of more effective medicaments.

Various techniques are available for the quantification of proteins nowadays, the majority of which rely on the use of antibodies (e.g. ELISA, protein microarrays). Quantitative immunoassays are widely used because of their accuracy and the fact that they can be implemented even in small laboratories due to their low cost and simplicity. Nevertheless, these techniques suffer from a certain number of drawbacks, non-linearity in quantitation and formation of unspecific reactions to name a few. Moreover, quantitation of different proteins normally necessitates separate experiments, not to mention that only a limited number of TF-specific antibodies are commercially available, limiting thereby the applicability of these methodologies to small scale studies of several TFs at the time. In this regard, there is a strong need for a robust methodology that could bypass such limitations, possibly pushing the current limits in terms of sensitivity and specificity even farther. Until recently, mass spectrometry-based methodologies simply lacked the necessary sensitivity to be used for the identification of low abundant proteins.

I.II THE ADIPOGENIC MODEL

The terminal adipocyte differentiation is the last phase of adipogenesis, during which pre-adipocytes develop into mature adipocytes through a cascade of gene expression events. It is a natural process, consequence of normal cell turnover on one hand, as well as a necessity for fat mass storage in case of excessive weight gain. The study of adipogenesis has a clear medical relevance: excess fat mass, characterized by an increase in cell size and number, dramatically increases the risk of developing series of pathologies, including metabolic syndrome symptoms and cancer[11]. Several studies have established a basic framework of the GRN orchestrating the terminal phase of adipogenesis[12, 13]. Although substantial effort has been devoted to identify TFs and co-TFs involved in adipogenesis[14, 15], almost no information is available on adipogenic TF levels in the nucleus or in the cytoplasm. In this regard, understanding adipogenic TF contribution to target gene expression output during adipocyte differentiation may give further insight on how the adipogenic regulatory network behaves under different physiological or pathological conditions, opening possibly new venues for disease diagnostic and cure. The murine embryonic 3T3-L1 preadipocyte cell line recapitulates most of the aspects of terminal adipocyte differentiation observable *in vivo*[16]. Their homogeneity, synchronous development, as well as their availability makes this cell line the perfect model candidate for the study of adipogenesis.

I.III *STATE OF THE ART* QUANTITATIVE PROTEOMICS

Noticeable improvements in mass spectrometer detection limits have opened the doors to the quantitative analysis of proteins that are expressed at very low levels in cells, such as TFs. The majority of these techniques employ metabolic (SILAC), chemical (ICAT, iTRAQ) or enzymatic (digestion with $^{16}O/^{18}O$) stable isotope labelling to introduce a predictable mass shift between peptides from two or more experimental conditions (each methodology is reviewed in Ong and Mann[17]). Relative quantification is achieved by comparing “heavy” to “light” peptide signal intensities. These applications allowed capture of temporal changes and comparison among proteomes, and have become in the past decade a gold standard in biomedical applied proteomics[18, 19].

The increasing need for researchers to obtain accurate protein measurements has spurred an increasing interest to develop methodologies aimed at quantifying fractions of the proteome in absolute amounts. Despite such efforts, absolute quantification remains rather challenging from a technical perspective when compared to relative quantification[20]. Nevertheless, recent improvements in sensitivity and throughput have allowed for the routine implementation of absolute quantification methodologies. In essence, one can segregate MS-based quantification methodologies into two major classes: those requiring stable isotope labelling and those that do not necessitate labelling, so called “label-free”. Each of them carries a number of advantages as well as limitations. Label-free strategies are proven to provide a rather dynamic range of quantification, are of simple implementation and amicable to up-scaling. Unfortunately, this may come at the cost of accuracy and linearity when based on spectral counting. Recently, considerable effort has been devoted to overcome such limitations[21], and combine the benefits of label-free with the sensitivity of targeted MS approaches[22].

Stable isotope-labelling quantification entails the spiking of “heavy”-labelled peptides, such as AQUA[23], in which selected chemically synthesized isotope-labelled peptides are carefully quantified and used as standards. “Heavy”-to-“light” peptide ratios define the amount of the endogenous protein present within the biological sample. Issues with the cost of these peptides and with its storage have pushed for an amelioration of the technique, in which concatenated peptides (QconCAT)[24] are utilized. One of the criticisms though regarding QconCAT constructs resides in the digestion of its tryptic peptides, which does not mirror the digestion of endogenous proteins. For this particular reason, many laboratories have oriented their methodologies towards the expression of full-length proteins, expressed either *in vivo* (Absolute SILAC)[25] or *in vitro* (PSAQ)[26], which are spiked at some stage of sample preparation within the complex mixture. The main advantage is that all tryptic peptides generated from the protein, except the C-teminal one, can be readily monitored. This application tries to overcome issue related to peptide detection, because a large fraction of protein-specific peptides may not be identifiable due to sample complexity, solubility and ionization issues. Furthermore, by selecting only a small subset of peptides the methodology is more sensitive to post-translational modification. Methods based on full-length protein expression allow also for a more accurate quantification and a more robust statistical assessment; in addition precipitation issues related to peptide storage are systematically bypassed. It is generally agreed that the spiking of the labelled standard should be introduced as early as possible within the stages of sample fractionation, as to secure that both the standard and its endogenous counterpart are subject to the same artefacts. Therefore, sample losses or differential proteolytic treatments that may affect downstream measurements are minimized. Recently, a novel implementation of the absolute SILAC methodology, named SILAC-PrEST[27], introduces the use of a solubilisation tag to quantify in absolute terms the amount of recombinant PrESTs (Protein Epitope Signature Tags) produced *in vivo* as to quantify 40 selected proteins in HeLa cells. One of the benefits of the ABP (albumin binding protein) solubilisation tag resides in the fact that most of its tryptic peptides can be used for quantification, increasing the overall robustness of the approach. In addition, the experiments have been carried by switching the "heavy" versus "light" isotope incorporation of the PrESTs and of the sample, further corroborating its robustness. Ultimately, the two steps of quantification of the recombinant PrESTs and of their endogenous counterparts have been collapsed in one single experiment, simplifying the workflow as a whole.

To sum up, novel applications in quantitative proteomics and advances in MS technology development allowed us to accurately measure protein amounts in a given mixture in absolute terms. One particular technique, named Selected Reaction Monitoring (SRM), also known as Multiple Reaction Monitoring (MRM), is becoming a benchmark in targeted proteomics approaches, for it allows the detection and quantification of predetermined sets of proteins, based on selected peptide fragmentation reactions, in complex samples with previously unseen sensitivity and specificity. This in turn allows for a more in-depth analysis in the proteome, particularly for those proteins that are expressed at such low levels that fail detection with canonical MS approaches. The most important aspect of SRM, when it comes to quantification, is the consistency and the uniqueness of the peptides selected. Only peptides that uniquely identify a protein of choice, and that are consistently detected in different MS runs should be utilized; such peptides are termed "proteotypic" [28]. Moreover, when selecting such peptides, one has to be careful as to select the highest responding peptides for each protein of interest. There is no *gold standard* for the identification of such peptides. Nevertheless, several bioinformatic tools that guide the user in the selection of proteotypic peptide candidates based on a set of physico-chemical properties are currently available[29] (e.g. Pinpoint: Thermo Scientific, Waltham, USA). The best responding peptides are usually selected for validation (e.g. ESP Predictor[30]). In recent years, SRM coupled to stable-isotope labelling techniques has been adopted for estimating cellular protein levels in large-scale proteomic analyses[31, 32]. Most of such efforts were aimed at quantifying large fraction of the proteome, covering the largest possible dynamic range. As the complexity of the model system studied increases, mainly due to technical limitations, venturing in the lower levels of the dynamic range becomes extremely challenging. Quantitative information available on TFs appears to be rather scarce compared to the biological importance of this family of DNA-binding proteins. Pioneering work has been accomplished in bacteria and low eukaryotes first[33, 34] before measuring copy-numbers per cell in mouse and human[32, 35, 36]. Mann and co-workers were able to accurately quantify proto-oncogene c-Fos at approximately 5-6,000 copies per cell, and Zfp828 at approximately 70-75,000 copies per cell utilizing the SILAC-PrEST methodology in HeLa cells[27]. Aebersold and colleagues have estimated of the lowest copy-number per cell of several transcription factors including Zfp608, Zfp335, Zbtb40, Zbtb48, and E2F7 at less than 500 copies per cell utilizing a stable-isotope labelling SRM-based technique[32]. Thus, although transcription factors are being fished out in the context of large-scale efforts aimed at determining copy-numbers for a fraction of an organism's entire proteome, no comprehensive study focused at quantitatively monitoring TFs dynamic behaviour in biological process has been carried out.

The adipogenic regulatory network has been widely studied because it has a clear medical relevance. A significant effort has been devoted to establish the basic framework of the GRN orchestrating the terminal phase of adipogenesis[12, 13]. In particular, several transcriptomic studies have been implemented to understand what are the specific cellular mechanisms that take place during this period of time in the cell's life[37, 38]. Interestingly, recent genome-wide binding analyses have revealed that a few TFs alone are responsible for a wide range of gene expression events in the early stages of differentiation as well as in the mature adipocytes[39]. The poor mRNA-protein activity correlations experimentally observed have fostered and interest to employ proteomic methodologies to obtain protein expression profiles during adipocyte differentiation, tackling thus the problem from a different angle. Kim and co-workers examined the changes in protein expression resulting from hypoxia and normoxia in a 3T3-L1 murine cell line[40]. Ahmed and colleagues monitored the changes in mRNA and protein and profile of adipose tissue in response to drug treatment (rosiglitazone) to identify several potential protein targets for insulin sensitization[41]. Using a 5-plexed SILAC-based MS approach Pandey and co-workers identified 882 nuclear and secreted proteins at 5 different time-points of adipogenesis. For about half of them relative quantitative measurements were obtained[42]. In the above mentioned studies, TFs are strongly underrepresented if identified at all. The knowledge of TF levels in absolute terms is absolutely needed to effectively understand the dynamics of adipocyte biology.

Last year, an interesting application, adapted by the developers to the quantification of TFs, has been built by the MacCoss laboratory at the University of Washington (U.S.A), which consists of a high-throuput, cost-effective methodology for the discovery of optimal precursor- and fragment-ions to be utilized in targeted proteomics assays based on the use of *in vitro*-synthesized full-length proteins. Absolute quantification of *in vitro*-expressed TFs is accomplished via two GST (Glutathione-S-transferase) signature peptides. Using their approach, MacCoss and colleagues were able to experimentally derive optimal transitions for 96 human TFs, which expression and enrichment was verified for 44 TFs in total. The utility of the derived ion transitions to quantify endogenous TFs was tested by measuring relative abundances of 6 candidates between 4 human cell lines[43]. To conclude, although there is a growing interest in adopting proteomic-based methodology, cost-effective and sensitive methodology to achieve absolute TF quantification are currently lacking. Thus, to date, no study targeting key regulatory TFs that provides a quantitative analysis of the dynamics of a cellular process of interest such as adipocyte differentiation in absolute terms has been published.

REFERENCES

1. Kalra, I.S., Alam, M.M., Choudhary, P.K. & Pace, B.S. Kruppel-like Factor 4 activates HBG gene expression in primary erythroid cells. *British journal of haematology* **154**, 248-259 (2011).
2. Akinyeke, T.O. & Stewart, L.V. Troglitazone suppresses c-Myc levels in human prostate cancer cells via a PPARgamma-independent mechanism. *Cancer biology & therapy* **11**, 1046-1058 (2011).
3. Bintu, L. et al. Transcriptional regulation by the numbers: applications. *Current opinion in genetics & development* **15**, 125-135 (2005).
4. Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome biology* **4**, 117 (2003).
5. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and cellular biology* **19**, 1720-1730 (1999).
6. Maier, T., Guell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS letters* **583**, 3966-3973 (2009).
7. Choi, J.H. et al. Antidiabetic actions of a non-agonist PPARgamma ligand blocking Cdk5-mediated phosphorylation. *Nature* **477**, 477-481 (2011).
8. Karamouzis, M.V., Gorgoulis, V.G. & Papavassiliou, A.G. Transcription factors and neoplasia: vistas in novel drug design. *Clinical cancer research : an official journal of the American Association for Cancer Research* **8**, 949-961 (2002).
9. Langlois, M.C., Beaudry, G., Zekki, H., Rouillard, C. & Levesque, D. Impact of antipsychotic drug administration on the expression of nuclear receptors in the neocortex and striatum of the rat brain. *Neuroscience* **106**, 117-128 (2001).
10. Emery, J.G., Ohlstein, E.H. & Jaye, M. Therapeutic modulation of transcription factor activity. *Trends in pharmacological sciences* **22**, 233-240 (2001).
11. Ailhaud, G. Adipose tissue as a secretory organ: from adipogenesis to the metabolic syndrome. *Comptes rendus biologies* **329**, 570-577; discussion 653-575 (2006).
12. Farmer, S.R. Transcriptional control of adipocyte formation. *Cell metabolism* **4**, 263-273 (2006).
13. Rosen, E.D. & MacDougald, O.A. Adipocyte differentiation from the inside out. *Nature reviews. Molecular cell biology* **7**, 885-896 (2006).
14. Oishi, Y. et al. Kruppel-like transcription factor KLF5 is a key regulator of adipocyte differentiation. *Cell metabolism* **1**, 27-39 (2005).
15. Fujimori, K. & Amano, F. Forkhead transcription factor Foxa1 is a novel target gene of C/EBPbeta and suppresses the early phase of adipogenesis. *Gene* **473**, 150-156 (2011).

16. Green, H. & Meuth, M. An established pre-adipose cell line and its differentiation in culture. *Cell* **3**, 127-133 (1974).

17. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology* **1**, 252-262 (2005).

18. Gronborg, M. et al. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Molecular & cellular proteomics : MCP* **5**, 157-171 (2006).

19. Afkarian, M. et al. Optimizing a proteomics platform for urine biomarker discovery. *Molecular & cellular proteomics : MCP* **9**, 2195-2204 (2010).

20. Brun, V., Masselon, C., Garin, J. & Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *Journal of proteomics* **72**, 740-749 (2009).

21. Griffin, N.M. et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature biotechnology* **28**, 83-89 (2010).

22. Hansson, J. et al. Time-resolved quantitative proteome analysis of in vivo intestinal development. *Molecular & cellular proteomics : MCP* **10**, M110 005231 (2011).

23. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6940-6945 (2003).

24. Pratt, J.M. et al. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nature protocols* **1**, 1029-1043 (2006).

25. Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *Journal of proteome research* **7**, 1118-1130 (2008).

26. Brun, V. et al. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Molecular & cellular proteomics : MCP* **6**, 2139-2149 (2007).

27. Zeiler, M., Straube, W.L., Lundberg, E., Uhlen, M. & Mann, M. A protein epitope signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP* (2011).

28. Craig, R., Cortens, J.P. & Beavis, R.C. The use of proteotypic peptide libraries for protein identification. *Rapid communications in mass spectrometry : RCM* **19**, 1844-1850 (2005).

29. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).

30. Fusaro, V.A., Mani, D.R., Mesirov, J.P. & Carr, S.A. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature biotechnology* **27**, 190-198 (2009).

31. Costenoble, R. et al. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Molecular systems biology* **7**, 464 (2011).

32. Beck, M. et al. The quantitative proteome of a human cell line. *Molecular systems biology* **7**, 549 (2011).

33. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795-806 (2009).

34. Malmstrom, J. et al. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460**, 762-765 (2009).

35. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

36. Vogel, C. et al. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular systems biology* **6**, 400 (2010).

37. Soukas, A., Socci, N.D., Saatkamp, B.D., Novelli, S. & Friedman, J.M. Distinct transcriptional profiles of adipogenesis in vivo and in vitro. *The Journal of biological chemistry* **276**, 34167-34174 (2001).

38. Burton, G.R., Nagarajan, R., Peterson, C.A. & McGehee, R.E., Jr. Microarray analysis of differentiation-specific gene expression during 3T3-L1 adipogenesis. *Gene* **329**, 167-185 (2004).

39. Nielsen, R. et al. Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes & development* **22**, 2953-2967 (2008).

40. Choi, S. et al. Comparative proteome analysis using amine-reactive isobaric tagging reagents coupled with 2D LC/MS/MS in 3T3-L1 adipocytes following hypoxia or normoxia. *Biochemical and biophysical research communications* **383**, 135-140 (2009).

41. Ahmed, M., Neville, M.J., Edelmann, M.J., Kessler, B.M. & Karpe, F. Proteomic analysis of human adipose tissue after rosiglitazone treatment shows coordinated changes to promote glucose uptake. *Obesity* **18**, 27-34 (2010).

42. Molina, H. et al. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *Journal of proteome research* **8**, 48-58 (2009).

43. Stergachis, A.B., Maclean, B., Lee, K., Stamatoyannopoulos, J.A. & Maccoss, M.J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nature methods* **8**, 1041-1043 (2011).

I.IV DNA-CENTERED APPROACHES TO CHARACTERIZE REGULATORY PROTEIN–DNA INTERACTION COMPLEXES

Our capabilities to elucidate the dynamic mechanisms by which TFs regulate the expression of a target gene is strongly influenced by the methodology we adopt. Until recently, protein-centered methodologies have been widely implemented particularly due to the availability of TF-specific antibodies. Recently, there has been a growing consensus in favour of the notion of DNA directly influencing DNA-binding protein complex formation. This paradigm shift catalysed the interest of understanding how DNA nuclear composition dictates TF recruitment and protein complex formation, with the aim of providing a holistic explanation of dynamic gene regulatory mechanisms. This review summarizes the state-of-the-art and focuses on novel areas of improvement of gene-centered approaches, with a particular attention to the years to come.

DNA-centered approaches to characterize regulatory protein–DNA interaction complexes

Jovan Simicevic and Bart Deplancke *

Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences, Institute of Bioengineering, Station 15, 1015 Lausanne, Switzerland. E-mail: bart.deplancke@epfl.ch; Tel: +41-21-6931821

Received 5th August 2009 , Accepted 6th October 2009

First published on the web 2nd December 2009

Gene regulation is mediated by site-specific DNA-binding proteins or transcription factors (TFs), which form protein complexes at regulatory loci either to activate or repress the expression of a target gene. The study of the dynamic properties of these regulatory DNA-binding complexes has so far been dominated by protein-centered methodologies, aiming to characterize the DNA-binding behavior of one specific protein at a time. With the emerging evidence for a role of DNA in allosterically influencing DNA-binding protein complex formation, there is renewed interest in DNA-centered approaches to capture protein complexes on defined regulatory loci and to correlate changes in their composition with alterations in target gene expression. In this review, we present the current state-of-the-art in such DNA-centered approaches and evaluate recent technological improvements in the purification as well as in the identification of regulatory DNA-binding protein complexes within or outside their biological context. Finally, we suggest possible areas of improvement and assess the putative impact of DNA-centered methodologies on the gene regulation field for the forthcoming years.

WHY DNA-CENTERED METHODS?

Differential gene expression is central to most fundamental biological processes and is controlled by site-specific DNA binding protein complexes. The latter transcriptional complexes, of which transcription factors (TFs) are the core members, function by integrating extra- and intracellular cues through protein–protein or protein–ligand interactions and translating these cues into a gene regulatory output by binding to gene regulatory elements.[1] Signal integration can thereby be directly mediated by the TF itself, for example, through post-translational modification (PTM) of TF domains which modulates its activity[2] or cellular location,[3] or can be controlled indirectly through interaction with co-regulators. These higher-order interactions can result in, or can also be the result of PTMs, and can then determine whether the TF-containing complex acts as an activator or repressor of gene expression. This concept has perhaps been best characterized for nuclear receptor TFs, for which multiple PTM-dependent co-activator and co-repressor complexes have been identified (*e.g.* reviewed in ref. 4 and 5). It is currently unclear to what extent these higher-order protein interactions and resulting PTMs can affect DNA binding specificities and affinities. Given that TFs have often been observed to act both as activators and repressors,[4,6,7] it is possible that the associated DNA binding complex influences DNA binding specificity or affinity, for example to distinguish genes that need to be repressed from those that need to be activated. This could occur through modulation not only of the TF DNA binding domain,[8] but also of regions located outside the DNA binding domain that can alter DNA recognition and affinity through protein domain intercommunication.[9–11] In recent years however, there has been increasing evidence for the reverse notion of the DNA dictating complex formation rather than the DNA binding complex differentiating between gene targets.[12–15] In other words, while the implicated TF(s) is still responsible for target gene identification, the nucleotide composition of the respective TF binding site allosterically influences co-regulator recruitment and thus whether the resulting DNA binding complex will activate or repress gene expression (Fig. 1). So far, this phenomenon has been elucidated for only a couple of TFs, including Oct-1,[15] NFκB[13] and glucocorticoid receptor,[14] but it is possible that this regulatory principle extends to many, if not most, other TFs. To validate this, it will be important to perform a comprehensive analysis of transcriptional complexes while bound to DNA, ideally without using protein-specific antibodies as this would significantly limit the experimental scope since such antibodies are available for only a low number of TFs and co-regulators. This DNA-centered approach to transcriptional regulation would also allow assessment of the dynamic properties of these complexes, as the same "DNA bait" could be used in distinct biological contexts. With the rapidly growing amount of experimentally defined regulatory element data (*e.g.* the Fantom and Encode consortia[16,17]), there is a wealth of suitable DNA bait candidates which could provide instrumental insights into gene regulatory mechanisms. The use of actual regulatory elements as DNA baits rather than TF binding site-representing double-stranded oligonucleotides may thereby be more informative given the often relatively poor correlation between *in vitro*-derived TF binding sites and *in vivo*-observed binding events.[18]

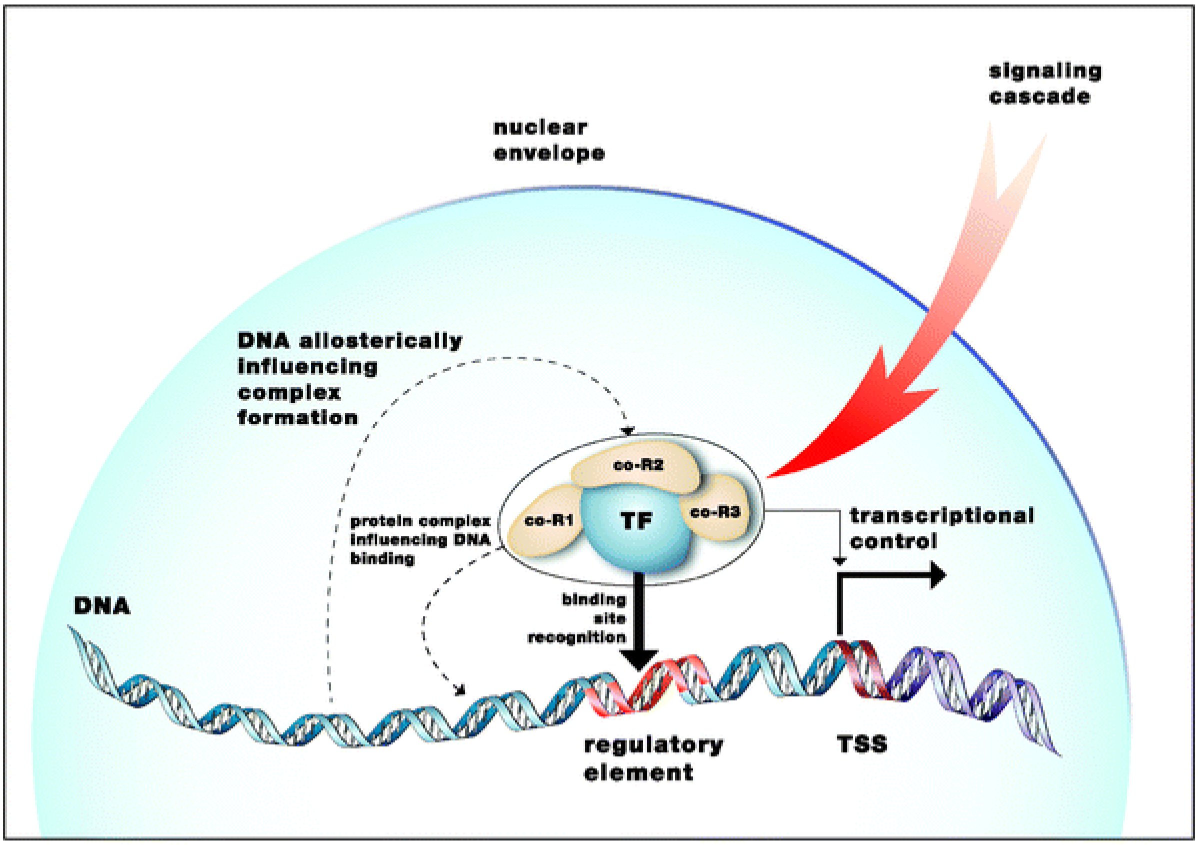


Fig. 1 Drawing illustrating the complex interplay between TF binding site recognition and DNA-binding protein complex formation at a specific regulatory locus.

The practical realization of such DNA-centered analyses has so far been difficult, as there are inherent difficulties associated with studying TF function such as their low expression and involvement in many transient and context-dependent interactions. Nevertheless, in recent years, important experimental progress has been made, which promises to significantly improve our ability to study the dynamic properties of transcriptional complexes in a DNA-centered fashion. Here, we provide a critical overview of these advances by highlighting their technical improvements over previously available DNA-centered methods and by pinpointing the remaining limitations. In addition, we briefly compare their output against protein-centered DNA binding complex detection methods, and highlight the advantages and disadvantages of each strategy.

DNA AFFINITY CHROMATOGRAPHY

GENERAL CONCEPT

The most familiar DNA-centered method is DNA affinity chromatography. This approach to study TF-containing complexes is based on DNA bait-mediated protein purification, which is achieved by exploiting the inherent capacity of TFs to bind to DNA. DNA is thereby either absorbed or linked covalently to a chromatographic support before being used for DNA affinity chromatography. Originally, heterogeneous non-specific DNA such as salmon or herring DNA was linked to a cellulose or sepharose chromatographic support (reviewed in ref. 19). This approach is not optimal for the purification of specific DNA binding complexes because of the abundant prevalence of contaminant proteins which non-specifically bind either to the support material or to the DNA. Rather, this approach is now commonly used as one of the many steps involved in TF purification as it efficiently removes contaminating proteins from complex protein mixtures. To subsequently isolate selected TFs or DNA binding complexes, specific DNA sequences are preferred. These are typically double-stranded oligonucleotides either in single-copy or concatemerized format, which represent TF consensus binding sites[20] or very small DNA regions with known DNA binding function identified, for example, through DNAse I footprinting.[21] Since TFs have an affinity several orders of magnitude greater for their consensus binding site sequence compared to non-specific DNA, the use of TF-specific double-stranded oligonucleotides allows a relatively straightforward purification of the respective TF and associated proteins from complex protein mixtures. However, this approach has also important limitations. First, while binding site concatemerization has been the preferred format for TF purification, it also introduces novel DNA sites, increasing the probability that other proteins will bind to the DNA bait and thus reducing purity.[19] Second, DNA binding and complex assembly occurs *in vitro* and since also stringent washing is required, only proteins that bind with high affinity to the respective TF will be retained, making this approach not ideal to study the dynamic properties of DNA binding complexes within their endogenous context. Third, it requires prior knowledge of specific TF binding sites. For a large number of TFs such corresponding binding sites are still unavailable, limiting the scope of this method.[1] Consequently, the approach is TF-centered, and will therefore not provide a comprehensive view of the factors controlling the transcription of your gene of interest. Fourth, such DNA bait typically represents just one of the possible binding site possibilities. This is important given the *in vivo* observation in both prokaryotic[22] and eukaryotic systems[1] that TFs also bind to sub-optimal binding sites. Thus, while the use of high affinity binding sites will provide a significant insight into DNA-binding protein complexes involving the respective TF, it will be by no means comprehensive. Moreover, it is becoming increasingly clear that multiple binding sites with varying TF affinities can cooperate, for example through DNA looping, to stabilize the TF-containing complex.[23] Thus, individual sites would fail to capture this complex. Finally, double-stranded oligonucleotides or other short DNA fragments usually fail to preserve the same DNA topology as that of the endogenously occurring TF binding site, which has also been shown to affect DNA binding. For example, p53 binding to DNA was enhanced with increasing negative superhelix density.[24]

RECENT ADVANCES

In recent years, there have been attempts to overcome many limitations by considering the use of single regulatory elements (or at least short fragments thereof) such as enhancers or promoters as DNA baits.[25–28] This approach does not require *a priori* knowledge of TF binding site properties, and since elements are typically linked to a specific gene, it also provides information of immediate relevance to how the respective gene may be transcriptionally controlled. Recent examples of the DNA affinity chromatography approach include the isolation of a *Drosophila* TF, DEAF-1, binding to the enhancer of an immunity gene,[26] as well as several proteins binding to promoter fragments of, respectively, the human *ESRRA* and *MTA*2 genes.[27] In both studies, DNA was immobilized onto a solid phase by biotin-labelling the DNA and coupling it to either streptavidin-coated columns or magnetic beads. This is in contrast with the technique of DNA trapping used by Jiang and co-workers[29] in which a 250 bp region of the human *c-jun* promoter with a single stranded $(GT)_5$ tail was annealed to single-stranded $(AC)_5$-Sepharose. Several pre-initiation components such as RNA polymerase II, TBP, and the TFIF subunit RAP 74 were retrieved as well as the TF SP1. DNA trapping typically supports better purification as streptavidin-coated supports are known to bind to contaminating proteins (Table 1). In addition, DNA trapping allows a non-denaturating elution of bound proteins,[29] which, because of the strength of the interaction, is not possible with DNA immobilization where high temperatures and denaturating agents such as SDS are required to elute proteins. On the other hand, DNA immobilization is less time consuming and amenable to automatization due to its relatively simple workflow, and is therefore often the method of choice. As indicated above, a critical aspect of both techniques is the use of competitor DNA such as salmon sperm DNA, poly(dI:dC) or scrambled bait DNA to eliminate proteins that have low affinity for the DNA bait but would otherwise be retained because of the high concentration of bait DNA.

Table 1 Summary of the strengths (+) and weaknesses (−) of the discussed DNA-centered methods to characterize regulatory protein–DNA interaction complexes

			Time-consuming	Unambiguous protein purification/identification	Unambiguous complex purification/identification	Prone to artificial DNA binding	Unbiased (truly DNA-centered) interaction screen	Protein complex characterization in its natural context
Homogeneous DNA				–	–	–		–
Heterogeneous DNA	Short DNA sequences as DNA bait	TF-specific double-stranded oligonucleotide trapping	–	+	–			–
		TF-specific double-stranded oligonucleotide immobilization		+	–			–
		Concatemers		+	–	–		–
	Long DNA sequences as DNA bait	DNA trapping	–	+	+		+	–
		DNA immobilization		+	+		+	–
PICh			–	+	+		+	+
Supershift (Electrophoretic Mobility Super Shift Assay)			–	+	–	–	–	–
High-throughput yeast one-hybrid				+	–	–		–

DNA-BINDING PROTEIN IDENTIFICATION

ANTIBODIES

The most straightforward and therefore widespread strategy to identify captured TFs or TF-containing protein complex members is based on the use of antibodies. Their application is thereby not restricted to DNA affinity chromatography,[28] as other protein–DNA interaction detection approaches also benefit greatly from antibody availability. An excellent example is the supershift assay in which the identity of a DNA-binding protein is confirmed only when a protein-specific antibody reduces the electrophoretic mobility of a protein–DNA interaction complex.[30] A significant advantage of such a gel shift procedure over other protein–DNA interaction detection methods is its ability to distinguish single from multimeric forms of bound protein and to immediately relate this information to the respective DNA bait. For example, using supershift assays, Tantin and colleagues[31] determined that DNA baits were more likely to induce di- or multimerization of the TF Oct-4 when they contained at least three Oct-4 half sites. Thus, and as discussed already above, binding site cooperativity can influence the formation of distinct TF complex configurations, with each possibly having a differential impact on how the respective target gene is transcriptionally controlled. Nevertheless, despite their utility, antibodies restrict the scope of the assay as only a limited number of highly specific DNA-binding protein antibodies are currently available. Moreover, antibody implementation requires an *a priori* assumption about the identity of interacting proteins, making this approach protein-centered. Thus, while several protein-centered methods have already contributed in significant fashion to our understanding of the molecular mechanisms underlying protein–DNA interactions *in vitro* and *in vivo* (reviewed extensively in ref. 1), we will not discuss them here given this review's focus on DNA-centered protein–DNA interaction approaches. Instead, we will briefly discuss new efforts to eliminate the protein-centered bias of current DNA bait-based techniques such as gel shift and DNA affinity chromatography by linking them to *de novo* protein detection and identification methods such as two-dimensional gel electrophoresis (2-DE)[32] and mass spectrometry.[33] Since it is desirable still to confirm the identity of 2-DE-detected proteins using mass spectrometry, we will briefly focus on the latter technology.

MASS SPECTROMETRY

Traditionally, the detection and identification of DNA-binding proteins or complexes by mass spectrometry has always been difficult owing to the low cellular abundance of the majority of these types of proteins.[34] In recent years, mass spectrometry has become increasingly sensitive, driven by fast-paced technological advances in instrumentation. This significant increase in mass accuracy and resolving power now allows for the first time a more detailed functional analysis of such lowly expressed proteins as their spectral peaks become increasingly distinguishable from background noise in complex mixtures. Based on overall sensitivity, two mass analyzers stand out. The first is the

Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR).[35] The second is the Orbitrap, which also uses an FT-based strategy.[36] For example, Mann and co-workers[27] have used an FT-ICR to identify sequence-specific DNA-binding proteins in HeLa S3 cells purified by DNA affinity chromatography. Interestingly, when comparing the eluted protein SDS-PAGE profiles from the wild-type and negative control DNA bait, there was virtually no difference and thus no clear bands were revealed corresponding to true specific DNA-binding proteins. Nonetheless, because of the sensitivity of FT-ICR, candidate DNA-binding proteins that were more abundant in the wild-type *versus* negative control samples were in the end identified (see also below). Protein-centered approaches aiming to characterize TF-specific protein interaction partners or complexes are also benefiting greatly from the recent sensitivity increase as evidenced by the fact that many of the detected TF interactors were themselves TFs.[37,38] Thus, we are entering an exciting era in which proteins such as TFs that have traditionally been for the most part off-limit become increasingly accessible and thus characterizable.

DNA-BINDING DYNAMICS

To achieve a comprehensive, mechanistic understanding of gene regulation, it is essential to not only determine the identity of regulatory DNA-binding complex members, but to also chart compositional complex changes in relation to alterations in target gene expression. This need to monitor protein complex assembly dynamics either with other proteins or with DNA has prompted the development of quantitative proteomics approaches (*e.g.* reviewed in ref. 39). The latter involve the labelling of proteins with isotopically distinguishable tags enabling a protein abundance comparison between two or more biological samples. Brand *et al.*[37] used isotope-coded affinity tagging (ICAT) to monitor the compositional changes of the protein complex involving the TF NF-E2p18/MafK during erythroid differentiation. Results uncovered more than 100 potential protein interactors and indicated that MafK acts as a dual-function TF, exchanging dimerization partners upon induction of differentiation, leading to the replacement of interacting co-repressors with co-activators and up-regulation of the expression of its target gene β-globin. To answer questions related to the molecular mechanisms underlying this protein partner exchange, the next step would be to monitor the compositional changes of only those MafK-containing complexes that are bound to DNA. Although this is in principle feasible by monitoring the DNA occupation of individual complex members at distinct time points using chromatin immunoprecipitation,[37,40] the unavailability of antibodies for the majority of proteins limits, as indicated previously, the scope of such assays and thus prevents a functional analysis of the majority of detected protein interaction partners. Moreover, similar to other recent mass spectrometry-based TF-protein interaction detection techniques such as the streptavidin-mediated isolation of biotinylated TF complexes,[38] the approach used by Brand and colleagues[37] is again strictly TF or protein-centered and may therefore miss crucial factors that may influence MafK complex assembly on the DNA and thus β-globin gene regulation in general without physically interacting with MafK, but for example by altering DNA accessibility.[41] With this experimental mindset, Mittler *et al.*[27] combined the "stable isotope labelling with amino acids in cell culture" (SILAC) technique with DNA affinity chromatography to detect protein–DNA complex assembly differences on wild-type *versus* mutated TF binding sites or short regulatory element fragments. For both types of DNA baits, a significant number of putative binding proteins were found. Since most of these were captured in approximately equal amounts by the wild-type and negative control bait, they could however be eliminated. The identity of the remaining proteins was in line with predictions and proved the value of their method. In addition, these researchers were able to identify proteins, many of which were previously not described, that preferentially bind to methylated *versus* non-methylated CpG sites on the *MTA*2 gene promoter. While the latter method clearly increases our ability to determine TF binding profiles, it still suffers from the previously mentioned important limitation that DNA binding and putative complex assembly occurs *in vitro* and thus information regarding the complex composition at the corresponding endogenous locus is lost.

ALTERNATIVE DNA-CENTERED APPROACHES

PICH

To enable the *in vivo* assessment of regulatory DNA-binding complexes at specific gene loci, Déjardin and Kingston[42] have now developed the PICh (Proteomics of Isolated Chromatin segments) method. This method is a drastic departure from previous methods typically based on DNA affinity chromatography as it is better described as a reverse chromatin immunoprecipitation, since it uses cross-linking to fix protein complexes on DNA, but rather than using a protein-specific antibody to identify bound DNA regions, it employs a DNA element-specific probe to pull down the associated protein complex (Fig. 2). The probe is an oligonucleotide containing locked nucleic acid (LNA) residues. These have an altered backbone that favours base stacking, thereby significantly increasing the stability of probe–DNA interactions. After the probe has hybridized to chromatin cross-linked to protein complexes, it is captured on streptavidin magnetic beads through a desthiobiotin molecule covalently linked to the probe. Desthiobiotin is a biotin analog with weaker affinity for avidin which therefore permits a more gentle competitive elution using biotin, limiting the co-elution of non-specific factors. Thus, by maintaining the DNA-bound protein complex in its natural state and because probes can be designed against any locus, PICh provides in principle the possibility to correlate protein complex composition changes with alterations in the expression of any target gene. However, PICh has so far only been used to detect proteins associated with human telomere sequences, which, with around 100 copies per cell, are rather abundant in the genome and therefore compensate for the relatively low protein detection sensitivity of the method. Evolving the method to allow screening of less abundant chromatin loci or even unique regulatory elements is now the next challenge and could involve a reconsideration of probe design, the use of even more sensitive mass spectrometers, or the integration of quantitative proteomics techniques.

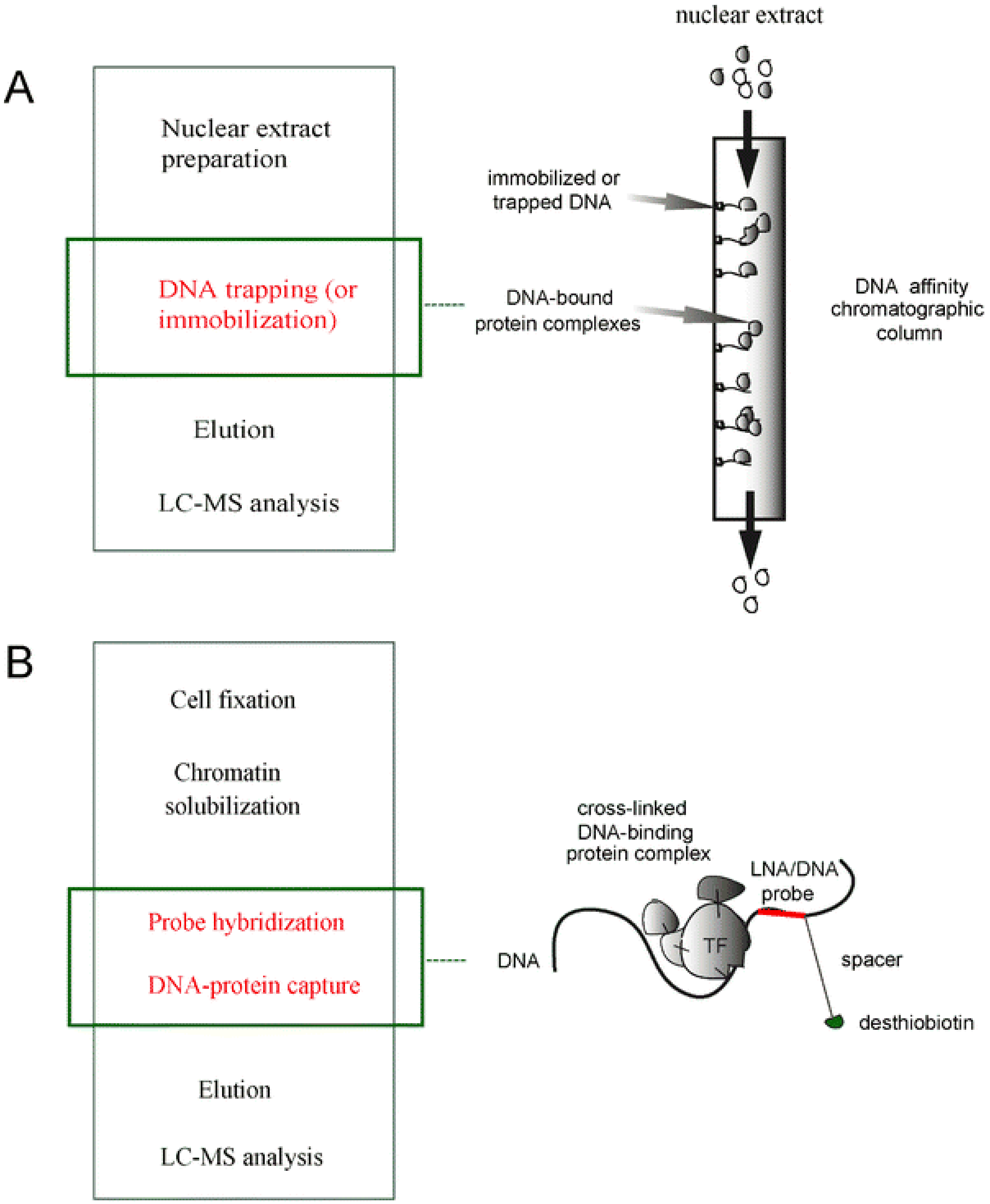

A
Nuclear extract preparation
DNA trapping (or immobilization)
Elution
LC-MS analysis
nuclear extract
immobilized or trapped DNA
DNA-bound protein complexes
DNA affinity chromatographic column
B
Cell fixation
Chromatin solubilization
Probe hybridization
DNA-protein capture
Elution
LC-MS analysis
cross-linked DNA-binding protein complex
LNA/DNA probe
DNA
TF
spacer
desthiobiotin

Fig. 2 Drawing illustrating the protein complex purification workflow using DNA affinity chromatography (A) and PICh (B).

HIGH-THROUGHPUT YEAST ONE-HYBRID ASSAY

While PICh has the promise to revolutionize the gene regulation field, it remains to be seen how much the method will live up to expectations. Consequently, DNA affinity chromatography as well as other alternative methods will remain useful to study the dynamic properties of regulatory DNA-binding complexes in DNA-centered fashion. One other alternative method is the high-throughput yeast one-hybrid system, which allows the screening of regulatory elements of interest for interacting TFs or TF dimers.[43–45] Although the latter technique does not allow the detection of DNA-binding complexes and is performed in yeast and thus outside the endogenous context, it provides the unique possibility to scan the whole regulatory protein repertoire for binding to a DNA bait of choice depending on the completeness of the screened TF library.[46,47]

CONCLUSION

In recent years, there has been renewed interest in obtaining a complete understanding of the complex mechanisms underlying gene regulation, driven by recent discoveries illustrating the complex interplay between all components involved (DNA, TFs, co-regulators *etc*.) in guiding the formation of functional regulatory complexes, which either activate or repress gene expression. Consequently, there are revived efforts to improve current technologies to enable an increasingly more accurate and comprehensive study of gene regulatory complex formation. Specifically, there is a significant need to monitor the formation of such complexes at defined regulatory loci in distinct biological contexts, hence the renewed interest in DNA-centered protein–DNA interaction detection technologies. Although progress in this area has been made as discussed in this review, we are still far from the complete and functional characterization of DNA-binding protein complexes and from the ability to relate changes in their composition to expression changes of their respective target genes. In this regard, we are eagerly looking forward to novel developments in the *in vivo* quantitative proteomics field, to improved TF and protein complex purification methods, and to further increases in the sensitivity of mass spectrometers, which are quickly becoming the “gold standard” in the analysis of DNA-binding functional protein complexes.

ABBREVIATIONS

TF	Transcription Factor
PTM	Post-Translational Modification
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
2-DE	Two-dimensional Gel Electrophoresis
FT	Fourier Transform
FT-ICR	Fourier Transform Ion Cyclotron Resonance
ICAT	Isotope-Coded Affinity Tags
SILAC	Stable Isotope Labelling with Amino Acids in Cell Culture
PICh	Proteomics of Isolated Chromatin Segments
LNA	Locked Nucleic Acid
Co-R	Co-regulatory element

ACKNOWLEDGEMENTS

We thank the members of the Deplancke lab for reading of the manuscript, and Michele Zinnai for his significant help with the preparation of the figures. Funding is provided by the Swiss National Science Foundation, the SystemsX Initiative, and the European Union (IRG, FP7); institutional support from the Ecole Polytechnique Fédérale de Lausanne.

REFERENCES

1 B. Deplancke, *Briefings Funct. Genomics Proteomics*, 2009, **8**, 12–27.

2 S. H. Yang, E. Jaffray, R. T. Hay and A. D. Sharrocks, *Mol. Cell*, 2003, **12**, 63–74.

3 N. C. Reich and L. Liu, *Nat. Rev. Immunol.*, 2006, **6**, 602–612.

4 M. G. Rosenfeld, V. V. Lunyak and C. K. Glass, *Genes Dev.*, 2006, **20**, 1405–1428.

5 B. W. O'Malley, J. Qin and R. B. Lanz, *Curr. Opin. Cell Biol.*, 2008, **20**, 310–315.

6 N. Soontorngun, M. Larochelle, S. Drouin, F. Robert and B. Turcotte, *Mol. Cell. Biol.*, 2007, **27**, 7895–7905.

7 J. S. Reece-Hoyes, B. Deplancke, M. I. Barrasa, J. Hatzold, R. B. Smit, H. E. Arda, P. A. Pope, J. Gaudet, B. Conradt and A. J. M. Walhout, *Nucleic Acids Res.*, 2009, **37**, 3689–3698.

8 C. W. Garvie, J. Hagman and C. Wolberger, *Mol. Cell*, 2001, **8**, 1267–1276.

9 V. Chandra, P. Huang, Y. Hamuro, S. Raghuram, Y. Wang, T. P. Burris and F. Rastinejad, *Nature*, 2008, 350–356.

10 H. Gronemeyer and W. Bourguet, *Sci. Signal.*, 2009, **2**, pe34-.

11 M. A. Pufall, G. M. Lee, M. L. Nelson, H.-S. Kang, A. Velyvis, L. E. Kay, L. P. McIntosh and B. J. Graves, *Science*, 2005, **309**, 142–145.

12 N. Ha, K. Hellauer and B. Turcotte, *Nucleic Acids Res.*, 1996, **24**, 1453–1459.

13 T. H. Leung, A. Hoffmann and D. Baltimore, *Cell*, 2004, **118**, 453–464.

14 S. H. Meijsing, M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto, *Science*, 2009, **324**, 407–410.

15 A. Tomilin, A. Reményi, K. Lins, H. Bak, S. Leidel, G. Vriend, M. Wilmanns and H. R. Schöler, *Cell*, 2000, **103**, 853–864.

16 The FANTOM Consortium, *Science*, 2005, **309**, 1559–1563.

17 The ENCODE Project Consortium, E. Birney, J. Stamatoyannopoulos, A. Dutta, R. Guigo, T. Gingeras, E. Margulies, Z. Weng, M. Snyder, E. Dermitzakis, R. Thurman, M. Kuehn, C. Taylor, S. Neph, C. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. Greenbaum, R. Andrews, P. Flicek, P. Boyle, H. Cao, N. Carter, G. Clelland, S. Davis, N. Day, P. Dhami, S. Dillon and M. Dorschner, *Nature*, 2007, **447**, 799–816.

18 C. E. Massie and I. G. Mills, *EMBO Rep.*, 2008, **9**, 337–343.

19 H. Gadgil, L. A. Jurado and H. W. Jarrett, *Anal. Biochem.*, 2001, **290**, 147–178.

20 M. Yaneva and P. Tempst, *Anal. Chem.*, 2003, **75**, 6437–6448.

21 E. Ruteshouser and B. de Crombrugghe, *J. Biol. Chem.*, 1992, **267**, 14398–14404.

22 D. C. Grainger and S. J. W. Busby, *Biochem. Soc. Trans.*, 2008, **036**, 754–757.

23 L. Saiz and J. M. G. Vilar, *Curr. Opin. Struct. Biol.*, 2006, **16**, 344–350.

24 E. B. Jagelská, V. c. Brazda, P. Pecinka, E. Palecek and M. Fojta, *Biochem. J.*, 2008, **412**, 57–63.

25 N. Ahmad and J. B. Lingrel, *Biochemistry*, 2005, **44**, 6276–6285.

26 D. E. Reed, X. M. Huang, J. A. Wohlschlegel, M. S. Levine and K. Senger, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 8351–8356.

27 G. Mittler, F. Butter and M. Mann, *Genome Res.*, 2009, **19**, 284–293.

28 A. Krehan, H. Ansuini, O. Bocher, S. Grein, U. Wirkner and W. Pyerin, *J. Biol. Chem.*, 2000, **275**, 18327–18336.

29 D. Jiang, R. A. Moxley and H. W. Jarrett, *J. Chromatogr., A*, 2006, **1133**, 83–94.

30 L. M. Hellman and M. G. Fried, *Nat. Protoc.*, 2007, **2**, 1849–1861.

31 D. Tantin, M. Gemberling, C. Callister and W. Fairbrother, *Genome Res.*, 2008, **18**, 631–639.

32 J. A. Stead, J. N. Keen and K. J. McDowall, *Mol. Cell. Proteomics*, 2006, **5**, 1697–1702.

33 E. Nordhoff, A.-M. Krogsdam, H. F. Jorgensen, B. H. Kallipolitis, B. F. C. Clark, P. Roepstorff and K. Kristiansen, *Nat. Biotechnol.*, 1999, **17**, 884–888.

34 J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann and N. M. Luscombe, *Nat. Rev. Genet.*, 2009, **10**, 252–263.

35 M. B. Comisarow and A. G. Marshall, *J. Mass Spectrom.*, 1996, **31**, 581–585.

36 A. Makarov, *Anal. Chem.*, 2000, **72**, 1156–1162.

37 M. Brand, J. A. Ranish, N. T. Kummer, J. Hamilton, K. Igarashi, C. Francastel, T. H. Chi, G. R. Crabtree, R. Aebersold and M. Groudine, *Nat. Struct. Mol. Biol.*, 2004, **11**, 73–80.

38 J. Wang, S. Rao, J. Chu, X. Shen, D. N. Levasseur, T. W. Theunissen and S. H. Orkin, *Nature*, 2006, **444**, 364–368.

39 A.-C. Gingras, M. Gstaiger, B. Raught and R. Aebersold, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 645–654.

40 J. Sun, M. Brand, Y. Zenke, S. Tashiro, M. Groudine and K. Igarashi, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1461–1466.

41 E. Segal and J. Widom, *Nat. Rev. Genet.*, 2009, **10**, 443–456.

42 J. Déjardin and R. E. Kingston, *Cell*, 2009, **136**, 175–186.

43 B. Deplancke, A. Mukhopadhyay, W. Ao, A. M. Elewa, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, J. S. Reece-Hoyes, I. A. Hope, H. A. Tissenbaum, S. E. Mango and A. J. Walhout, *Cell*, 2006, **125**, 1193–1205.

44 B. Deplancke, D. Dupuy, M. Vidal and A. J. Walhout, *Genome Res.*, 2004, **14**, 2093–2101.

45 G. Chen and J. A. Shin, *Anal. Biochem.*, 2008, **382**, 101–106.

46 J. S. Reece-Hoyes, B. Deplancke, J. Shingles, C. A. Grove, I. A. Hope and A. J. Walhout, *Genome Biol.*, 2005, **6**, R110.

47 V. Vermeirssen, B. Deplancke, M. I. Barrasa, J. S. Reece-Hoyes, H. E. Arda, C. A. Grove, N. J. Martinez, R. Sequerra, L. Doucette-Stamm, M. R. Brent and A. J. Walhout, *Nat. Methods*, 2007, **4**, 659–664.

Footnote

†This article is part of a *Molecular BioSystems* themed issue on Computational and Systems Biology.

II. RESULTS

II.I DEVELOPMENT OF A GENE-CENTERED PROTEOMIC PLATFORM FOR THE SYSTEMATIC IDENTIFICATION OF DNA BINDING PROTEINS AND COMPLEXES

The purpose of this project was to adapt a proteomic platform to capture DNA-binding proteins or complexes in a systematic and quantitative manner within their functional context *in vivo*. Contrary to most protein-DNA interaction techniques that are used today, the platform is gene-centered, in that it takes advantage of the growing amount of experimentally defined transcriptionally active DNA element data to identify associated DNA binding proteins or complexes. By incorporating a stable metabolic labelling technique (SILAC), it thereby aims to provide quantitative information regarding the binding behaviour of these DNA binding proteins or complexes. The information collected was intended to be used to build a comprehensive protein-DNA interaction map that elucidates the dynamic and quantitative aspects of gene regulation.

Development of a gene-centered proteomic platform for the systematic identification of DNA binding proteins and complexes

MOTIVATION

Gene expression is a complex mechanism that requires the input of several signalling cascades involving a large number of components. The components of these cascades have been extensively characterized, providing valuable information on how living cells are able to respond to their environment [1, 2]. Signalling ultimately results in differential gene expression. The interface between these signalling pathways and the protein layer that mediates gene regulation, the so-called gene regulatory networks, has however been poorly characterized because many of the implicated regulatory proteins such as transcription factors (TFs) and co-regulators tend to be present at very low concentrations in cells, making the study of this important class of proteins with today's technologies quite challenging (figure 1). It is our major goal to understand in deeper detail the mechanisms that dynamically and quantitatively govern gene expression within their native context. To achieve this goal, we need to comprehensively characterize the regulatory protein interface between signalling pathways and the genome and therefore intend to develop a technology that enables us to monitor TF and co-regulator changes in a systematic manner.

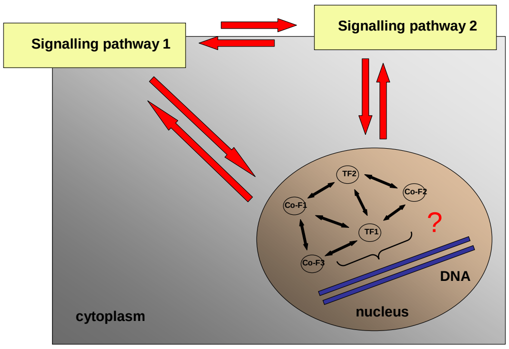
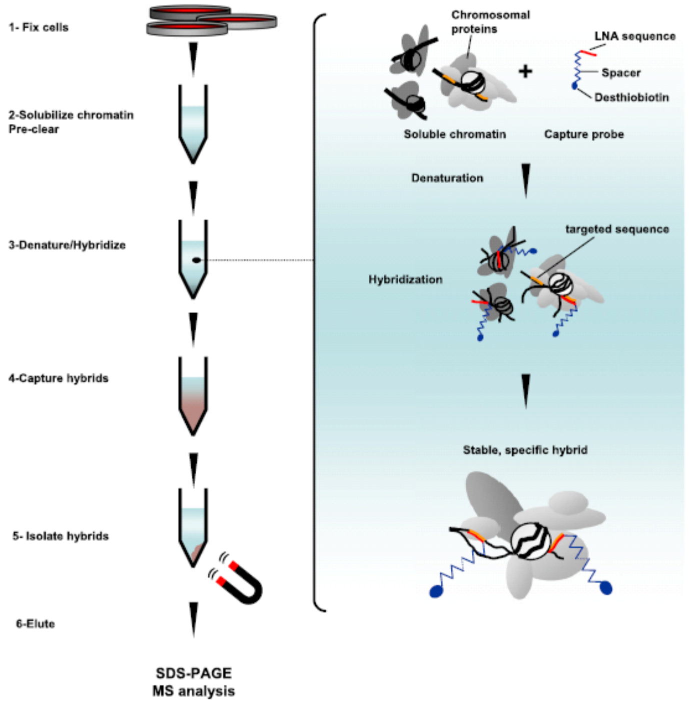


Figure 1. Model of gene regulation emphasizing the relative lack of knowledge regarding the composition and dynamic properties of the regulatory protein interface indicated by a question mark.

The recent outburst of genomic information has spurred the development of novel methods to analyze and characterize the protein and gene components of gene regulatory networks and to understand how they interact at the systems level. To date, most of these methods are TF-centered, in that they use TFs to retrieve DNA sequences bound by them [3, 4]. Chromatin immunoprecipitation (ChIP) is a powerful technology to assess whether a protein of interest is bound to a given genomic region "in vivo" [5]. Yet ChIP relies heavily on the use of antibodies and thus is limited to analysis of the factors that have previously been tested, and unfortunately does not establish a comprehensive description of protein complex composition. Moreover, TF-centered methods do not take advantage of the growing amount of experimentally defined regulatory element data that has recently been generated, for example as part of the Encyclopedia of DNA elements Project (ENCODE, [6]), or other genome-wide regulatory element mapping projects [7, 8].

To address these limitations and gain insight into locus-specific protein complex composition, a gene-centered strategy would be required enabling the purification of an endogenous segment of chromatin, e.g. a regulatory element of interest, in sufficient quantity and purity to identify the in vivo associated proteins. Such a technology would permit a detailed correlation between composition at a locus and phenotype, leading to a deeper understanding of gene regulation mechanisms. Mann and colleagues developed an "in vitro" gene-centered quantitative interaction screen that utilizes short promoter fragments (wild-type (WT) versus mutated) to analyse sequence specificity for the capture of TFs [9]. Since this methodology uses a very short DNA "bait" sequence, it allows for the characterization of only a small subset of DNA binding proteins, thereby failing to reflect the appropriate cellular endogenous context.

Recently, a new technique called PICh (Proteomics of Isolated Chromatin segments) was developed, in which specific nucleic acid probes are used to target specific sequences within genomic DNA [10] (figure 2). In doing so, PICh enables the specific isolation of sequence specific DNA binding complexes, extracted directly from live cells, thereby maintaining the cellular environment as close as possible to the physiological context ("in vivo"). The identification of such complexes is performed by mass spectrometry (MS). To date, PICh has only been used to detect proteins associated with telomere sequences, which are present in multiple copies within cells. We aim to adapt the technique to the much more challenging task of screening individual DNA sequences, such as enhancers or promoters for interacting protein complexes. This is challenging because such unique sequences are present only twice in diploid cells. We therefore need to increase significantly the sensitivity of PICh to permit the characterization of single copy regulatory sequences bound complexes.



(Déjardin & Kingston; Cell, 2009)

Figure 2. Workflow of the PICh protocol

In order to understand dynamic aspects of gene regulation, we also aim to monitor protein complex-DNA interaction phenomena in biological samples showing differential expression of a particular gene of interest. To optimize the PICh assay for our purposes, we have chosen a well-studied gene regulation model, the β-globin cluster, which will be explained in the next section.

THE MODEL

The β-globin gene family is probably one of the best studied families of genes. Several TFs and transcriptional complexes have already been identified as modulators of globin gene expression, particularly in mouse and human, and the corresponding gene regulatory elements have been well characterized (reviewed in [11]). In addition, these genes are tissue-specific and developmentally regulated. The β-globin gene cluster is located on chromosome 11 and consists of five functional genes, arranged from 5' to 3' reflecting their developmental sequential expression. ε-globin is normally expressed in the embryonic yolk sac while A-gamma and G-gamma globin are expressed only during fetal development. Around birth, the production of γ-globin declines whereas the production of β-globin sharply increases. The combination of two alpha genes and two beta genes comprises the normal adult haemoglobin.

Located 6-20 kb 5´ of the human ε-globin gene, a series of erythroid-specific DNase I hypersensitive sites (HSs), each of 200 to 400 bp in size, aid in controlling the expression of downstream cis-linked globin genes. These five sites form the globin locus control region (LCR) which can up-regulate or down-regulate expression in a cooperative manner. HSII lies some 11 kb 5´ of the ε-globin gene, and is well known to stimulate the transcription of embryonic as well as fetal globin genes in erythroid cell [12]. How HSII interacts with distant cis-linked globin promoter sequences to stimulate expression at different developmental stage is not fully understood. What has been proposed is that complex mechanisms of looping and tracking take place during globin gene expression [13] [14]. The ε-globin gene is situated 11 kb 3´ of HSII and its transcription is stimulated by the enhancing capabilities of this particular site. The core of the ε-globin promoter is about 200 bp long. Many TFs and co-factors are known to interact with the HSII element of the LCR (e.g. GATA1, NF-E2, Sp1) and with the promoter of the embryonic ε-globin gene (e.g. AP1, NF-E1)(figure 3). Figure 3 shows the overall organization of the human β-globin gene locus, including the LCR and the five globin genes, including DNA binding motifs (circles), TFs (squares), and co-factors (triangles). TFs, by binding to specific motifs through selective recruitment of co-factors, can either activate or repress gene expression [15, 16]. We therefore aim to identify TFs and co-regulators that are already known to bind to β-globin cis-linked regulatory sequences, namely to the 5 sites in the LCR and to β-globin promoters, and complement such knowledge with newly identified interactors, including other TFs or co-regulators.

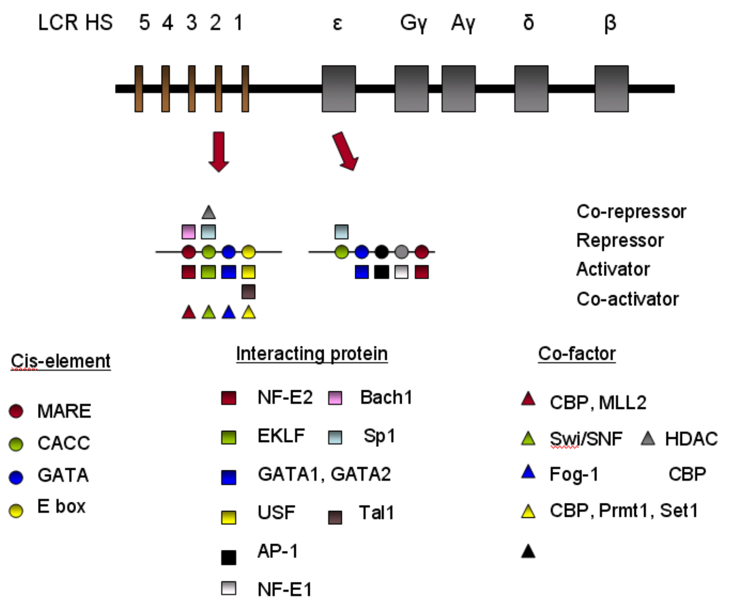


Figure 3. Summary of TFs and co-regulators known to interact with the LCR element HSII and with the embryonic ε-globin promoter.

Importantly, human cell lines in which globin genes are dynamically expressed are easily available, providing a relatively easy-to-use model system to study globin gene regulatory mechanisms (e.g. K562 cells, [17]). The K562 is an immortalized human myelogenous leukaemia cell line that was derived from a 53 year old patient in blast crisis [18, 19]. Being an immortalized myeloid leukemia line, K562 cells do not show a classical "erythroid" behaviour; instead they present mixed characteristics of early-stage erythrocytes [20], granulocytes [21] and monocytes [22]. K562 cells constitutively express embryonic and fetal globin genes (epsilon and gamma), but not β-globin [23]. Globin expression can be increased up to ten fold by exposing K562 cells to inducing agents such as hemin [24], hydroxyurea, and sodium butyrate.

Thus, the 5 elements known to mediate globin expression as well as promoter regions will be selected and used as baits to capture interacting proteins or complexes in conditions where β-globin like genes are expressed (K562) versus silenced (here the HEK293 cell line will be used as a negative control as globin genes are not expressed in these cells), revealing specific patterns of regulation. These experiments should yield sufficient information to generate crude background protein profiles and thus enable the elimination of false positive proteins from the analysis by intelligently discarding those proteins that are constantly present throughout different experimental settings.

As mentioned in the introduction, the PICh technique (figure 2) was recently developed to characterize protein composition on specific loci within the chromatin in a functional “in vivo” context. Interestingly, the PICh technique can be applied on a high-throughput set up, due to the fact that the LC (liquid chromatography)-MS steps are amenable to automatization (e.g. MudPIT (multidimensional protein identification technology) [25]).

For the optimization of hybridization to the target sequences, the chromatin capturing probe is composed of mixed LNA/DNA oligonucleotides. Locked nucleic acid (LNA) residues (figure 4) have an altered backbone that favors base stacking, thereby significantly increasing the stability of probe-chromatin interactions [26].

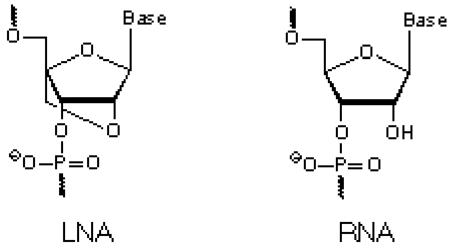


Figure 4. Comparison of LNA and RNA nucleotides

A desthiobiotin molecule is covalently linked to the 5´ of the oligonucleotide thorough a long spacer. The use of desthiobiotin (a biotin analog with weaker affinity for avidin) is justified by the fact that the structure of the molecule permits a competitive gentle elution using biotin [27], limiting the co-elution of non-specific factors.

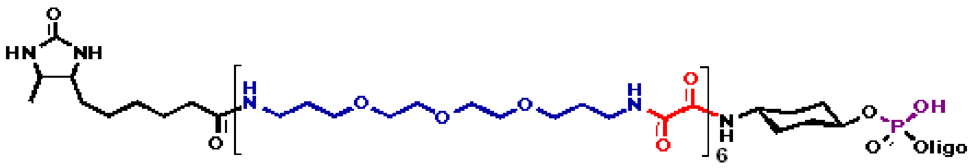


Figure 5. The chemical structure of the molecular probe. A desthiobiotin moiety is covalently linked to a long phosphoramidate spacer linked to the 5´ end of a 25mer oligo.

To minimize the steric hindrance (which is detrimental for yields) observed upon immobilization of chromatin [28, 29], a very long spacer has to be placed between the immobilization tag and the LNA/DNA probe (figure 5). It has been previously shown that tether lengths of at least 40 atoms for surface-bound oligonucleotides are crucial for optimum target hybridization. It is also possible to make longer linkers by adding spacers one after the other to the end of the oligonucleotide prior to adding a terminal desthiobiotin. Excessively long tethers will have the adverse effect of interfering with hybridization [30].

Thus, we plan to adapt the PICh technology to target single copy regulatory sequences using the above mentioned chromatin capturing probe technology. Those probes have to be highly specific to avoid hybridization to other loci in the genome leading to an incorrect interpretation of the results.

TF COMPLEX DETECTION

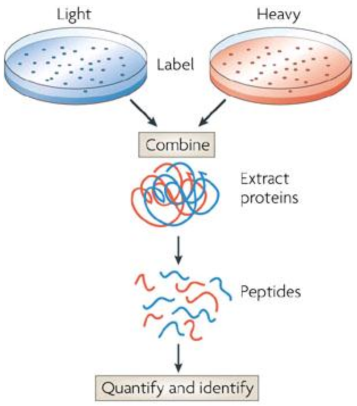
One of the larger difficulties associated with protein identification and quantitation is the co-purification of often abundant “sticky” proteins unrelated to the biological process of study. These “false positives” may obscure the correct interpretation of the proteomic data [31]. It is possible to overcome this problem via the systematic analysis and comparison of protein profiles obtained after each DNA affinity chromatography-based purification. Thus, by screening several DNA baits, a DNA affinity background protein profile could be generated. Non-specific proteins can then be systematically discarded allowing the identification of proteins that are relatively unique to the DNA sequence of interest. It is clear that this strategy will only succeed if the protein detection method of choice delivers data in a standardized format and therefore supports direct comparison of protein profiles between different assays.

The recent transition to spectrum-based protein identification (in addition to sequence-based identification) may in this context be highly beneficial. Indeed, spectral library searching has recently shown promising results, in terms of computational time, identification rate and accuracy, [32]. Therefore, high quality peptide matches obtained from a sequence search approach can be extracted from a first set of experiments to build an appropriate spectral library. Run after run, tandem mass spectra can be screened against the library, and only spectra with no interpretation would be submitted to a sequence search engine. By generating a spectral library “on the fly”, either from search results from a previous experiment where a similar sample preparation process was employed, or from a fraction of the samples to be analyzed, a significant percentage of the spectra can be rapidly and unambiguously identified. As mentioned in the introduction, the low abundance of TFs is a crucial limiting factor. Generating a spectral library may not be sufficient to clearly identify TF peptides due to the fact that TF peaks may be hidden in the background noise or the resulting spectra may be too complicated to be interpreted. To address this issue, we also intend to develop a proteotypic peptide library, in collaboration with the laboratory of Biomolecular Mass Spectrometry of Prof. Tsybin, to ease the bioinformatic burden of TF identification by furnishing clean spectra belonging to purified TFs.

Thus, we plan to detect loci specific protein complexes (captured by PICh) by filtering out non-specific interactors in a multiple screening process. Spectrum-based protein identification will be facilitated by the availability of a TF proteotypic peptide collection of spectra.

COMPARATIVE PROTEOMICS

An important novel aspect of this project is our goal of monitoring protein expression changes within our models. To render our approach quantitative, a biological differential labelling technique named SILAC (Stable Isotope Labelling of Amino acids in Cell cultures) will be combined with the PICh method [33]. SILAC will enable relative quantitation of DNA-binding protein complexes between different biological states, such as “expressed” versus “non-expressed”. In the simplest version of this technique, two cells populations are differentially labelled by growing them separately in a light medium versus a medium enriched with heavy, non-radioactive isotopes such as ^{13}C and ^{15}N (often in either arginine or lysine). Proteins obtained from such cultures are then pooled, trypsinized, and subjected to LC-MS/MS (figure 6). Metabolic incorporation of heavy amino acids into the proteins results in a proportional mass shift of the corresponding peptides. This mass shift can be detected by mass spectrometers. Therefore, the relative protein abundance in each experimental condition can then be inferred by combining the two samples and comparing their peak intensities.



(Gingras et al.; Nature Reviews; 2007)

Figure 6. SILAC metabolic labelling

Thus, based on the differential expression of proteins of interest using SILAC, we will be able to better understand the dynamic aspects of gene regulation within specific biological processes.

IMPLEMENTATION

As previously described, the major aim of the project is to develop a consistent gene-centered proteomic platform to capture TFs and co-factors in the form of functional complexes in a systematic manner within their natural environment. The model we decided to use is the β-globin gene cluster in the K-562 cell line.

CAPTURE OF HSII AND E-GLOBIN PROMOTER BINDING PROTEINS

As already mentioned, the HSII site and the ε-globin promoter regulatory regions were chosen as a basis for the development of the PICh proteomic platform. We believe that exposed DNA sequences will facilitate the process of hybridization. However, it is possible that proteins bound to the target may render the DNA inaccessible, while tight packing of chromatin in heterochromatin by either methylation or acetylation may hinder the accessibility to the target sequence as well. On the other hand, the original PICh report targeted core telomere sequences and did obtain good results, indicating that it is as yet unclear in how much bound proteins may interfere with probe hybridization. Nevertheless, to alleviate potential interference problems, we designed probes to regulatory regions where the probability of finding proteins bound to them is relatively low. It impossible to ensure "a priori" that the target sequence will be "free of proteins". Interestingly, the likelihood of observing TFs bound to regulatory sequences decreases proportionally to the distance from the core region. Therefore it would be good practice, if TFs DNA-binding pattern are known, to target relatively "empty" motifs, with the hope that the hybridization process takes place without any interference (figure 7).

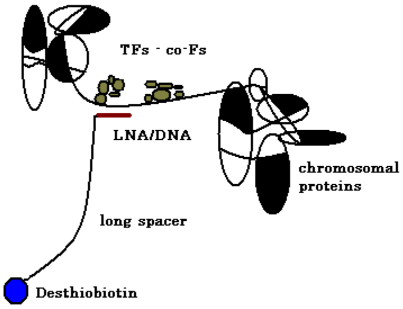


Figure 7. Schematic representation of the probe-chromatin interaction

Once hybridization to chromatin takes place, the complex probe-regulatory (hybrid) sequence will be isolated by purification and sonicated. The resulting sequences are expected to have a length of 400 to 500 bp. The hybrids will be affinity-captured using streptavidin coated magnetic beads, and eluted using free biotin as a competitor. Extensive cross-linking with formaldehyde will ensure a better preservation of protein-DNA and protein-protein complexes. Consecutively, the process of cross-linking will be reversed to allow for the proteins bound to the regulatory sequence to be separated and purified before resuming the classical proteomics workflow.

E-GLOBIN DIFFERENTIAL EXPRESSION USING SILAC

The quantitative analysis of ε-globin expression will be implemented using the two cell populations (K562 and HEK293) in three different conditions: K562 cells alone, K562 cells grown in the presence of hemin (up-regulated), and HEK293 cells. Hemin is used as it increases the expression of globin genes up to ten fold (figure 8).

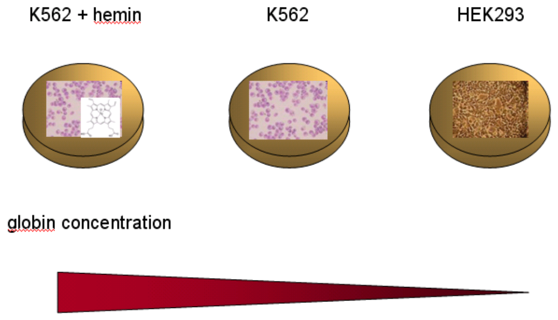
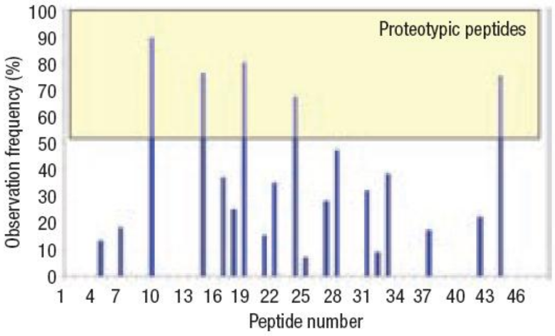


Figure 8. Schematic representation of globin concentration in K562 cells (alone, and in presence of hemin) and in HEK293 cells.

Here, we are interested in observing differential protein composition at specific regulatory loci, in order to understand in a dynamic setting how locus protein composition determines gene expression.

BUILDING A TF PROTEOTYPIC PEPTIDE LIBRARY

Tandem mass spectrometry (MSMS) is routinely used to identify naturally occurring peptide sequences from complex mixtures of proteins. This technique is utilized to obtain structural information of proteins by fragmenting the ions produced in the source of a mass spectrometer and identifying the resulting ions. Structural information of the intact protein can be obtained by pulling together the collected data through bioinformatics. Another use of tandem mass spectrometry is that specific compounds can be detected in complex samples by analysing their fragmentation pattern (selected reaction monitoring or SRM). In this technique, the analysers are set to detect a specific ionic transition belonging to the protein of interest. The mass of the parent as well as the product ion have to be known. SRM is used to confirm unambiguously the presence of a specific protein in a complex mixture. The main advantage is its high specificity and sensitivity. In most instances, as few as two to three identified peptides are sufficient to positively reveal the presence of a particular protein in a sample [34]. Some peptides are more prone to uniquely identify a protein than others, meaning that they are frequently observed. Most mass spectrometers show an identification bias toward analysing peptide species with the most intense MS signal. However, signal intensity often does not relate to frequency of observation. These "proteotypic" peptides should possess particular physico-chemical properties that allow them to be detected at higher frequency (figure 9).



(J. of Proteome Research; Vol. 6, No. 3; 2007)

Figure 9. Diagram showing the frequency of peptide observation.

Ideally, by examining the peptides found in experimental data sets collected by multiple screenings, a comprehensive proteotypic peptides library can be generated.

The development of an open source public repository for peptide tandem mass spectrometry data, the Global Proteome Machine Database (GPMDB), has made it possible to retrieve a list of proteotypic peptides for a limited number of species, based on experimental observations. This repository contains more than four million annotated peptide mass spectra, contributed by many laboratories. Other proteotypic peptide databases include PeptideAtlas (www.peptideatlas.org), SBEAMS (www.sbeams.org), and PRIDE (www.ebi.ac.uk/pride). Until today, no comprehensive TF proteotypic peptide database has been developed. Moreover, due to the difficulties in detecting TF peptides, many

TF entries in these databases remain empty. As previously explained, building a library of TF proteotypic peptides will be a useful tool for the identification and absolute quantification of TFs.

The high sensitivity needed for the detection of transcription factor peptides will be provided by a Fourier Transform Ion Cyclotron Resonance (FT-ICR) analyser (laboratory of Biomolecular Mass Spectrometry, EPFL) as it has been known to have the highest sensitivity among mass spectrometers.

Clearly, proteotypic peptides open exciting new avenues not only for improvements in speed and accuracy of protein identifications, but also for cross-comparisons of quantitative proteomics data, bypassing the use of protein tagging or the use of internal calibrants. The goal is to identify proteotypic peptides for our vast collection of transcription factors to be used in a spectrum based protein identification methodology.

PRELIMINARY RESULTS

DEVELOPMENT OF A MOUSE TF DATABASE

In order to study the complex interplay between DNA and its binding transcription factor and co-factors, our laboratory has cloned already approximately 900 mouse TFs out of the 1,200 planned, representing a good majority of TFs encoded by the mouse genome. cDNA sequences were retrieved from either Genbank or Riken databases, and used as templates for the design of longest open reading frames (lORFs). Each lORF, which encode for a TF of interest, was computed by means of bioinformatics (due to the large number of TFs). For this purpose, I developed a series of object-oriented “perl” scripts in a Linux environment, and subsequently screened the lORFs for the presence of DNA binding domains (BDs) using a program developed by Jacques Rougemont (Bioinformatics and Biostatistics Core Facility, EPFL). The predicted lORFs were compared to the ones present in different databases such as Genbank, Refseq, and Riken. For the majority of TFs, the two sequences were exactly the same. In some instances though, the two sequences showed variations. Those differences were due either to incomplete gene sequencing or differential intron incorporation. Figure 10 shows the decision tree in the choice of the most appropriate sequence to be used for TF cloning. The chosen sequences were used for Gateway-compatible amplification of the insert using specific Gateway primers (see below).

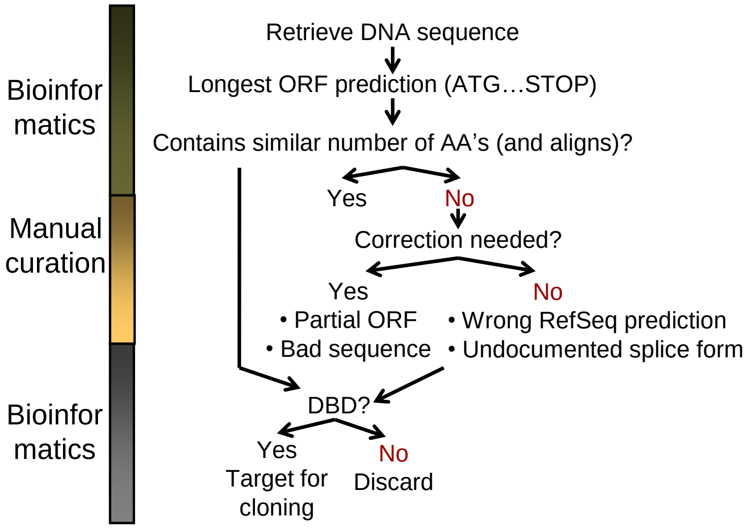


Figure 10. Diagram showing the lORF decision process.

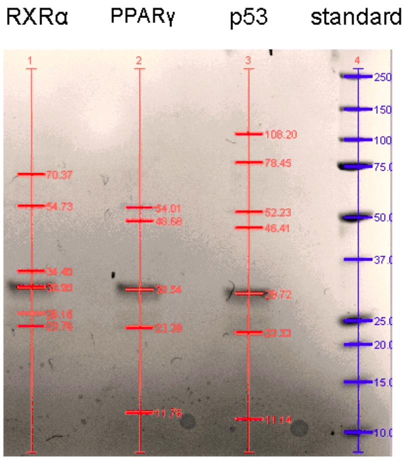
TF "IN VITRO" EXPRESSION USING "GATEWAY" CLONING

To simplify cloning procedures, our laboratory has adopted the "Gateway" cloning technology, which uses a particular plasmid construction strategy to rapidly clone one DNA sequence into multiple destination plasmids. "Gateway" cloning greatly reduces the labour-intensive and time-consuming procedures of classical plasmid construction, and bypasses the utilization of restriction enzymes. This makes the technology useful in many applications, specifically for protein expression. To be able to isolate the proteins, a GST tag was fused to the C-terminus for downstream isolation. We chose the C-terminus as successful GST-based protein purification then implies that the complete protein is expressed.

TFs of interest were expressed using a cell-free eukaryotic protein expression method (in vitro transcription translation system). Basically, in vitro translation is accomplished using a crude lysate from a given organism; providing the translational machinery, tRNAs, accessory enzymes and other factors. Compared to traditional cell-based expression methods, cell-free "in vitro" protein synthesis procedures are extremely rapid, making them suitable for screening expression templates. Reduced reaction volumes and short process time as well as improvements in translation efficiency make the use of a cell-free system suitable for high-throughput strategies as well. Moreover, even proteins that are toxic for cells in high amounts, such as TFs, can be easily expressed.

SELECTION OF PROTEOTYPIC PEPTIDES OF PPARΓ

In the context of discovering proteotypic peptides, peptides that constantly and uniquely identify a given protein, PPARγ (peroxisome proliferator-activated receptor) RXRα, and p53 were chosen for preliminary testing. PPARγ was cloned and expressed using the methods explained above. GST fused PPARγ was expressed using a rabbit reticulocyte lysate and purified using agarose beads coupled to protein A first and using streptavidin coated magnetic beads with biotinilated anti-GST IgG antibodies in a second series of experiments. Proteins in the lysate were separated by molecular weight with SDS-PAGE, and stained with Coomassie Blue (figure 11) and silver nitrate (figure 12). The presence of PPARγ was confirmed by Western Blot (figure 13).



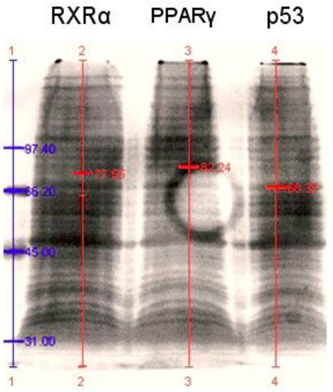


Figure 11. (on the left) Coomassie Blue stained SDS-PAGE gel of PPARγ, RXRα, and p53 expressed in rabbit reticulocytes

Figure 12. (on the right) Silver stained SDS-PAGE gel of PPARγ, RXRα, and p53 expressed in rabbit reticulocytes

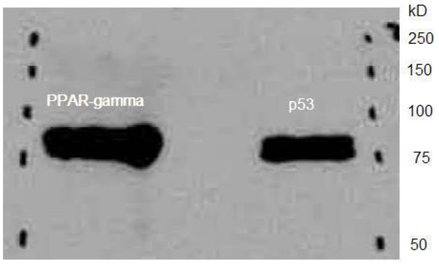


Figure 13. Western Blot of PPARγ and p53 on a nitricellulose membrane using mouse and goat anti-mouse GST abs.

The figures show that Coomassie Blue staining is not sensitive enough to detect protein bands belonging to the expressed proteins. Silver staining on the other hand appears to be much more sensitive, but it is difficult to clearly distinguish bands of interest from the background noise, indicating that proteins are expressed but likely at low levels. The three expressed proteins bands were excised from the gel in figure 2. Peptides were extracted from the fraction after trypsination, and run on an ESI ion-trap with a negative control (blank gel fraction) and a positive control (mouse anti-GST IgG heavy chain). Protein identification with Mascot showed no traces of either PPARγ, RXRα or p53.

The expression yields obtained with the rabbit reticulocyte “in vitro” cell free expression system were simply too low. A wheat germ “in vitro” cell free expression system is more appropriate. Wheat germ has shown a much higher yield in previous preliminary experiments (www.promega.com), and should thus be a better system for the expression of the amount of TF protein needed for MS identification.

DESIGN OF THE PROBES

The gene-centered proteomic approach will firstly be implemented to fish out two regulatory regions (HSII and the epsilon-globin promoter) using the specifically design probes presented above (section 4.a). Two 25mer LNA containing oligonucleotides were designed to have a melting temperature of 76°C (Déjardin & Kingston; 2009). The relative concentration of LNA bases within the oligo determines the melting temperature. The locked nucleic acids provider reported an increase in the Tm of the primer of 2 to 6 °C per incorporation of LNA residue (www.exiqon.com) when compared with DNA. Such a melting temperature ensures that the probe hybridizes to genomic DNA without disrupting the structure of the protein complex bound to the double stranded DNA.

	Sequence (LNA residues are capitalized)	Genomic position	Chromosome	Position from ATG	Tm (°C)
HSII	actCtAggctgaGaacAtctggGca	5,266,142-5,266,118	11	N/A	76
HBE1 promoter	gaGagCtaGaaCtggGtgAgatTct	5,248,184-5,248,208	11	-457	76

Table1. Sequence and genomic information of HSII and HBE1.

Table 1 shows the sequences and characteristics of the two probes that have been designed to target the HSII and HBE1. In order to achieve the desired melting temperature (76°C) LNA residues had to be carefully introduced every 3-4 bp.

FUTURE PERSPECTIVES

It was our goal to extend the platform to the study of less characterized models of gene regulation. Currently our laboratory is studying the regulatory networks underlying adipogenesis, and our method would strongly contribute to the understanding of the dynamic process of differentiation in mouse 3T3-L1 pre-adipocyte cells. Along the lines of the newly developed proteomic platform, promoters of TFs of interest will be screened for interacting proteins or complexes. In later stages, additional up or downstream regulatory elements could be included. The goal will be the creation of a list of high-confidence PDIs, which will consequently be used to generate a gene regulatory network underlying adipocyte differentiation. Once the validity of the platform is established for cell cultures, it would be very interesting to move a step closer to real "in vivo" models by expanding our methodology to the systematic analysis of DNA binding complexes directly in tissue. Recently, Krueger and Mann have extended the SILAC technique to enable proteomic quantitative analysis from tissue samples. Mice were fed with a $^{13}C_6$-lysine diet, without adverse effects. This "SILAC mouse" can therefore to be used as "Wild Type" reference to the analysis of a plethora of knock-out mice. Again, Krueger *et al.* have successfully used different organs from a SILAC mouse to the "in vivo" quantitation proteomic study of the integrin pathway [35]. I am currently exploring the possibility of obtaining specific tissues from SILAC mice to be used to validate the results obtained using cells lines, bringing to the table a different perspective on "in vivo" proteomic studies. Tissues of interest include mouse adipose (namely for the study of the differentiation process of 3T3 cells), and mouse blood (for the dynamic developmental study of globin genes).

The knowledge obtained with our gene-centered proteomic platform can be employed for future modelling efforts to make predictions on how gene regulatory networks behave under different physiological or pathological conditions in different organisms. Our hope is that this technology will serve as a blueprint, and be utilized on a routine basis for the systematic analysis of DNA binding protein complexes within their functional context.

CONCLUSIONS

Our initial efforts lead us to obtaining promising preliminary results. In particularly, we were able to build a comprehensive TF clone library that could be very useful in the context of increasing protein identification confidence or serve as a valid resource to promptly generate TF-specific proteotypic peptides simply by producing TFs of choice "in vitro". To this end, using three well-studied TF as trailblazers, we explored different possibilities to optimize TF expression. It is of crucial importance to utilize a transcription/translation system that can produce proteins in very high yields in order to obtain enough material for downstream applications. In our case, the rabbit reticulocyte protein expression kit we tested did not live up to its expectation. After a careful consideration of "in vitro" expression systems commercially available, we decided to adopt one capable of producing proteins in large amounts. For this purpose, the high-yield wheat germ-based expression system from Promega was selected. However, all things considered, the most crucial aspect of this project is the hybridization of the LNA/DNA probes onto their corresponding target regions. Although the two probes we had devised did successfully bind to genomic DNA (therefore in the absence of DNA-binding proteins) proving that the designing of the probes was well conducted, we experienced major issues in probe hybridization when we switched to using chromatin. Since protein-complex composition as well as chromatin landscape may interfere with probe anchoring, finding accessible target regulatory regions might become a rather difficult task. Our several attempts were not sufficient to achieve a successful landing of the probes onto their targets. It appears that specifically avoiding core regulatory regions, which are often highly occupied, might not be sufficient on its own to increase the likelihood of having the probe hybridizing onto its target sequence. Our initial assumption was that the DNA-binding protein landscape density decreases as we move away from the core regulatory regions, which appears not necessarily always to be the case. Other factors (such as the actual chromatin conformation around the specific target site) play an important role in the process. The latter are unfortunately rather difficult to predict. One possible way to address this problem would be to design several probes targeting different sites within the same regulatory region, hoping that at least one hybridization event takes place. Unfortunately, such a procedure may be impractical and rather uneconomical, due to the high cost of the single probes. If we consider the original development of the PICh technique, which entailed the targeting of telomeric repeat sequences, one probe was used to target multiple genomic loci. A human cell contains 92 telomeres (23 pairs of chromosomes), therefore one probe could hypothetically hybridize onto 92 different genomic sites. In the current work, we aimed at selecting regulatory regions that are present in two copies in diploid cells, representing thereby an increase in sensitivity of a factor 50. In this regard, as to not loose sensitivity we were obliged to start with large amounts of cells to obtain enough material for the subsequent analysis, and planned to use the most sensitive MS-based technology available to tackle the issue of

TF-complex composition identification and subsequent quantification. To sum up, the major bottleneck encountered in the development of this project related to the successful hybridization of the LNA/DNA probes onto their target sequences. After several unsuccessful attempts, we deemed that utilizing one probe only per regulatory region was simply not enough to warrant probe-DNA binding. In essence, the attractiveness of the methodology as a whole is drastically reduced if multiple probes per regulatory element have to be utilized. Moreover, the difficulties we encountered in the implementation of the PICh approach seemed to be shared and acknowledged by many colleagues (thus raising doubts about the reproducible nature of the PICh method). In the light of these facts, we decided to abandon the project and concentrate our efforts on the quantification of TF.

REFERENCES

1. Kim MO, Si Q, Zhou JN, Pestell RG, Brosnan CF, Locker J, Lee SC: **Interferon-beta activates multiple signaling cascades in primary human microglia**. *J Neurochem* 2002, **81**(6):1361-1371.

2. Roberts RE: **The role of Rho kinase and extracellular regulated kinase-mitogen-activated protein kinase in alpha2-adrenoceptor-mediated vasoconstriction in the porcine palmar lateral vein**. *J Pharmacol Exp Ther* 2004, **311**(2):742-747.

3. Blais A, Dynlacht BD: **Constructing transcriptional regulatory networks**. *Genes Dev* 2005, **19**(13):1499-1511.

4. Mak HC, Pillus L, Ideker T: **Dynamic reprogramming of transcription factors to and from the subtelomere**. *Genome Res* 2009.

5. Pillai S, Dasgupta P, Chellappan SP: **Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications in vivo**. *Methods Mol Biol* 2009, **523**:323-339.

6. Encode Consortium: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project**. *Nature* 2007, **447**(7146):799-816.

7. Fantom Consortium: **The Transcriptional Landscape of the Mammalian Genome**. *Science* 2005, **309**(5740):1559-1563.

8. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD *et al*: **In vivo enhancer analysis of human conserved non-coding sequences**. *Nature* 2006, **444**(7118):499-502.

9. Mittler G, Butter F, Mann M: **A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements**. *Genome Res* 2009, **19**(2):284-293.

10. Dejardin J, Kingston RE: **Purification of proteins associated with specific genomic Loci**. *Cell* 2009, **136**(1):175-186.

11. Dean A: **Chromatin remodelling and the interaction between enhancers and promoters in the {beta}-globin locus**. *Brief Funct Genomic Proteomic* 2004, **2**(4):344-354.

12. Bouhassira EE, Krishnamoorthy R, Ragusa A, Driscoll C, Labie D, Nagel RL: **The enhancer-like sequence 3' to the A gamma gene is polymorphic in human populations**. *Blood* 1989, **73**(4):1050-1053.

13. Ptashne M: **Gene regulation by proteins acting nearby and at a distance**. *Nature* 1986, **322**(6081):697-701.

14. Chakalova L, Carter D, Debrand E, Goyenechea B, Horton A, Miles J, Osborne C, Fraser P: **Developmental regulation of the beta-globin gene locus**. *Prog Mol Subcell Biol* 2005, **38**:183-206.

15. Gusev VD, Nemytikova LA, Chuzhanova NA: **[A rapid method for detecting interconnections between functionally and/or evolutionary close biological sequences]**. *Mol Biol (Mosk)* 2001, **35**(6):1015-1022.

16. Muchardt C, Seeler JS, Nirula A, Gong S, Gaynor R: **Transcription factor AP-2 activates gene expression of HTLV-I**. *Embo J* 1992, **11**(7):2573-2581.

17. Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA: **Proximity among Distant Regulatory Elements at the [beta]-Globin Locus Requires GATA-1 and FOG-1**. *Molecular Cell* 2005, **17**(3):453-462.

18. Lozzio CB, Lozzio BB: **Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome**. *Blood* 1975, **45**(3):321-334.

19. Drexler HG, Matsuo Y: **Malignant hematopoietic cell lines: in vitro models for the study of natural killer cell leukemia-lymphoma**. *Leukemia* 2000, **14**(5):777-782.

20. Gahmberg CG, Jokinen M, Andersson LC: **Expression of the major red cell sialoglycoprotein, glycophorin A, in the human leukemic cell line K562**. *J Biol Chem* 1979, **254**(15):7442-7448.

21. Klein E, Ben-Bassat H, Neumann H, Ralph P, Zeuthen J, Polliack A, Vanky F: **Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia**. *Int J Cancer* 1976, **18**(4):421-431.

22. Lozzio BB, Lozzio CB, Bamberger EG, Feliu AS: **A multipotential leukemia cell line (K-562) of human origin**. *Proc Soc Exp Biol Med* 1981, **166**(4):546-550.

23. Wada-Kiyama Y, Peters B, Noguchi CT: **The epsilon-globin gene silencer. Characterization by in vitro transcription**. *J Biol Chem* 1992, **267**(16):11532-11538.

24. Mookerjee B, Arcasoy MO, Atweh GF: **Spontaneous delta- to beta-globin switching in K562 human leukemia cells**. *Blood* 1992, **79**(3):820-825.

25. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR, 3rd: **Direct analysis of protein complexes using mass spectrometry**. *Nat Biotechnol* 1999, **17**(7):676-682.

26. Vester B, Wengel J: **LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA**. *Biochemistry* 2004, **43**(42):13233-13241.

27. Hirsch JD, Eslamizar L, Filanoski BJ, Malekzadeh N, Haugland RP, Beechem JM, Haugland RP: **Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation**. *Anal Biochem* 2002, **308**(2):343-357.

28. Griesenbeck J, Boeger H, Strattan JS, Kornberg RD: **Affinity purification of specific chromatin segments from chromosomal loci in yeast**. *Mol Cell Biol* 2003, **23**(24):9275-9282.

29. Sandaltzopoulos R, Blank T, Becker PB: **Transcriptional repression by nucleosomes but not H1 in reconstituted preblastoderm Drosophila chromatin**. *Embo J* 1994, **13**(2):373-379.

30. Morocho AM, Karamyshev V, Shcherbinina O, Polushin N: **Biotin-labeled oligonucleotides with extraordinarily long tethering arms**. *Methods Mol Biol* 2005, **288**:225-240.

31. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R: **The study of macromolecular complexes by quantitative proteomics**. *Nat Genet* 2003, **33**(3):349-355.

32. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R: **Development and validation of a spectral library searching method for peptide identification from MS/MS**. *Proteomics* 2007, **7**(5):655-667.

33. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics**. *Mol Cell Proteomics* 2002, **1**(5):376-386.

34. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T *et al*: **Computational prediction of proteotypic peptides for quantitative proteomics**. *Nat Biotechnol* 2007, **25**(1):125-131.

35. Kruger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fassler R, Mann M: **SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function**. *Cell* 2008, **134**(2):353-364.

II.II ABSOLUTE COPY NUMBER ANALYSIS OF TRANSCRIPTION FACTORS DURING CELLULAR DIFFERENTIATION USING A MULTIPLEX, TARGETED PROTEOMICS APPROACH

The main goal of the research project is the development of an MS-based quantitative assay aimed at monitoring the dynamic changes in TF abundance underlying biological processes of interest. We intend to establish a multi-layered system for the efficient development of a sensitive medium-throughput SRM assay based on the use of TF proteins expressed in vitro for optimal transition selection. Absolute quantification of endogenous TFs is achieved by spiking heavy labeled TF proteins expressed extemporaneously in a cell-free system. Labeled proteins are quantified *in situ* via a unique engineered light reference-peptide tag. A faster implementation of this technology utilizes additional reference-peptide variants to expand the number of TFs that can be simultaneously analyzed. We ultimately focus on generating a quantitative model of gene regulation during terminal adipogenesis.

The supplementary tables and the supplementary note can be found in the Annex

Absolute copy number analysis of transcription factors during cellular differentiation using a multiplex, targeted proteomics approach

Jovan Simicevic[1,5], Adrian W. Schmid[2,5], Benjamin Zoller[3], Sunil K. Raghav[1] ,Irina Krier[1], Carine Gubelmann[1], Frédérique Lisacek[4], Felix Naef[3], Marc Moniatte[2,]*, Bart Deplancke[1,*]

[1]Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[2] Proteomics Core Facility, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[3] Laboratory of Computational Systems Biology, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[4]SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

[5]These authors contributed equally to this work

ABSTRACT

The regulatory properties of transcription factors (TFs) are largely dictated by their cellular abundance. Thus, deriving absolute TF copy numbers is crucial to understand how these proteins control gene expression. Here, we present a novel and sensitive selected reaction monitoring (SRM)-based mass spectrometry assay, allowing us to simultaneously determine the absolute copy numbers of up to 10 proteins. We apply this approach to profile the levels of key TFs during adipogenesis, revealing that their abundance differs dramatically (from 250 to >300,000 copies per nucleus), but that their dynamic range during differentiation varies at most five-fold. We also formulate a genome-wide TF DNA binding model to explain the significant increase in PPARγ binding sites during the final differentiation stage, despite a concurrent saturation in PPARγ copy number. This model provides unique, quantitative insights into the relative contributions of binding energetics, copy number, and chromatin state in dictating TF DNA occupancy profiles.

Understanding how the expression of large sets of genes is orchestrated at a systems level is a topic of fundamental importance in biology. Differential gene expression is controlled by gene regulatory networks, which consist of functional interactions between regulatory state (e.g. transcription factors (TFs)) and genomic (e.g. gene promoters, enhancers) components[1]. A major current interest is to derive models of gene regulatory networks that elucidate or predict the dynamic transcriptional mechanisms underlying complex gene expression programs[2-4]. An important drawback of existing models is that relatively little quantitative data has so far been used in their calibration. In particular, accurate measurements on the absolute, molecular abundance of most TFs are still very sparse[2, 5], even though such information is key to understand most biochemical and regulatory processes involving this type of proteins[6, 7]. This conundrum can be explained by the relatively low expression of TFs in cells[8], which make their direct quantification by mass spectrometry (MS) or other assays a substantial challenge. Consequently, only a few MS-based studies have to date been able to provide TF copy numbers in higher eukaryotes[9, 10]. However, since these studies were not specifically designed to target TFs, they did not tackle the difficult task of tracking the expression of TFs of interest over the course of specific biological processes.

Newly emerging proteomic approaches combining selected reaction monitoring (SRM) with isotope dilution quantification strategies now promise to alleviate this scarcity in absolute copy number data as they enable targeted, quantitative analyses[10-13]. These approaches all share the same principle of targeting a subset of detectable peptides which are specific to the protein of interest (i.e. proteotypic peptides). Quantification is thereby achieved by comparing their extracted MS signals to those of accurately quantified, isotopically-labeled peptides (usually having the same sequences), which are utilized as internal standards. Two recently published quantification strategies introduced the use of a single reference peptide serving as a cost-effective surrogate and allowing the accurate measurement of synthetically produced, full-length proteins[14, 15]. This intermediate information is crucial in the following step, where the complete set of protein-specific peptides can in turn be used to accurately quantify the endogenous protein within a complex sample. This has the advantage that both the endogenous and the synthetically produced counterpart are enzymatically digested together, leading to a more accurate and reproducible quantification[16, 17]. However, while powerful, this entire workflow currently only permits the direct quantification of one protein per assay, making it a costly and time-consuming exercise. This is because quantifying more than one protein at a time necessitates that the standards are quantified in a separate step. Moreover, precipitation and resolubilization issues inherently linked to protein/peptide storage may interfere with the overall accuracy of the methodology. Together, these considerations prompted us to develop an SRM-based assay in which the internal protein standard quantification step is performed *in situ* (i.e. both the protein standard and its endogenous counterpart are simultaneously quantified within the same assay). In addition, we managed to multiplex this procedure, an important step towards addressing the need for assay upscaling, which remains one of the main limitations in current SRM-based assays as recently highlighted by Picotti and Aebersold[18]. As such, our assay is capable of monitoring and determining absolute amounts of pre-determined sets of up to 10 TFs (or other lowly abundant proteins of interest) per single, analytical run.

Here, we applied this workflow to assess the dynamic changes in absolute TF levels during the terminal phase of adipogenesis. While quantitative proteomic studies have been performed to explore the adipogenic proteome, these lacked the required sensitivity to detect and quantify core adipogenic TFs[19-21]. Adipogenesis is in part orchestrated by two master regulators, PPARγ and RXRα, and can to a reasonable extent be mimicked *in vitro* using the mouse pre-adipocyte 3T3-L1 cell line[22, 23]. We were able to successfully monitor the temporal changes in nuclear PPARγ and RXRα abundance during 3T3-L1 differentiation, thereby obtaining values on absolute copy numbers per cell. This first accomplishment prompted us to apply the multiplex variant of our methodology to derive absolute measurements for 10 TFs within this cellular system in one single run. Herein, we demonstrate that the monitoring and quantification of sets of lowly abundant proteins can be achieved without compromising detection sensitivity. Finally, we provide a first insight into how TF abundance data can be modeled in conjunction with large-scale DNA occupancy and chromatin state data to further our understanding of gene regulatory mechanisms mediating cellular differentiation. Specifically, our model reveals how the DNA binding profile of PPARγ is in large part influenced by its own copy number and local chromatin state, thus underwriting the importance of both protein levels and chromatin remodeling in dictating TF DNA binding behavior.

RESULTS

SELECTION OF TF-SPECIFIC PROTEOTYPIC PEPTIDES AND CORRESPONDING FRAGMENT-IONS FOR TARGETED SRM ASSAYS

We initially turned to fragmentation spectra databases such as the NIST (http://www.nist.gov/nvl/) and PeptideAtlas[24] to select the ideal pool of peptide candidates for the design of our SRM-based assay, aiming to monitor nuclear PPARγ and RXRα abundance during terminal adipogenesis. However, only a limited amount of information regarding TF-specific peptides could be retrieved. We therefore had to devise a strategy that did not rely on the use of publicly available information to discover proteotypic peptides of TFs. To alleviate concerns regarding the detectability of TFs, we implemented three different bi-dimensional peptide fractionation pipelines on 3T3-L1 total nuclear protein extracts at day four of differentiation (a time point at which the adipogenic master regulators PPARγ and RXRα are known to be highly expressed[25, 26]) with the precise aim of discovering TF-specific peptide candidates that could be utilized in downstream quantitative analyses. However, only two peptides for each TF were retrieved in spite of the higher resolving power provided by the extensive fractionation and the consecutive higher sensitivity expected in mass spectrometry (**Supplementary Fig. 1**). Thus, despite recent increases in overall sensitivity [27], MS-based shotgun data-dependent methodologies are still lacking the combination of acquisition speed and sensitivity necessary to directly analyze lowly abundant proteins such as TFs within complex matrices derived from mammalian cells. The results of these preliminary experiments partly explain the scarcity of TF-specific peptides that we initially observed in databases. In response, we and others[28] are currently undertaking significant efforts to address this TF data paucity, but it is clear that current TF-specific peptide repositories are still far from being complete. An alternative strategy based on the combined use of proteotypic peptide prediction tools and crude peptides synthesis[29] did also not improve the quality of our acquired data. In fact, the majority of the synthetic peptides were either hardly detectable or not consistently detectable over the time course of the experiment. We attribute this behavior to peptide solubility and ionisability issues. We therefore searched for a more direct route to access detectable TF peptides, opting for a full-length protein expression-based strategy, which allows for a better accounting of the local environment surrounding the peptide cleavage sites. The enhanced wheat germ *in vitro* transcription-translation system proved to be straight-forward to use and led to the highest protein production yield (**Online Methods** and data not shown). Ten TFs (including PPARγ and RXRα) that exhibit gene expression at some point during terminal adipogenesis were retained for downstream applications based on their high yields upon *in vitro* expression. Importantly, this expression system also tags proteins with an in frame C-terminal glutathione S-transferase (GST) sequence for full-length expression validation and efficient purification (**Supplementary Table 1**). Highly enriched GST-tagged proteins with the correct, expected molecular weight were successfully obtained after optimization of the synthesis, capture, wash and recovery conditions as exemplified in **Supplementary Figure 2**. In-gel digestion followed by high resolution LC-MS/MS subsequently enabled us to generate a collection of experimentally detected TF-specific tryptic peptides for the 10 TFs (**Supplementary Table 2**). The pool of peptides identified in the preliminary screen was submitted to a series of manual curation steps to ensure maximum sensitivity and selectivity, which was subsequently validated by SRM (**Online Methods**). Specifically, we selected peptides that exhibit reliable detectability, digestability, homogeneity, stability and uniqueness (whenever possible).

IMPLEMENTATION OF THE SRM-BASED METHODOLOGY

In SRM, the quantification of proteins is generally achieved by spiking a set of pre-determined, isotopically-labelled proteotypic peptides into a complex sample at known concentrations to serve as standards[30]. It would thereby be optimal to dispose of the complete set of protein-specific peptides, as peptide behavior in the mass spectrometer is rather difficult to predict, but the cost associated with accurate quantification of synthetic peptides makes this difficult to implement (e.g.[31]). Moreover and perhaps more importantly, the differences in physicochemical behavior between peptides makes the use of individual peptides less attractive, since peptide solubility will affect the measurement accuracy[32]. Building on pioneering work[14, 15, 33], we therefore set out to utilize full-length, isotopically-labelled, *in vitro*-expressed TFs directly as standards for quantification of endogenous TFs utilizing a modified version of the same workflow that we have adopted for transition selection. For this purpose, we added an in-frame proteotypic reference-peptide tag previously employed for quantitative interaction studies, SH-Quant[15], at the N-terminus of the protein constructs. This tag is utilized for the *in situ* quantification of heavy TFs utilizing a synthetic and accurately quantified light SH-Quant counterpart. In this manner, one single peptide only (the SH-Quant) is used for the absolute quantification of a heavy, *in vitro*-expressed standard TF, thus simplifying and reducing the cost of the assays even further. In turn, quantification of the endogenous TF in absolute amounts is achieved by utilizing TF-specific peptides (**Fig. 1** and **Fig. 2a**).

ABSOLUTE QUANTIFICATION OF TWO ADIPOGENIC MASTER REGULATORS: PPARΓ AND RXRA

To evaluate PPARγ and RXRα peptide transitions, two separate pilot SRM assays were implemented as depicted in **Figure 1** and exemplified in **Figure 2b (**details in **Supplementary Figs. 3-5**).

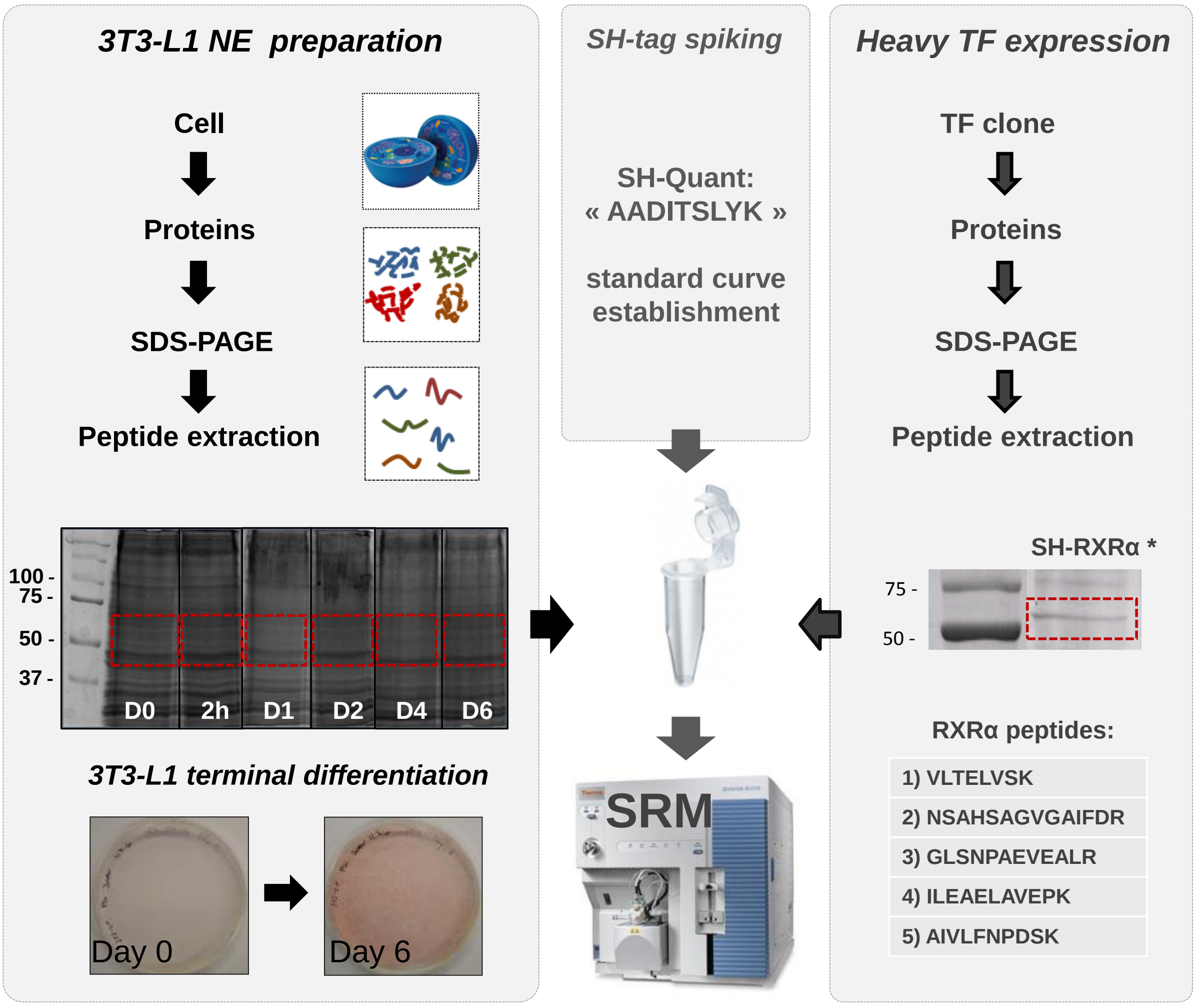


Figure 1: Workflow for the absolute quantification of TFs in 3T3-L1 cells.

The left panel shows the preparation of 3T3-L1 total nuclear protein extract. Cells are lysed at each differentiation time point (0h or Day 0 (D0), 2h, Day 1 (D1), Day 2 (D2), Day 4 (D4), and Day 6 (D6)) after which nuclear proteins are extracted. The resulting protein mixture is separated by SDS-PAGE, stained with Coomassie Blue, and bands where selected TFs (RXRα here) are expected to be located are excised from the gel. The right panel shows the preparation of *in vitro*-expressed SH-tagged TFs. The constructs are expressed in their heavy-labeled version (*) using a wheat germ-based *in vitro* transcription-translation kit, purified by GST affinity, separated by SDS-PAGE, and stained with Coomassie Blue. Bands containing the heavy-labeled constructs (SH-RXRα-GST* here) are excised from the gel and sliced. Each nuclear extract band to be quantified is then mixed with one gel slice of the *in vitro*-expressed TF construct, spiked after dehydration with known amounts of light SH-Quant tag and in-gel digested together. The resulting peptide mixtures are then quantified by SRM using validated proteotypic peptides selected from information previously collected on each *in vitro*-expressed TF via shotgun-MS runs. Each TF quantification requires a separate experiment in this configuration.

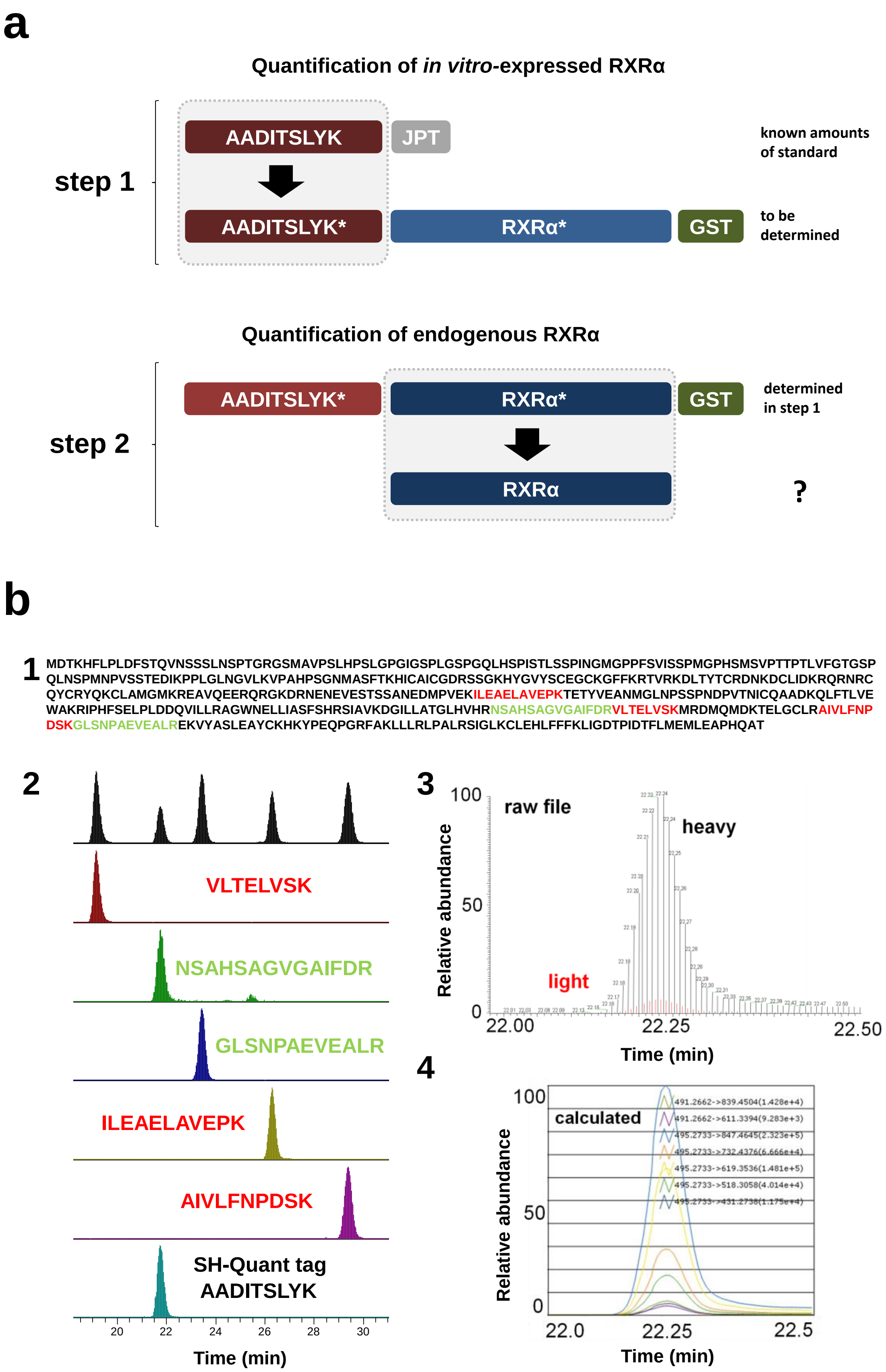

a
Quantification of *in vitro*-expressed RXRα
step 1
AADITSLYK
JPT
known amounts of standard
AADITSLYK*
RXRα*
GST
to be determined
Quantification of endogenous RXRα
step 2
AADITSLYK*
RXRα*
GST
determined in step 1
RXRα
?
b
1
MDTKHFLPLDFSTQVNSSSLNSPTGRGSMAVPSLHPSLGPGIGSPLGSPGQLHSPISTLSSPINGMGPPFSVISSPMGPHSMSVPTTPTLVFGTGSP
QLNSPMNPVSSTEDIKPPLGLNGVLKVPAHPSGNMASFTKHICAICGDRSSGKHYGVYSCEGCKGFFKRTVRKDLTYTCRDNKDCLIDKRQRNRC
QYCRYQKCLAMGMKREAVQEERQRGKDRNENEVESTSSANEDMPVEKILEAELAVEPKTETYVEANMGLNPSSPNDPVTNICQAADKQLFTLVE
WAKRIPHFSELPLDDQVILLRAGWNELLIASFSHRSIAVKDGILLATGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNP
DSKGLSNPAEVEALREKVYASLEAYCKHKYPEQPGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQAT
2
VLTELVSK
NSAHSAGVGAIFDR
GLSNPAEVEALR
ILEAELAVEPK
AIVLFNPDSK
SH-Quant tag
AADITSLYK
20
22
24
26
28
30
Time (min)
3
100
50
0
Relative abundance
raw file
heavy
light
22.00
22.25
22.50
Time (min)
4
100
50
0
Relative abundance
calculated
491.2662->839.4504(1.428e+4)
491.2662->611.3394(9.283e+3)
495.2733->847.4645(2.323e+5)
495.2733->732.4376(6.666e+4)
495.2733->619.3536(1.481e+5)
495.2733->518.3058(4.014e+4)
495.2733->431.2738(1.175e+4)
22.0
22.25
22.5
Time (min)

Figure 2. *In situ* protein quantification procedure and SRM-based monitoring of light and heavy RXRα peptides with their corresponding quantification tag (SH-Quant) spiked together into a nuclear extract sample.

(a) The scheme outlines the two step procedure used for the SRM-based quantification strategy (RXRα is used here as an example). In step 1, the amount of heavy-labeled, *in vitro*-expressed SH-RXRα-GST is quantified by isotope dilution using a pre-determined amount of light, accurately quantified SH-Quant reference-peptide (SH-tag). Specifically, quantification of the *in vitro*- expressed heavy construct is achieved based on the ratio of the light versus the heavy SH-tag peptide. Hence, the calculation of the heavy SH-tag reveals the amount of total heavy RXRα found in the assay. In step 2, RXRα-specific heavy signature peptides are used to accurately determine the amounts of endogenous RXRα present within the 3T3-L1 nuclear extract by comparing heavy versus light (endogenous) peptides. The two steps are accomplished at the same time to further simplify the assay and to limit measurement variations that may occur during differential treatment of the samples. Typical peptide fragmentation signatures, as well as the LC co-elution of the heavy TF-specific peptides serve as strong identification criteria in SRM-based mass spectrometry. **(b)** Subpanel 1: Positioning of the five proteotypic-selected peptides within RXRα. For the sake of clarity, their sequences are color-coded (peptides ending with a lysine in red, peptides ending with an arginine in green). Subpanel 2: Total ion extraction of five proteotypic peptides of the light (endogenous) and heavy forms as well as the SH-Quant tag. Clear separation of all peptides was achieved using a short LC gradient. Subpanel 3: raw file showing a zoom-in of the SH-Quant tag (AADITSLYK) in its light and heavy forms (colored in red and black respectively) and the calculated peak area using Pinpoint (subpanel 4). The calculated peak intensities of selected fragment ions of the light and heavy tag are also presented.

Subsequently, the complete workflow was applied to monitor the absolute, nuclear abundance of PPARγ and RXRα during terminal adipogenesis. The technical robustness of the whole workflow is highlighted by the low coefficient of variation (CV) obtained within one biological sample (**Supplementary Fig. 6**). The resulting data, as summarized in **Figure 3** and **Supplementary Table 3**, provide unique insights into the dynamic protein copy number per cell (nucleus) profiles of PPARγ and RXRα. We found that there are approximately two- to six-fold more RXRα than PPARγ protein copies in the nucleus depending on the differentiation time point. Prior to the induction of differentiation, nuclear TF copy numbers are at their lowest with ~2,000 for PPARγ and ~10,000 for RXRα. The number of RXRα molecules then peaks two days after differentiation (~32,000) after which it declines from ~23,000 at day four to ~13,000 at day six. In contrast, the greatest number of nuclear PPARγ molecules was found at day four (~8,500) with a small decline at day six (~7,500).

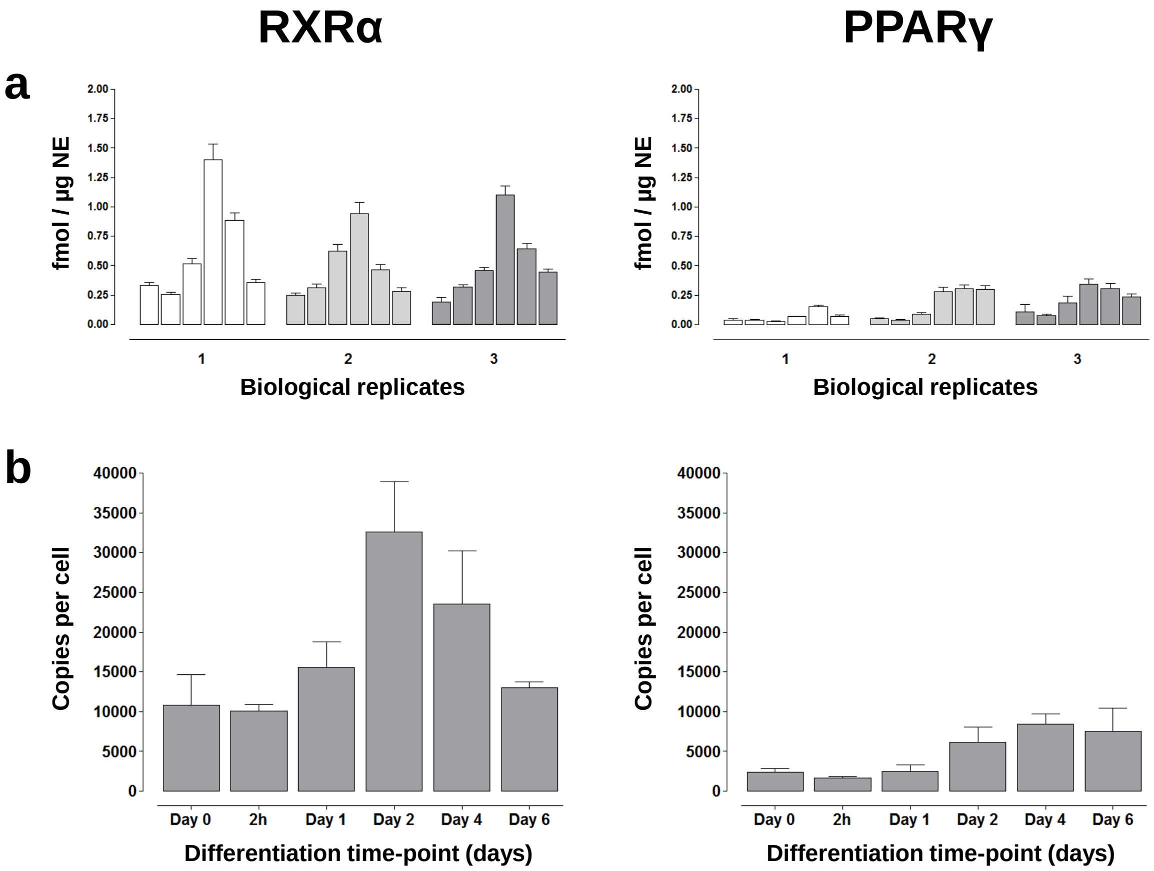


Figure 3. Summary of RXRα and PPARγ levels quantified by SRM.

(a) A graphical summary of the concentration (expressed as fmol/µg nuclear extract (NE)) of RXRα (left) and PPARγ (right) in three individual biological replicates (bars represent the mean ± SD of three technical replicates). **(b)** Absolute RXRα and PPARγ levels visualized as copies per cell (bars represent the mean ± SEM of three biological replicates).

To validate our results, we performed Western blot analysis utilizing anti-PPARγ and anti-RXRα antibodies for each time-point (**Supplementary Fig. 7**). The detected TF copy number profiles mirror those observed by our SRM-based quantification approach, although it is clear that only relative changes can be assessed using immunoblotting and that this technique appears to inflate (PPARγ) or deflate (RXRα) the actual, dynamic changes. Together, our results indicate that nuclear PPARγ and RXRα protein copy numbers change substantially over the course of adipogenesis. In addition, they illustrate both the accuracy and sensitivity of our targeted proteomics approach.

The availability of absolute protein copy number data allowed us to examine the relationship between the number of genome-wide DNA binding events and TF molecules. Using RXRα and PPARγ ChIP-seq data from Nielsen et al.[34], we found that, until the fourth differentiation day, the number of TF molecules and corresponding binding events correlate well (R^2 = 0.96 for PPARγ; R^2 = 0.85 for RXRα; **Supplementary Fig. 8**), and that in general, there are substantially more RXRα than PPARγ binding events consistent with their corresponding TF copy numbers. The slightly lower correlation between the number of RXRα binding events and protein copies compared to that for PPARγ may reflect the ability of RXRα to bind DNA as a homodimer, which may reduce the binding site count per RXRα molecule. Nevertheless, the high correlations indicate that ChIP-seq occupancy data likely reflect endogenous conditions rather than being the result of differences in antibody-mediated protein recoveries as previously hypothesized[34]. A striking discrepancy was however observed for day six as, compared to previous days, the number of RXRα and PPARγ binding events significantly increases (>three-fold for both TFs), whereas the number of TF molecules saturates or even decreases (**Supplementary Fig. 8**). To reconcile these findings and to possibly provide a mechanistic explanation for this phenomenon, we generated a quantitative model to predict the number of PPARγ binding events in the genome given a number of TF molecules, ChIP detection threshold, and genome accessibility maps (**Supplementary Note,** the document can be found in the Annex). The chromatin mark histone three lysine 27 mono-acetylation (H3K27Ac) was recently found to be a good indicator of active or accessible regulatory regions given its substantial overlap with FAIRE (open chromatin) sites[35] and its utility in identifying active promoters and enhancers[36]. In addition, more than 80% of PPARγ binding sites are located within H3K27Ac-enriched regions[37], further strengthening the observation that the H3K27Ac mark is a reasonable proxy for chromatin accessibility. Consequently, we formulated the model to account for the distribution of specific and non-specific sites for PPARγ in H3K27Ac-enriched regions, as derived from recent genome-wide enrichment data for differentiation days zero, two, and six[37] (**Fig. 4a**). To assess the DNA binding events at thermodynamic equilibrium, we assumed that all the PPARγ proteins were on the DNA at either specific or non-specific sites, consistent with the consensus in the field that TFs tend to mostly reside on DNA *in vivo*[7]. For the sake of simplicity, we did not consider interactions with other TFs, or hindrance due to other proteins sitting on DNA (for a discussion on these topics, see **Supplementary Note**). The resulting model predicts the number of binding events given the number of PPARγ copies per cell (nucleus) for each time point as a function of the ChIP detection threshold expressed as percentage occupancy (or residency time) (**Fig. 4b**). While it is difficult to precisely estimate the latter threshold, a survey of our own and published ChIP data revealed that the signal associated with DNA binding is typically considered as positive from around 1% (as compared to input)[26, 38]. Interestingly, when we used ~1%

as detection threshold, our model predicts binding event numbers which closely mimic the experimentally derived data (**Fig. 4c**).

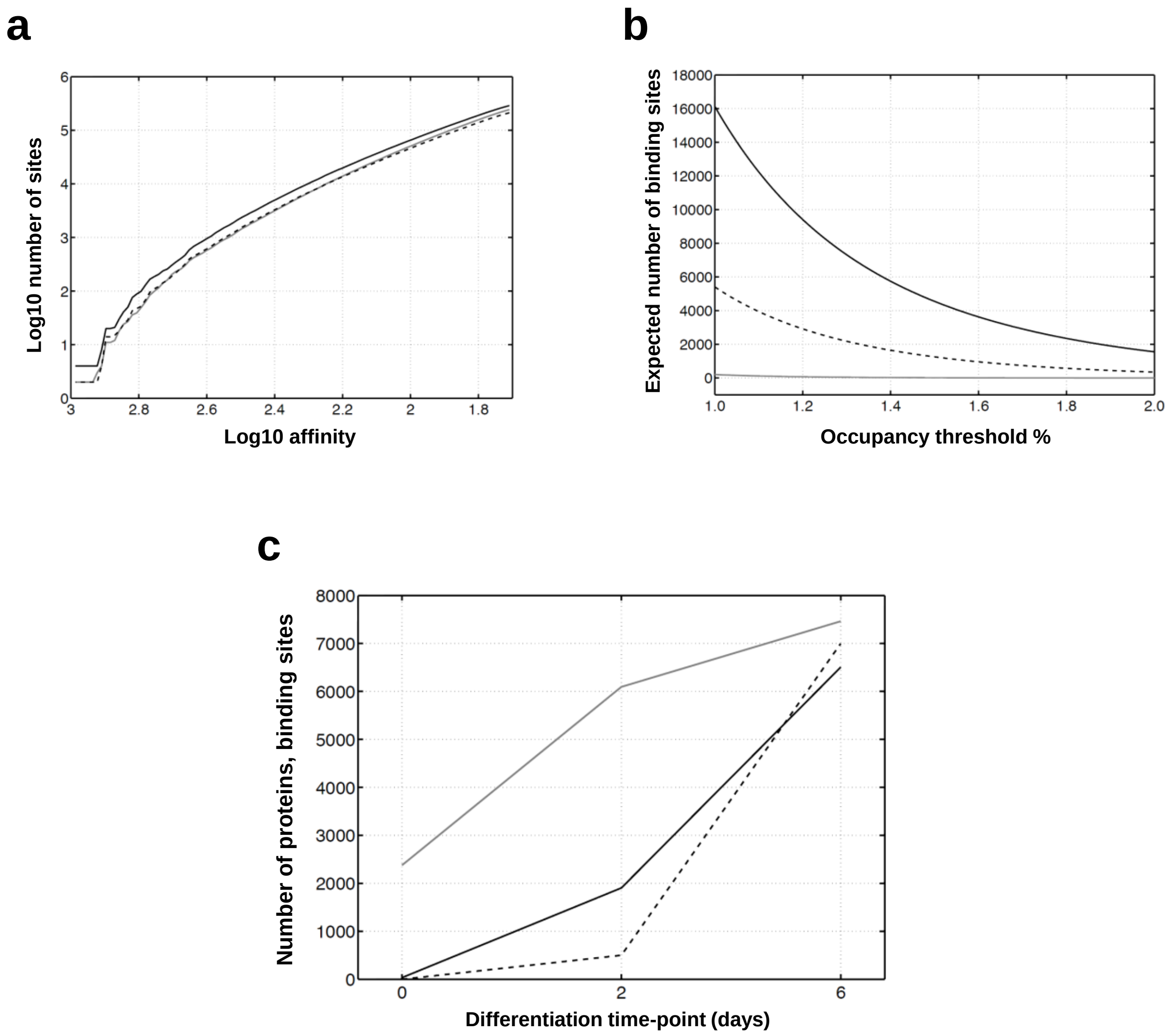


Figure 4. Quantitative modeling of genome-wide PPARγ DNA binding.

(a) Cumulative number of genomic sites in H3K27Ac regions at Day 0 (gray line), Day 2 (dashed line), and Day 6 (black line). The axis runs from the strongest (1000-fold stronger than non-specific sites) to medium affinity sites (60-fold stronger). While the numbers for Day 0 and Day 2 are comparable, Day 6 shows a 35% increase in this range. **(b)** Number of detected bound loci at Day 0 (gray line), Day 2 (dashed line), and Day 6 (black line) during 3T3-L1 terminal differentiation in function of the detection threshold on the expected occupancy. The model takes into account the measured PPARγ copies per cell (nucleus) and the distribution of accessible high affinity sites. Matching with the number of measured ChIP-seq sites predicts an occupancy threshold of 1.35%. **(c)** Temporal pattern for the prediction on the number of detected PPARγ-bound sites (dashed line), the actual number of measured sites (black line), and the protein copy number (grey line). Note that the number of sites shows an exponential-like increase, while the protein copy number graph reflects saturation.

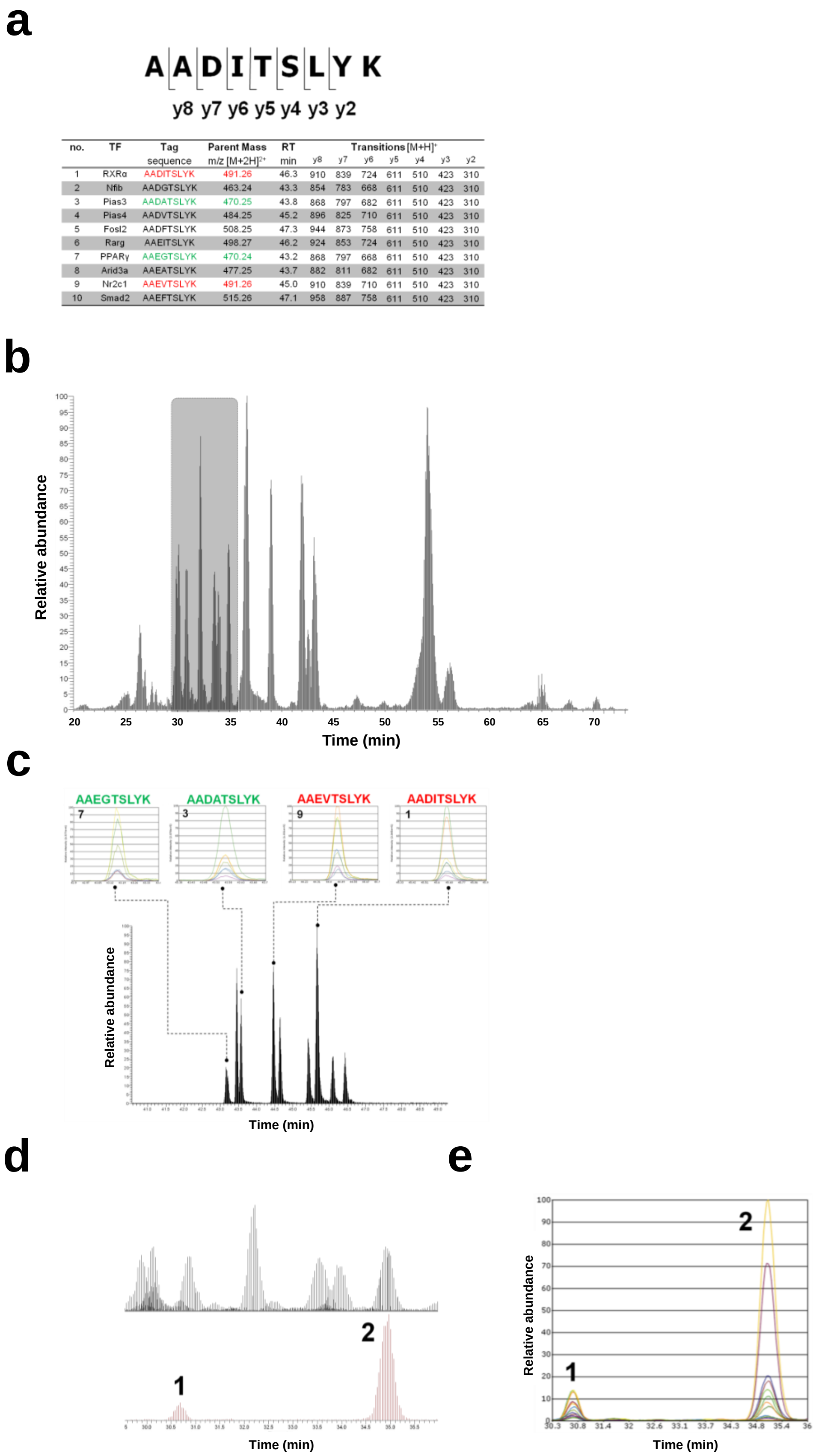
Thus, the model correctly predicts the temporal pattern in the number of binding sites over the course of terminal adipogenesis. A closer look into the determining factors revealed that the sharp increase in binding events at day six as compared to days zero and two is in part driven by the increase in PPARγ copy number between days two and six. However, we found that an equally important factor relates to shifts in the relative distributions of higher versus lower affinity sites in accessible genomic regions at the different differentiation time points (**Fig 4a**). Specifically, whereas the overall size of accessible regions is similar across all days, we found that cells undergo important chromatin remodeling between days six as compared to zero and two such that more medium-to-high affinity sites for PPARγ become available. This in turn allows for a substantial increase in detectable DNA binding events even though the TF copy number increase is relatively modest. Together, these analyses show the value of accurately quantifying TF copy numbers to generate quantitative DNA binding models from which emergent properties regarding gene regulatory mechanisms underlying a specific biological process can be derived.

MULTIPLEXING THE SRM ASSAY

Although our approach now enables us to monitor absolute protein copy numbers over time in a sensitive and reproducible manner, we sought to overcome the current limitation of quantifying only one protein per assay. For this purpose, we designed 9 quantotypic SH-Quant tag variants that, along with the original SH-Quant tag, allow the quantification of up to 10 proteins or TFs in one single integrated SRM assay, thereby increasing the practical usefulness of the methodology. This required changing one or two amino acids within the parent SH-Quant sequence for each variant, while retaining the best transitions for tag quantification and creating distinct peptide fragment signatures for tag identification purposes.

We subsequently generated 10 expression vectors enabling the coupling of each tag variant to a distinct, adipogenic TF. Our list includes: RXRα, NFIB, PIAS3, PIAS4, FOSL2, RARg, PPARγ, ARID3a, NR2C1 and SMAD22 (we included PPARγ and RXRα as validation). Corresponding transitions as well as the retention time of the parent ions are presented in **Figure 5a**. The performance as well as chromatographic separation of the nine, newly designed candidates and their parent SH-Quant sequence was tested in a series of SRM runs (**Fig. 5b**). Two SH-Quant tag-variants

“AADITSLYK” and “AAEVTSLYK” are isobaric and yet could be clearly distinguished by their chromatographic profile and by their distinct transitions (**Fig. 5c,d,e**).



no.	TF	Tag	Parent Mass	RT	Transitions [M+H]$^+$						
		sequence	m/z [M+2H]$^{2+}$	min	y8	y7	y6	y5	y4	y3	y2
1	RXRα	AADITSLYK	491.26	46.3	910	839	724	611	510	423	310
2	Nfib	AADGTSLYK	463.24	43.3	854	783	668	611	510	423	310
3	Pias3	AADATSLYK	470.25	43.8	868	797	682	611	510	423	310
4	Pias4	AADVTSLYK	484.25	45.2	896	825	710	611	510	423	310
5	Fosl2	AADFTSLYK	508.25	47.3	944	873	758	611	510	423	310
6	Rarg	AAEITSLYK	498.27	46.2	924	853	724	611	510	423	310
7	PPARγ	AAEGTSLYK	470.24	43.2	868	797	668	611	510	423	310
8	Arid3a	AAEATSLYK	477.25	43.7	882	811	682	611	510	423	310
9	Nr2c1	AAEVTSLYK	491.26	45.0	910	839	710	611	510	423	310
10	Smad2	AAEFTSLYK	515.26	47.1	958	887	758	611	510	423	310

Figure 5. Properties of the SH-Quant tag variants.

(a) The table provides a detailed summary of the different tags designed for the multiplex assay. The sequences of tags 2 to 10 derive from the original SH-Quant tag 1, and were designed with the intent of conserving intense y” fragment ions at positions y2 to y5 for quantification purposes. The fragmentation fingerprint of ions y6 to y8 allow for improved tag identification especially in the case of isobaric peptides. Overall, this targeted design resulted in a clear LC separation of isobaric tags 3 and 7 (m/z = 470.2; green) or tags 1 and 9 (m/z = 491.2; red). **(b)** LC-SRM spectrum of a multiplexed nuclear extract sample using a 120 minute gradient. **(c)** Chromatographic separation of all ten SH-Quant tags. Near complete chromatographic separation was achieved for most tags using a short LC gradient of 60 min. **(d)** Zoom-in view of the area highlighted in panel c, showing the extracted ion current (EIC) of isobaric SH-Quant tags 1 and 9 in a multiplex sample (peak 1: m/z = 491.2, peak 2: 491.2) and the calculation **(e)** of the peak areas of the light and heavy versions using Pinpoint. The above outlined sequence design resulted in a clear difference in physicochemical properties of two isobaric SH-Quant tags, as seen by their different elution times at 30.7min (AAEVTSLYK) and 35min (AADITSLYK) respectively in a complex sample. Identical elution times were observed for both light and heavy versions of all 10 SH-Quant tags.

A standard curve was established for each SH-Quant tag variant by spiking them separately within the same 3T3-L1 nuclear extract preparation that was next used for TF abundance measurements (**Supplementary Figs. 9-10**). This allowed us to determine the lower limit of quantification (LLOQ) for each tag (the original SH-Quant tag is presented as an example in **Supplementary Fig. 11**). The increased sample complexity also required optimization of the hardware setup for LC separation. Applying the multiplex version of our methodology, we were able to monitor endogenous levels of 10 TFs in 3T3-L1 nuclear extracts derived from three terminal differentiation time points (Day 0, Day 2, and Day 6) (**Figs. 6a,b**). As summarized in **Figure 6c** and **Supplementary Figure 12**, we found that the majority of the additionally analyzed TFs (e.g. PIAS4, NFIB, SMAD2, RARg, FOSL2) are expressed within approximately the same absolute copy number range as PPARγ or RXRα (**Supplementary Table 3**).

Importantly, measurements of PPARγ and RXRα copy numbers were coherent with data obtained with the non-multiplex approach, indicating that the up-scaling of the SRM assay does not interfere with measurement precision. The abundance of other TFs such as PIAS3, NR2C1, and ARID3a fell below the quantification threshold. Nevertheless, the overall quality of the fragment ions monitored for the heavy and light peptides (**Fig. 6d)** clearly allows for the unambiguous identification of these TFs.

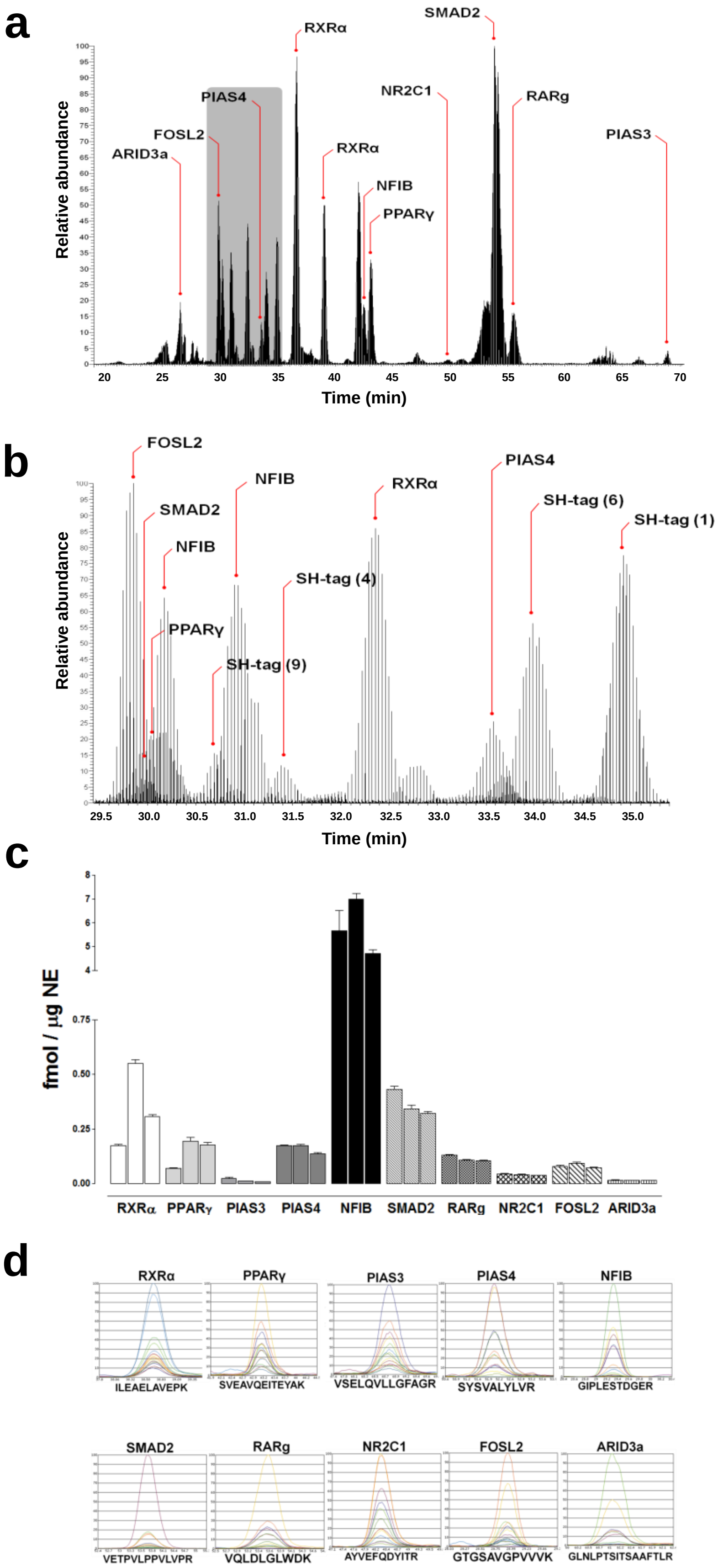

a
Relative abundance
Time (min)
ARID3a
FOSL2
PIAS4
RXRα
RXRα
NFIB
PPARγ
NR2C1
SMAD2
RARg
PIAS3
20
25
30
35
40
45
50
55
60
65
70
b
Relative abundance
Time (min)
FOSL2
SMAD2
NFIB
PPARγ
SH-tag (9)
NFIB
SH-tag (4)
RXRα
PIAS4
SH-tag (6)
SH-tag (1)
29.5
30.0
30.5
31.0
31.5
32.0
32.5
33.0
33.5
34.0
34.5
35.0
c
fmol / µg NE
0.00
0.25
0.50
0.75
4
5
6
7
8
RXRα
PPARγ
PIAS3
PIAS4
NFIB
SMAD2
RARg
NR2C1
FOSL2
ARID3a
d
RXRα
ILEAELAVEPK
PPARγ
SVEAVQEITEYAK
PIAS3
VSELQVLLGFAGR
PIAS4
SYSVALYLVR
NFIB
GIPLESTDGER
SMAD2
VETPVLPPVLVPR
RARg
VQLDLGLWDK
NR2C1
AYVEFQDYITR
FOSL2
GTGSAVGPVVVK
ARID3a
GLNLPTSITSAAFTLR

Figure 6. Simultaneous monitoring of the nuclear abundance of ten TFs during terminal adipogenesis.

(a) Multiplex SRM transition profiles of selected best responding tryptic peptides from all ten TFs during a 120 minute LC-SRM run. **(b)** Zoom-in (minutes 29-36) of the highlighted region in spectrum (a), illustrating the complexity of the mixture and the separation quality achieved with the multiplex SRM assay on nuclear extract sample. **(c)** Calculated endogenous levels of all ten TFs detected at Day 0, Day 2, and Day 4 (bars represent the mean ± SEM of 4-6 technical replicates). **(d)** Calculated peak areas of ten proteotypic peptides in their light and heavy versions.

DISCUSSION

In this study, we describe the development of a sensitive and multiplex, targeted MS assay, which uniquely allowed us to monitor the absolute copy number fluctuations of specific TFs of interest during a cellular differentiation process. Determining the copy number of this type of proteins has been of longstanding interest given that the DNA binding ability of a TF and hence its regulatory capacity strongly depends on its cellular concentration[6, 7, 39]. However, despite the fundamental importance of TFs in most biological processes, a recent comprehensive survey showed that only a handful of studies have so far provided estimates on the abundance of animal TFs[7]. These analyses were mostly achieved with indirect immuno-based measuring methods (e.g. [40]), whose additional drawback is the dependence on antibodies which are available for only a limited number of TFs. Nevertheless, these studies estimated TFs to be expressed in the range of 5,000 to several hundred thousand copies per nucleus[7], which is in line with our results.

Driven by this lack of accurate TF measurements and aided by recent improvements in analytical capacities of MS, we built an SRM-based workflow that combines high sensitivity and technological innovation to generate quantitative data in absolute terms. This workflow, which significantly extends recent efforts to specifically tailor MS methodologies to quantify lowly abundant proteins such as TFs (albeit so far only in relative terms)[28], offers several advantages. *First*, the only requisite of our SRM-based assay is the *in vitro* expression of the protein candidate, which, except for certain types of proteins, becomes relatively trivial given the availability of efficient cloning and expression systems and Open-Reading Frame (ORF) clone libraries[41, 42]. Thus, the assay capitalizes on the well-appreciated advantages of *in vitro* full-length protein expression, including the quick and economical production of nearly any protein of choice (isotopically labeled or not), as well as the ability to limit the accuracy loss linked to the use of single or concatenated peptides[18]. In addition, the ability to nearly extemporaneously produce heavy-labeled full-length protein standards, and to simultaneously quantify both these protein standards and their endogenous counterparts within the same assay bypasses the necessity for protein/peptide storage, another significant source of variability in any quantitative assay[18]. Finally, the elution profiles and ion intensities of the protein standard-specific heavy peptides are compared to those of their non-labeled counterparts, further increasing the protein identification confidence. This is of great importance when dealing with peptide signals of lowly abundant proteins such as TFs. *Second*, the SH-Quant tag used in our assay allows for a robust, sensitive, and reproducible *in situ* quantification of the internal standard protein and by extension allows for accurate dynamic monitoring of endogenous protein copy number fluctuations. *Third*, the introduction of variations in this reference tag enables the analysis of multiple proteins in one single assay. These quantotypic tag variants differ from the original tag at the most by one or two amino acids, thereby minimizing the putatively adverse impact on measurement accuracy which may arise due to significant differences in physico-chemical properties of the peptides. Indeed, the results obtained with the multiplex pipeline demonstrate that the overall accuracy and sensitivity compared to the "one protein at a time" approach is maintained, thus providing a cost-effective, relatively quick, and sensitive strategy to simultaneously quantify several proteins or even entire pathways. Our quantification strategy therefore constitutes a powerful and even cost-effective alternative to quantitative immunoassays (immunoblotting[43] or ELISA) to determine fluctuations in protein abundance over a wide concentration range with the additional advantage of being able to multiplex the assay, something that is much more difficult to do with ELISA[31]. This multiplex capacity should be of great value to those interested in examining or modeling absolute, dynamic changes of entire pathways or sets of biomarkers of interest in wildtype versus clinical, disease, or perturbation settings.

Here, we illustrate the importance of deriving this type of data by building a quantitative model aiming to clarify an intriguing discrepancy between TF binding site and copy number data. As such, we provide a quantitative explanation for the common observation that many TFs, including PPARγ, bind to significantly fewer sites in the genome than predicted based on the presence of their respective consensus motifs[37, 44, 45]. Whereas it was suggested that other factors may contribute to this binding site selectivity, here, we demonstrate that, at least for PPARγ, its DNA binding profile can be closely modeled by simply considering its own copy number, simple thermodynamic principles, and chromatin accessibility. The functional consequence of our findings is that the chromatin state constitutes a landing map for PPARγ DNA binding, consistent with the emerging notion of the importance of chromatin structure in shaping TF DNA binding patterns[45]. This raises the question as to which other factors may control the chromatin state over the course of terminal adipogenesis. One possible candidate is the adipogenic TF C/EBPβ, which was recently shown to mediate chromatin accessibility in this terminal differentiation process[46]. However, several other TFs may also qualify based on the differential enrichment of their respective motifs in adipocyte- compared to pre-adipocyte-specific H3K27Ac sites[37] and thus the precise identity of the responsible TF(s) remains to be elucidated.

In summary, our work represents an important effort to expand the analytical and practical capabilities of targeted proteomics approaches while safeguarding measurement accuracy. In a next phase, we look forward to further expanding the number of reference-peptide variants to boost the assay's multiplex capacity as part of the overall aim to elucidate and model the dynamic properties of entire biological processes or networks under different experimental conditions.

ACKNOWLEDGEMENTS

We would like to thank Marco Lai (Thermo Fischer Scientific) for his help on solving Pinpoint software-related issues; Matthias Gstaiger (ETHZ) for providing us with the SH-Quant peptide tag; Erik Ahrné (University of Basel) for computational assistance and providing feedback on the manuscript together with Paola Gilardoni and Sebastian Waszak (EPFL), Markus Mueller (SIB, University of Geneva), and Alex Schmidt (University of Basel); and Diego Chiappe, Romain Hamelin, Florence Armand (all EPFL) for technical and computational assistance. This work has been supported by funds from the Swiss National Science Foundation, by a Marie Curie International Reintegration Grant (to B.D.) from the Seventh Research Framework Programme, by a SystemsX.ch iPhD Fellowship (to J.S.), by the Roland Bailly Foundation (Geneva, CH), by the NCCR Frontiers in Genetics Program, and by Institutional support from the Ecole Polytechnique Fédérale de Lausanne (EPFL).

Author Contributions

Conceived and planned the study: J.S., A.S., M.M., B.D. Prepared the manuscript: J.S., A.S., B.Z., F.N., M.M., B.D. Performed wet bench experiments: J.S., S.R., C.G. Mass spectrometry data analysis: J.S., A.S. Modeling: B.Z., I.K., F.N., B.D. All the authors discussed the results and commented on the paper.

3T3-L1 cell culture, differentiation, and protein extraction

3T3-L1 cells were cultured and differentiated into adipocytes as detailed in Raghav *et al.*,[23] (**Supplementary Methods**).

Cloning and plasmids

Mouse TF ORFs were cloned in Gateway format essentially as described earlier[42]. To make the wheat germ (WG) *in vitro* transcription-translation expression system (Promega) compatible with the mouse TF ORF clones and to allow the purification of TFs, we modified the SP6 pF3A WG (BYDV) Flexi vector (Promega) to accommodate the Gateway reading frame A cassette (Invitrogen). A glutathione-S-transferase (GST) coding sequence containing a stop codon was subsequently incorporated in frame at the 3'-end of the second (i.e. 3') Gateway site using standard cloning techniques. TFs were then subcloned from the Gateway entry clone level into this "in house" modified pF3A-GST destination vector by standard Gateway cloning (**Supplementary Methods**).

In vitro protein expression and purification, gel staining, and immunoblotting

For selection of optimal TF transitions: the 10 selected TFs containing a GST tag at their C-terminus were expressed *in vitro* via the TnT SP6 High Yield Wheat Germ Protein Expression System (Promega) according to the manufacturer's protocol. GST-labeled TFs were subsequently purified via glutathione sepharose 4B beads (GE Healthcare) according to the manufacturer's protocol (**Supplementary Methods**). Purified proteins were separated by SDS-PAGE in a 10% resolving gel (Tris-Glycine) and stained either with silver nitrate or with Coomassie blue (SimplyBlue Safestain, Invitrogen) for band visualization. Validation of *in vitro*-expressed proteins was performed by Western blot (**Supplementary Methods**).

For TF quantification: the 10 selected TFs were expressed *in vitro* via the TnT SP6 Wheat Germ High Yield Master Mix Minus Amino Acids (Promega) in their isotopically-labeled version. 18 amino acids were mixed to reach a 80µM final concentration and added to the reaction together with isotopically-labeled Arginine (^{13}C,^{15}N-Arg) and Lysine (^{13}C,^{15}N-Lys)(Cambridge Isotope Laboratories) to a 1mM final concentration prior to resorting to the conventional expression protocol. Depending on the analyzed TF, 3X reactions were utilized (or 4X for TFs with low expression yields). Subsequently, the purification and fractionation techniques described above were applied.

Reference peptide tag incorporation

The SH-Quant peptide (AADITSLYK)-coding sequence along with the segment of the 5’ Gateway site comprised between the *Pvu*I and *Pvu*II restriction sites was synthesized by Geneart-Life Technologies. An *Spe*I restriction site was added at the 3’ of the tag-coding sequence to simplify the insertion of tag variants. The SH-Quant insert was introduced in the *Pvu*I and *Pvu*II restriction sites at the 5’ Gateway site of the pF3A-GST vector by double digestion with the two restriction enzymes and subsequently ligated to obtain the SH-pF3A-GST destination vector.

Reference peptide tag variants design for multiplexing

The additional tags 2 to 5 described in Fig. 5 were designed by a first permutation of Ile residue 4 from the original SH-Quant tag sequence with amino acids of different hydrophobicity and size (G, A, V, F) to introduce molecular weight and/or retention time differences while not altering the fragmentation pattern. Additional tags 6 to 10 were generated by a second permutation conservative of the charge by replacing Asp residue 3 by Glu in the 5 sequences generated as above. Hydrophilic residues Thr 5 and Tyr 8 which produce intense y” fragment ions were kept constant to maintain solubility and sensitivity. The IN VITRO expression vector SH-pF3A-GST was modified via site-directed mutagenesis (QuikChange Lightning, Stratagene) according to the manufacturer’s protocol to generate the SH-quant peptide variant-incorporating vectors (SH$_i$-pF3A-GST). The I4G, I4A, I4V, I4F, D3E, D3E- I4G, D3E-I4A, D3E-I4V, and D3E-I4F substitutions were introduced into SH-pF3A-GST using the forward and reverse primers presented in **Supplementary Table 4**, and sequence-verified. Proteotypicity of the newly designed tags was confirmed with Blast-P against the mouse protein database (http://blast.ncbi.nlm.nih.gov).

In vitro-expressed TF identification

In vitro-expressed proteins were in-gel digested using trypsin as follows: the samples were reduced in 10 mmol/l dithioerythritol (DTE) and alkylated in 55 mmol/l iodoacetamide (IAA) before being dried. The samples were then incubated with 12.5 ng/ml of trypsin overnight at 37°C. The resulting tryptic peptides were extracted from the gel slices, dried, and resuspended in 2% acetonitrile (ACN): 0.1% formic acid (FA) for LC-MS/MS analysis. A mass spectrometer (LTQ Orbitrap XL (Thermo Fisher Scientific) equipped with an ultraperformance LC (UPLC) system (nanoACQUITY; Waters) was used

(**Supplementary Methods**). Data search was performed using Mascot software (v. 2.3; Matrix Science) and Proteome Discoverer (v.1.3; Thermo Fisher Scientific) (**Supplementary Methods**).

Proteotypic peptide selection

The procedure to select optimal TF proteotypic peptide candidates is explained in the **Supplementary Methods**.

Nuclear extract sample preparation and spiking of SH-Quant reference tags

Typically, 400 µg (quantified using the BSA assay) of total protein nuclear extract (differentiation times: Day 0 to Day 6) were separated and fractionated by SDS-PAGE using 10% Tris-Glycine gels. Gel bands were excised around the migration height of the selected proteins: 50-60 kDa for RXRα and PPARγ, 30-40 kDa and 45-70 kDa for the multiplex analysis (all 10 TFs). Gel fractions were first reduced and alkylated using 10 mM dithioerythritol and 55mM iodoacetamide, followed by gel drying using speed vacuum. Gel fractions were re-suspended in 50 mM ammonium bicarbonate (pH 8.3) digestion buffer, containing trypsin (Trypsin-Gold, Promega) at a final concentration of 12 ng/µl. The accurately quantified SH-Quant tags (JPT Peptide Technologies) were added to each tube (200 fmol/tag) immediately after re-suspension in the digestion buffer. Digestion was performed overnight at 37 °C. Following digestion, peptides were extracted from gels and accurately aliquoted into four equal volumes (4x 100 µl) followed by sample drying using speed vacuum and stored at -20°C prior to analysis.

Liquid chromatography-mass spectrometry (LC-MS) analysis

Dried aliquots (theoretical concentration: 100 µg) were re-suspended in 20 µl LC-MS loading solvent (2% acetonitrile, 0.1% formic acid) yielding a final, theoretical concentration of nuclear extract of 5 µg/µl. Following re-suspension, samples were allowed to settle for 1h to increase overall peptide solubility prior to analysis. Typically, 5 µl of sample (theoretical total: 25 µg) was loaded and captured on a home-made capillary pre-column (Magic C18; 3 µm-200Å; 2 cm × 250 µm) prior to analytical LC separation (nanoACQUITY UPLC, Waters). Samples were separated using a 90-minute biphasic gradient starting from 100% A solvent (2% acetonitrile, 0.1% formic acid) to 90% B solvent (100% acetonitrile, 0.1% formic acid) on a Nikkyo (Nikkyo Technologies) nano-column (C18; 3 µm-100 Å; 15 cm length and 100 µm inner diameter, at a flow of 1 µl/min). The gradient was followed by a wash for 8 min at 90% solvent B and column re-equilibration of 12 min at 100% solvent A.

Selected reaction monitoring (SRM) of TFs

All samples were analyzed on a TSQ-Vantage triple-quadrupole mass spectrometer (Thermo Fisher, Scientific). A 0.7 FWHM resolution window for both Q1 and Q3 was set for parent- and product-ion isolation. Fragmentation of parent-ions was performed in Q2 at 1.5 mTorr, using collision energies calculated with the Pinpoint software (v1.1). Parent-ion selection was set for fully digested peptides on the doubly charged ion for both SH-Quant tags and target proteins. Generally, singly charged peptide fragment ions ranging from y''_3 to y''_{n-1} were preferably monitored, unless otherwise stated. A complete list of all monitored transitions is provided in **Supplementary Table 5**. A parent-ion retention time target window of 2 min (single quantification of RXRα and PPARγ) and 5 min (multiplexed quantification of ten TFs) was set for Q1 during a scheduled SRM run. A total of respectively 88 and 76 transitions were monitored for RXRα and PPARγ in the single protein quantification strategy, whereas a total of 492 transitions were typically monitored during a multiplex run. A cycle time of 0.5 s and 2 s was used for single and multiplex SRM runs respectively. A minimum dwell time of 10 ms or more was set for both single and multiplex SRM assays. Samples were analyzed as three biological and three technical replicates (n=3) in RXRα and PPARγ single protein quantification assays. One biological replicate was analyzed in 6 technical replicates as a proof-of-concept of the multiplex SRM assay.

Data analysis and absolute quantification of TFs

Calculation of absolute levels of TFs was performed using a two-step procedure as outlined in **Figure 2**. All data analyses were carried out using Pinpoint software (v.1.1). Peptide identification and peak area integration of all SH-Quant tags and targeted peptides as well as their transitions were verified manually in Pinpoint. Ten individual standard curves were established for all SH-Quant tags using a concentration range of six values recorded in three technical replicates (**Supplementary Fig. 10**). The performance-based definitions of lower limits of quantification (LLOQ) were applied as defined by the FDA (Guidance for Industry: Bioanalytical Method Validation: *http://www.fda.gov/cder/guidance/index.htm*). Briefly, the threshold for the limit of detection (LOD) was set at 20% of the CV, whereas the LLOQ was set at 15% of the CV using three technical replicates for single- and six technical replicates for multiplex quantification approaches. An example of the LLOQ found for an SH-Tag is provided in **Supplementary Figure 11**. An analysis of variance (ANOVA) coupled with a Neuwman-Keuls post-hoc test was applied for calculating the differences in endogenous levels of TFs between different differentiation time points, where the level of $P<0.05$ was considered as statistically significant. The calculation steps to derive absolute copy numbers per cell (nucleus) are presented in **Supplementary Table 3**.

Motif Analysis

Regions enriched with H3K27ac as determined by Mikkelsen *et al.,*[37] were scanned with FIMO[47] allowing for a *P* value of 10^{-3}. Each match in those regions was then scored for log-likelihood given by the position weight matrix of the PPARγ motif from the JASPAR CORE database[48], assuming uniform background.

PPARγ ChIP-seq data reanalysis

ChIP-seq data from Nielsen *et al*. ([34]) was processed as in Raghav *et al.*[23].

Quantitative model of genome-wide TF DNA binding

The estimation of the PPARγ DNA binding profile is based on equilibrium thermodynamics and is described in detail in the **Supplementary Note**.

REFERENCES

1. Simicevic, J. & Deplancke, B. DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. *Molecular bioSystems* 6, 462-468 (2010).

2. Kim, H.D., Shay, T., O'Shea, E.K. & Regev, A. Transcriptional Regulatory Circuits: Predicting Numbers from Alphabets. *Science* 325, 429-432 (2009).

3. Bussemaker, H.J., Foat, B.C. & Ward, L.D. Predictive Modeling of Genome-Wide mRNA Expression: From Modules to Molecules. *Annual Review of Biophysics and Biomolecular Structure* 36, 329-347 (2007).

4. Segal, E. & Widom, J. From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat Rev Genet* 10, 443-456 (2009).

5. Meng, J., Zhang, J.M., Chen, Y. & Huang, Y. Bayesian non-negative factor analysis for reconstructing transcription factor mediated regulatory networks. *Proteome science* 9 Suppl 1, S9 (2011).

6. Stormo, G.D. & Zhao, Y. Determining the specificity of protein–DNA interactions. *Nat Rev Genet* 11, 751-760 (2010).

7. Biggin, Mark D. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Developmental Cell* 21, 611-626 (2011).

8. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nature reviews. Genetics* 10, 252-263 (2009).

9. Zeiler, M., Straube, W.L., Lundberg, E., Uhlen, M. & Mann, M. A protein epitope signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & cellular proteomics : MCP* (2011).

10. Beck, M. et al. The quantitative proteome of a human cell line. *Molecular systems biology* 7, 549 (2011).

11. Costenoble, R. et al. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Molecular systems biology* 7, 464 (2011).

12. Picotti, P. et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature methods* 7, 43-46 (2010).

13. Craig, R., Cortens, J.P. & Beavis, R.C. The use of proteotypic peptide libraries for protein identification. *Rapid communications in mass spectrometry : RCM* 19, 1844-1850 (2005).

14. Singh, S., Springer, M., Steen, J., Kirschner, M.W. & Steen, H. FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. *Journal of proteome research* 8, 2201-2210 (2009).

15. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R. & Gstaiger, M. Quantitative interaction proteomics using mass spectrometry. *Nature methods* 6, 203-205 (2009).

16. Proc, J.L. et al. A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *Journal of proteome research* 9, 5422-5437 (2010).

17. Kuhn, E. et al. Inter-laboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Molecular & cellular proteomics : MCP* (2011).

18. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods* 9, 555-566 (2012).

19. Choi, J.H. et al. Antidiabetic actions of a non-agonist PPARgamma ligand blocking Cdk5-mediated phosphorylation. *Nature* 477, 477-481 (2011).

20. Ahmed, M., Neville, M.J., Edelmann, M.J., Kessler, B.M. & Karpe, F. Proteomic analysis of human adipose tissue after rosiglitazone treatment shows coordinated changes to promote glucose uptake. *Obesity* 18, 27-34 (2010).

21. Molina, H. et al. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *Journal of proteome research* 8, 48-58 (2009).

22. Rosen, E.D. & MacDougald, O.A. Adipocyte differentiation from the inside out. *Nature reviews. Molecular cell biology* 7, 885-896 (2006).

23. Raghav, S.K. et al. Integrative Genomics Identifies the Corepressor SMRT as a Gatekeeper of Adipogenesis through the Transcription Factors C/EBPbeta and KAISO. *Molecular cell* (2012).

24. Desiere, F. et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome biology* 6, R9 (2005).

25. Rosen, E.D. & MacDougald, O.A. Adipocyte differentiation from the inside out. *Nat Rev Mol Cell Biol* 7, 885-896 (2006).

26. Raghav, Sunil K. et al. Integrative Genomics Identifies the Corepressor SMRT as a Gatekeeper of Adipogenesis through the Transcription Factors C/EBPβ and KAISO. *Molecular Cell* 46, 335-350 (2012).

27. Thakur, S.S. et al. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular & cellular proteomics : MCP* 10, M110 003699 (2011).

28. Stergachis, A.B., Maclean, B., Lee, K., Stamatoyannopoulos, J.A. & Maccoss, M.J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nature methods* 8, 1041-1043 (2011).

29. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology* 4, 222 (2008).

30. Gevaert, K. et al. Stable isotopic labeling in proteomics. *Proteomics* 8, 4873-4885 (2008).

31. Whiteaker, J.R. et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotech* 29, 625-634 (2011).

32. Mirzaei, H., McBee, J.K., Watts, J. & Aebersold, R. Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics. *Molecular & cellular proteomics : MCP* 7, 813-823 (2008).

33. Dupuis, A., Hennekinne, J.A., Garin, J. & Brun, V. Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics* 8, 4633-4636 (2008).

34. Nielsen, R. et al. Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes & development* 22, 2953-2967 (2008).

35. Waki, H. et al. Global Mapping of Cell Type–Specific Open Chromatin by FAIRE-seq Reveals the Regulatory Role of the NFI Family in Adipocyte Differentiation. *PLoS genetics* 7, e1002311 (2011).

36. Cotney, J. et al. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Research* (2012).

37. Mikkelsen, T.S. et al. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 143, 156-169 (2010).

38. Rey, G. et al. Genome-Wide and Phase-Specific DNA-Binding Rhythms of BMAL1 Control Circadian Output Functions in Mouse Liver. *PLoS Biol* 9, e1000595 (2011).

39. Gaudet, J. & Mango, S.E. Regulation of Organogenesis by the Caenorhabditis elegans FoxA Protein PHA-4. *Science* 295, 821-825 (2002).

40. Bradley, M.N., Zhou, L. & Smale, S.T. C/EBPβ Regulation in Lipopolysaccharide-Stimulated Macrophages. *Molecular and Cellular Biology* 23, 4841-4858 (2003).

41. Lamesch, P. et al. hORFeome v3.1: A resource of human open reading frames representing over 10,000 human genes. *Genomics* 89, 307-315 (2007).

42. Hens, K. et al. Automated protein-DNA interaction screening of Drosophila regulatory elements. *Nat Meth* 8, 1065-1070 (2011).

43. Thuillier, P., Baillie, R., Sha, X. & Clarke, S.D. Cytosolic and nuclear distribution of PPARgamma2 in differentiating 3T3-L1 preadipocytes. *Journal of lipid research* 39, 2329-2338 (1998).

44. Farnham, P.J. Insights from genomic profiling of transcription factors. *Nat Rev Genet* 10, 605-616 (2009).

45. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 43, 264-268 (2011).

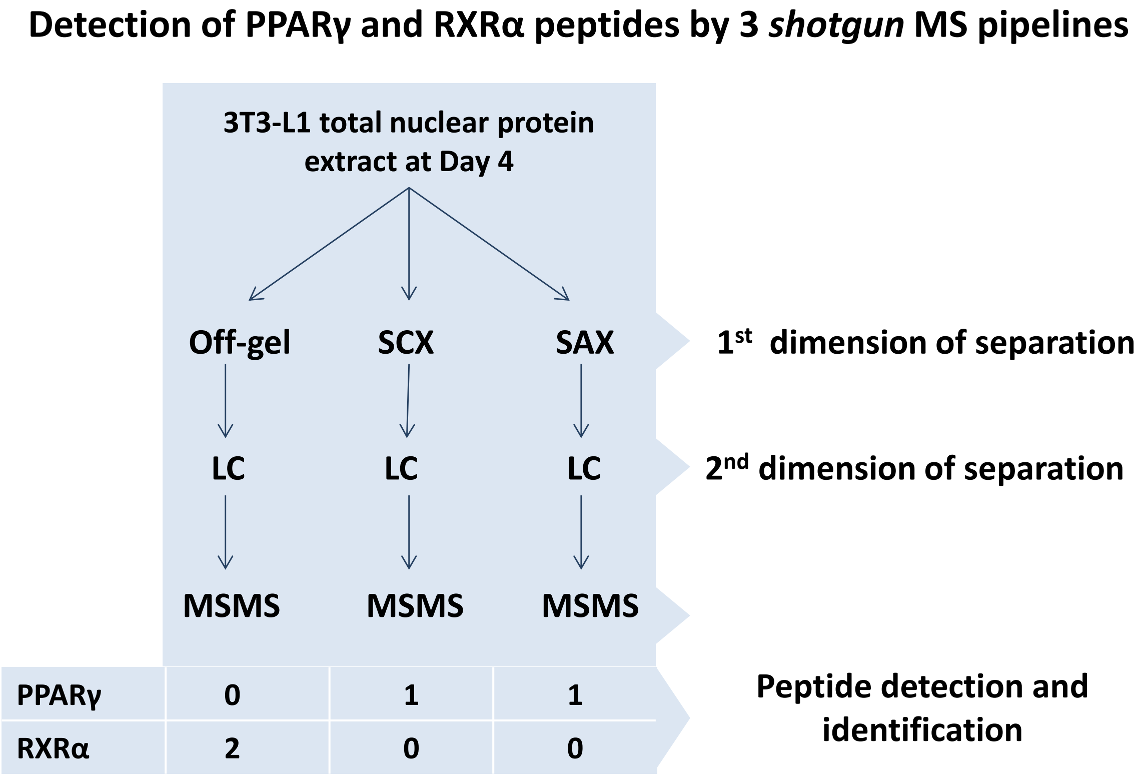
46. Siersbaek, R. et al. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *The EMBO journal* 30, 1459-1472 (2011).

47. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017-1018 (2011).

48. Vlieghe, D. et al. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic acids research* 34, D95-97 (2006).

Absolute copy number analysis of transcription factors during cellular differentiation using a multiplex, targeted proteomics approach

Supplementary Figures



Detection of PPARγ and RXRα peptides by 3 *shotgun* MS pipelines

3T3-L1 total nuclear protein extract at Day 4

Off-gel → LC → MSMS

SCX → LC → MSMS

SAX → LC → MSMS

1st dimension of separation

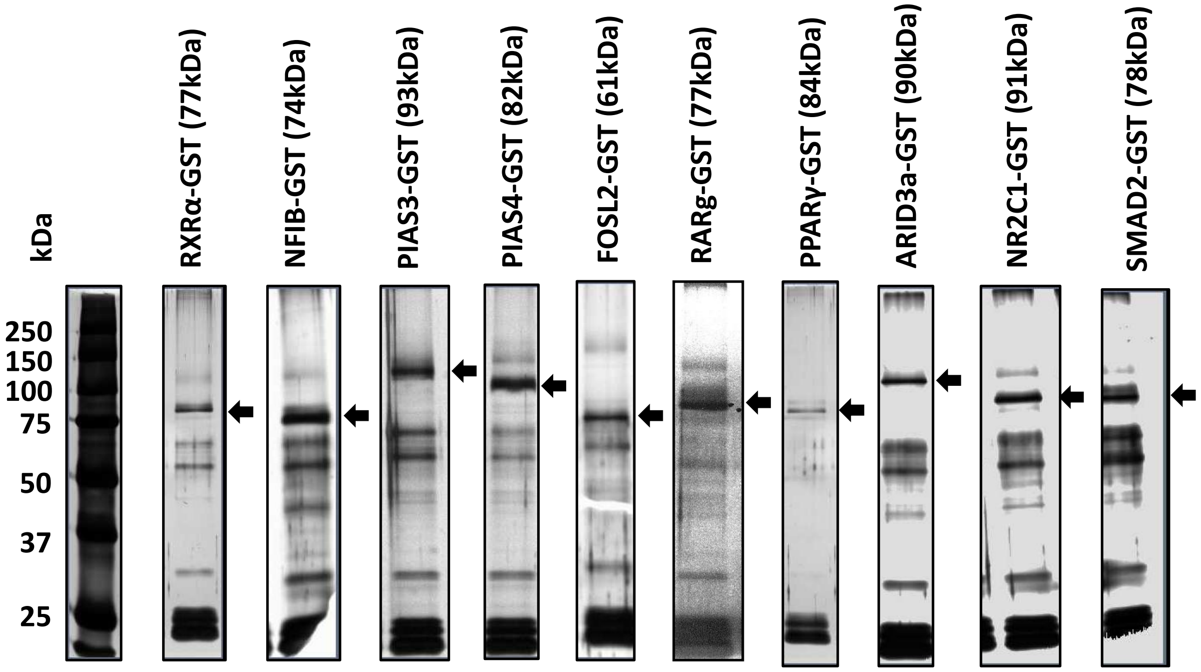
2nd dimension of separation

Peptide detection and identification

	Off-gel	SCX	SAX
PPARγ	0	1	1
RXRα	2	0	0

Supplementary Figure 1. Diagram representing the 3 *shotgun* MS pipelines tested utilizing 3T3-L1 total nuclear protein extract at Day 4 of differentiation.

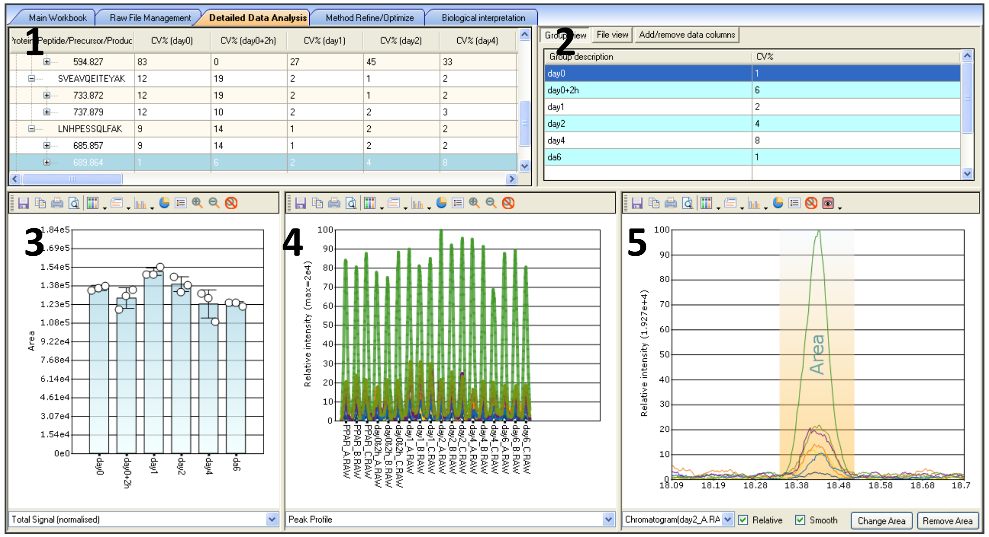
Each pipeline employs a different peptide separation technique in the first dimension: SAX (strong anion exchange, 6 fractions), SCX (strong cation exchange, 20 fractions), and separation of peptides by their isoelectric point using an “Off-gel” electrophoretic system (24 fractions). The peptides were subsequently submitted to LC-MSMS for identification. The two adipogenic master regulators PPARγ and RXRα were specifically targeted. A maximum of only 2 peptides per TF were ultimately observed.



Supplementary Figure 2. 10 *in vitro*-expressed GST-tagged TFs.
The 10 *in vitro*-expressed TFs were run on a denaturing SDS-PAGE gel stained with silver nitrate. The molecular weights of the TF constructs (molecular weight of the TF + molecular weight of the GST-tag (26 kDa)) are presented in parentheses and their respective bands are indicated by an arrow. Peptide detection and protein identification for each TF construct was performed by in-gel digestion and LC-MS analysis.

PPARγ peptide: LNHPESSQLFAK

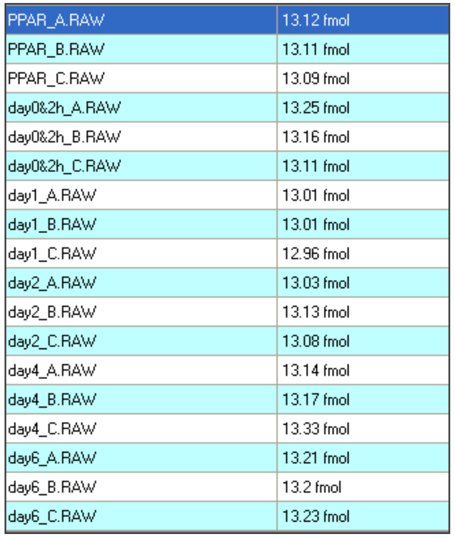
a



'rotein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
594.827	83	0	27	45	33
SVEAVQEITEYAK	12	19	2	1	2
733.872	12	19	2	1	2
737.879	12	10	2	2	3
LNHPESSQLFAK	9	14	1	2	2
685.857	9	14	1	2	2
689.864	1	6	2	4	8

Group description	CV%
day0	1
day0+2h	6
day1	2
day2	4
day4	8
da6	1

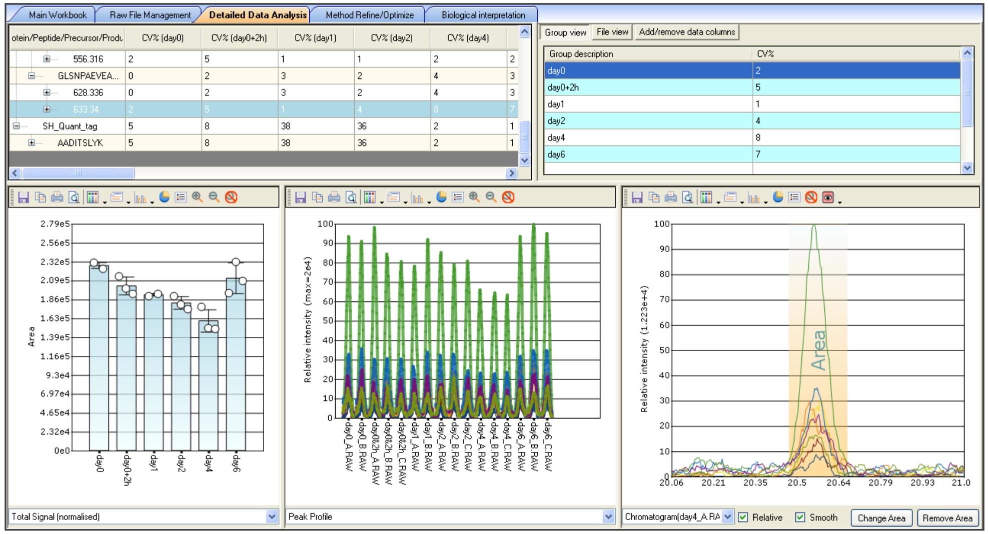
b



PPAR_A.RAW	13.12 fmol
PPAR_B.RAW	13.11 fmol
PPAR_C.RAW	13.09 fmol
day0&2h_A.RAW	13.25 fmol
day0&2h_B.RAW	13.16 fmol
day0&2h_C.RAW	13.11 fmol
day1_A.RAW	13.01 fmol
day1_B.RAW	13.01 fmol
day1_C.RAW	12.96 fmol
day2_A.RAW	13.03 fmol
day2_B.RAW	13.13 fmol
day2_C.RAW	13.08 fmol
day4_A.RAW	13.14 fmol
day4_B.RAW	13.17 fmol
day4_C.RAW	13.33 fmol
day6_A.RAW	13.21 fmol
day6_B.RAW	13.2 fmol
day6_C.RAW	13.23 fmol

RXRα peptide: GLSNPAEALR

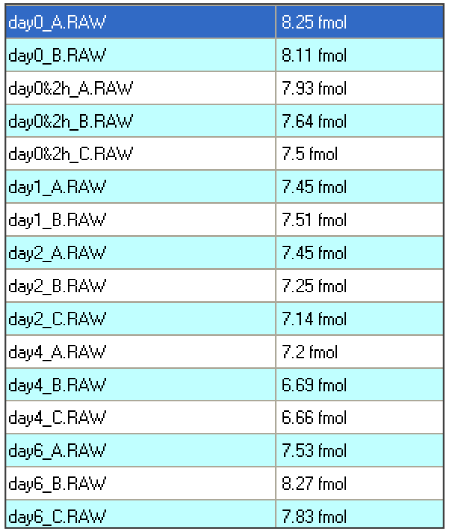
c



otein/Peptide/Precursor/Produ	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)	
556.316	2	5	1	1	2	2
GLSNPAEVEA...	0	2	3	2	4	3
628.336	0	2	3	2	4	3
633.34	2	5	1	4	8	7
SH_Quant_tag	5	8	38	36	2	1
AADITSLYK	5	8	38	36	2	1

Group description	CV%
day0	2
day0+2h	5
day1	1
day2	4
day4	8
day6	7

d

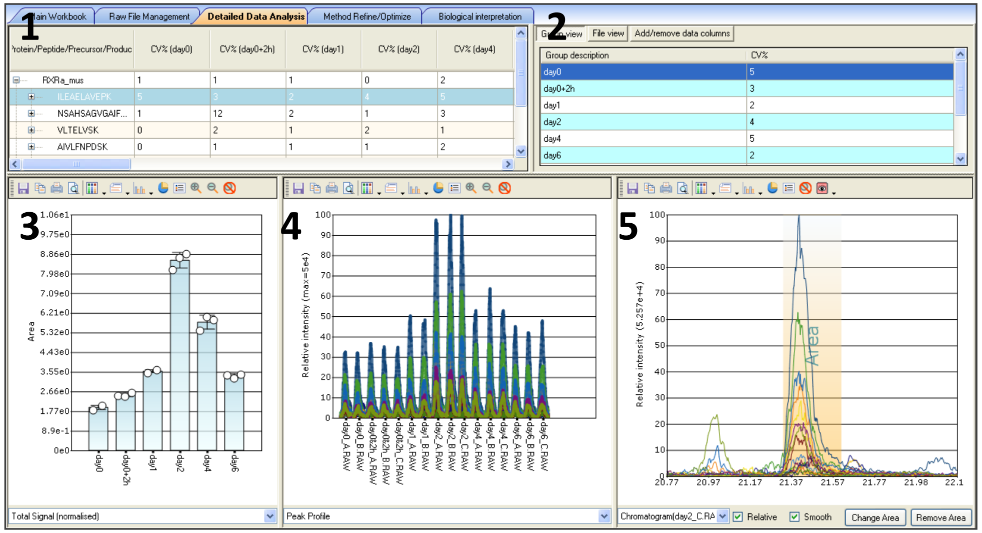


day0_A.RAW	8.25 fmol
day0_B.RAW	8.11 fmol
day0&2h_A.RAW	7.93 fmol
day0&2h_B.RAW	7.64 fmol
day0&2h_C.RAW	7.5 fmol
day1_A.RAW	7.45 fmol
day1_B.RAW	7.51 fmol
day2_A.RAW	7.45 fmol
day2_B.RAW	7.25 fmol
day2_C.RAW	7.14 fmol
day4_A.RAW	7.2 fmol
day4_B.RAW	6.69 fmol
day4_C.RAW	6.66 fmol
day6_A.RAW	7.53 fmol
day6_B.RAW	8.27 fmol
day6_C.RAW	7.83 fmol

Supplementary Figure 3. Evaluation of heavy PPARγ "LNHPESSQLFAK" and RXRα "GLSNPAEALR" peptides in 3T3-L1 nuclear extract.
Pinpoint screen prints, showing an example of the evaluation of heavy PPARγ (a) (LNHPESSQLFAK) and RXRα (c) (GLSNPAEALR) peptides, monitored at Day 0 through Day 6. The Pinpoint window provides a summary of the evaluation of the peptide of interest. Subpanel 1 shows an overview of all peptides monitored. Subpanel 2 shows the coefficients of variation per peptide at the different time points. Subpanel 3 shows the total normalized signal of all technical replicates (n=3). Subpanel 4 shows the peak profiles of all technical replicates. Subpanel 5 shows the chromatogram and calculated peak area of peptide transitions. Panel legend and numbering applies to panel (c) as well. Panels b and d show the calculated values of the PPARγ and RXRα heavy peptides spiked into nuclear extracts at Day 0 through Day 6.

RXRα peptide: ILEAELAVEPK

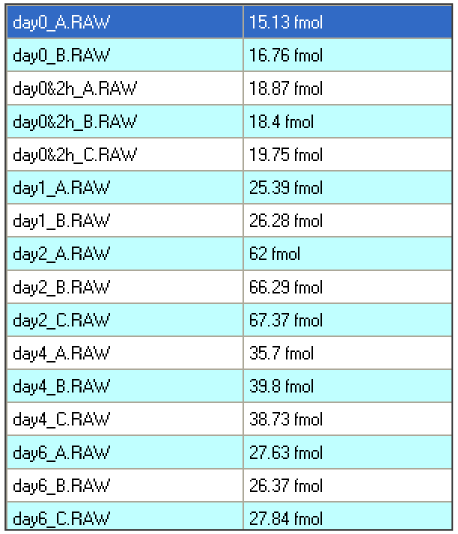
a



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
RXRa_mus	1	1	1	0	2
ILEAELAVEPK	5	3	2	4	5
NSAHSAGVGAIF...	1	12	2	1	3
VLTELVSK	0	2	1	2	1
AIVLFNPDSK	0	1	1	1	2

Group description	CV%
day0	5
day0+2h	3
day1	2
day2	4
day4	5
day6	2

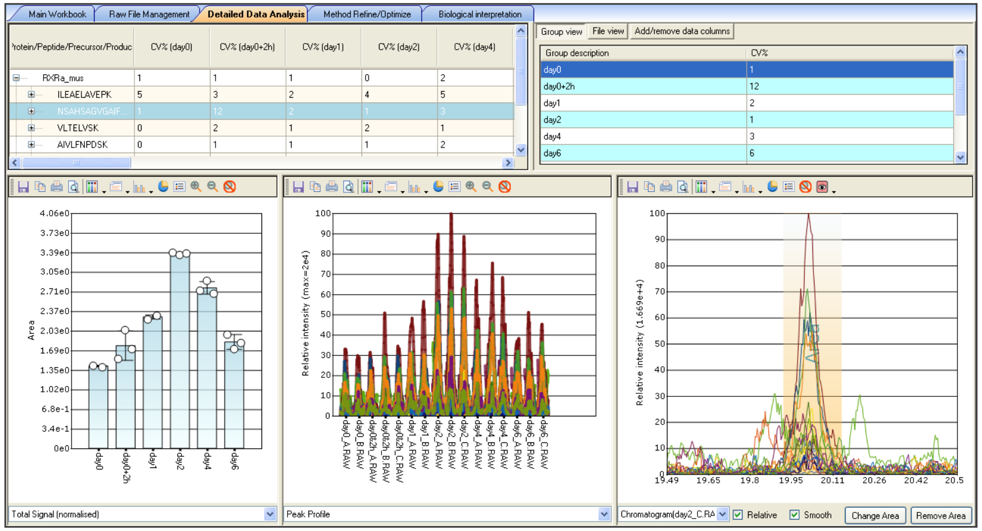
b



day0_A.RAW	15.13 fmol
day0_B.RAW	16.76 fmol
day0&2h_A.RAW	18.87 fmol
day0&2h_B.RAW	18.4 fmol
day0&2h_C.RAW	19.75 fmol
day1_A.RAW	25.39 fmol
day1_B.RAW	26.28 fmol
day2_A.RAW	62 fmol
day2_B.RAW	66.29 fmol
day2_C.RAW	67.37 fmol
day4_A.RAW	35.7 fmol
day4_B.RAW	39.8 fmol
day4_C.RAW	38.73 fmol
day6_A.RAW	27.63 fmol
day6_B.RAW	26.37 fmol
day6_C.RAW	27.84 fmol

RXRα peptide: NSAHSAGVGAIF

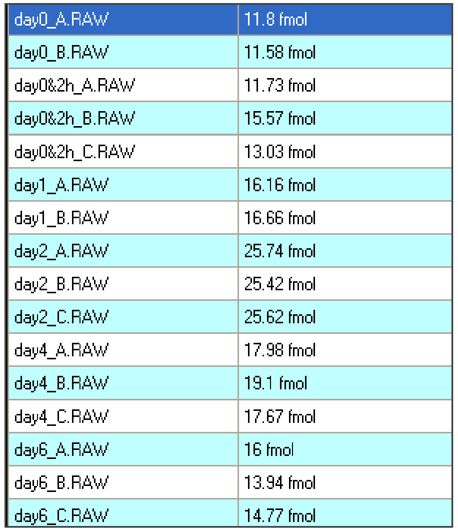
c



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
RXRa_mus	1	1	1	0	2
ILEAELAVEPK	5	3	2	4	5
NSAHSAGVGAIF...	1	12	2	1	3
VLTELVSK	0	2	1	2	1
AIVLFNPDSK	0	1	1	1	2

Group description	CV%
day0	1
day0+2h	12
day1	2
day2	1
day4	3
day6	6

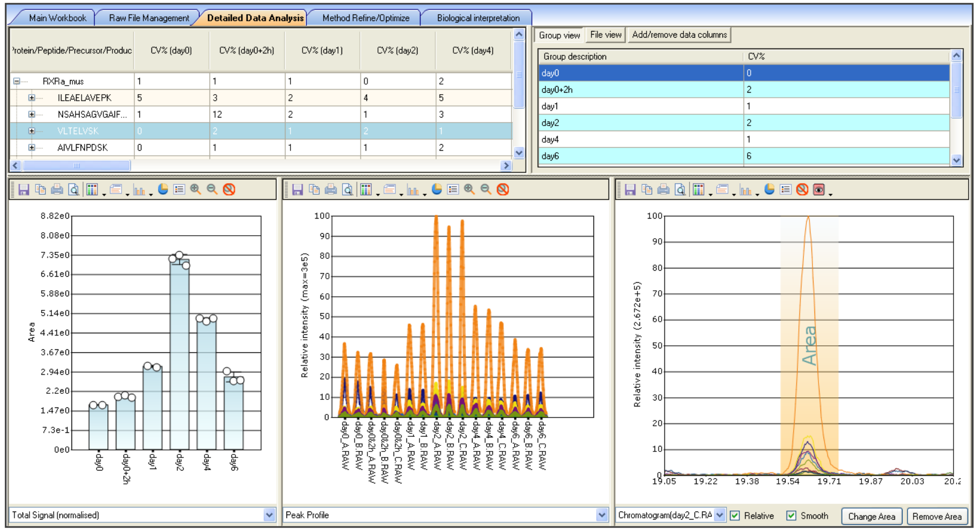
d



day0_A.RAW	11.8 fmol
day0_B.RAW	11.58 fmol
day0&2h_A.RAW	11.73 fmol
day0&2h_B.RAW	15.57 fmol
day0&2h_C.RAW	13.03 fmol
day1_A.RAW	16.16 fmol
day1_B.RAW	16.66 fmol
day2_A.RAW	25.74 fmol
day2_B.RAW	25.42 fmol
day2_C.RAW	25.62 fmol
day4_A.RAW	17.98 fmol
day4_B.RAW	19.1 fmol
day4_C.RAW	17.67 fmol
day6_A.RAW	16 fmol
day6_B.RAW	13.94 fmol
day6_C.RAW	14.77 fmol

e

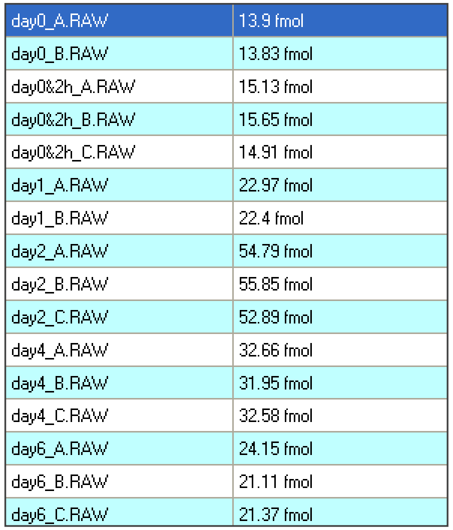
RXRα peptide: VLTELVSK



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
RXRa_mus	1	1	1	0	2
ILEAELAVEPK	5	3	2	4	5
NSAHSAGVGAIF...	1	12	2	1	3
VLTELVSK	0	2	1	2	1
AIVLFNPDSK	0	1	1	1	2

Group description	CV%
day0	0
day0+2h	2
day1	1
day2	2
day4	1
day6	6

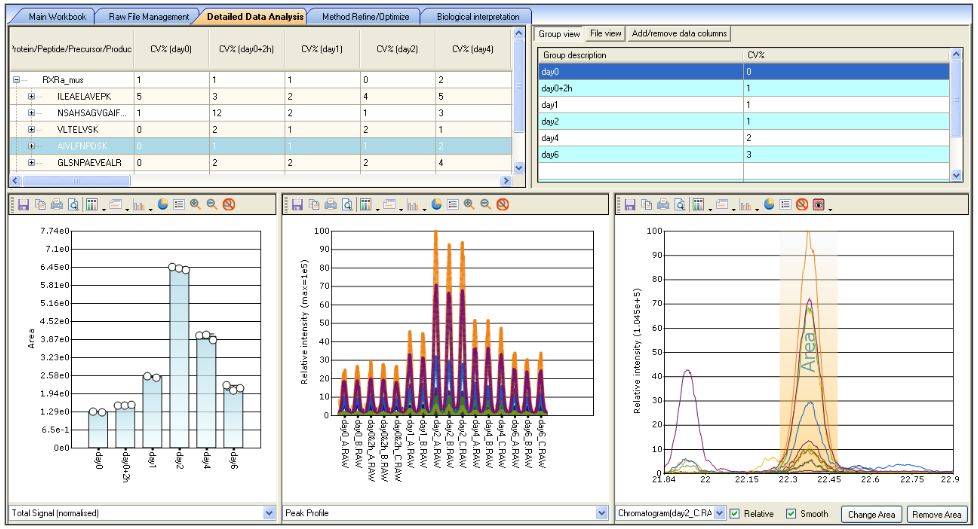
f



day0_A.RAW	13.9 fmol
day0_B.RAW	13.83 fmol
day0&2h_A.RAW	15.13 fmol
day0&2h_B.RAW	15.65 fmol
day0&2h_C.RAW	14.91 fmol
day1_A.RAW	22.97 fmol
day1_B.RAW	22.4 fmol
day2_A.RAW	54.79 fmol
day2_B.RAW	55.85 fmol
day2_C.RAW	52.89 fmol
day4_A.RAW	32.66 fmol
day4_B.RAW	31.95 fmol
day4_C.RAW	32.58 fmol
day6_A.RAW	24.15 fmol
day6_B.RAW	21.11 fmol
day6_C.RAW	21.37 fmol

g

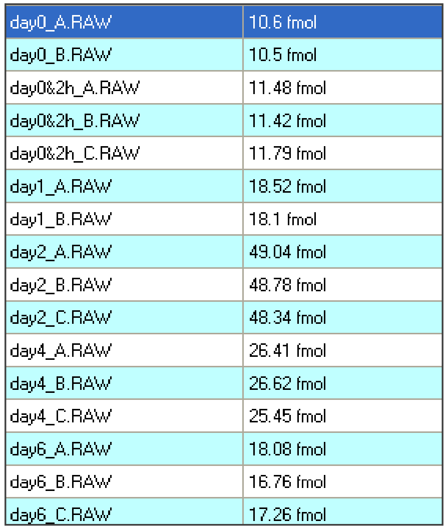
RXRα peptide: AIVLFNPDSK



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
RXRa_mus	1	1	1	0	2
ILEAELAVEPK	5	3	2	4	5
NSAHSAGVGAIF...	1	12	2	1	3
VLTELVSK	0	2	1	2	1
AIVLFNPDSK	0	1	1	1	2
GLSNPAEVEALR	0	2	2	2	4

Group description	CV%
day0	0
day0+2h	1
day1	1
day2	1
day4	2
day6	3

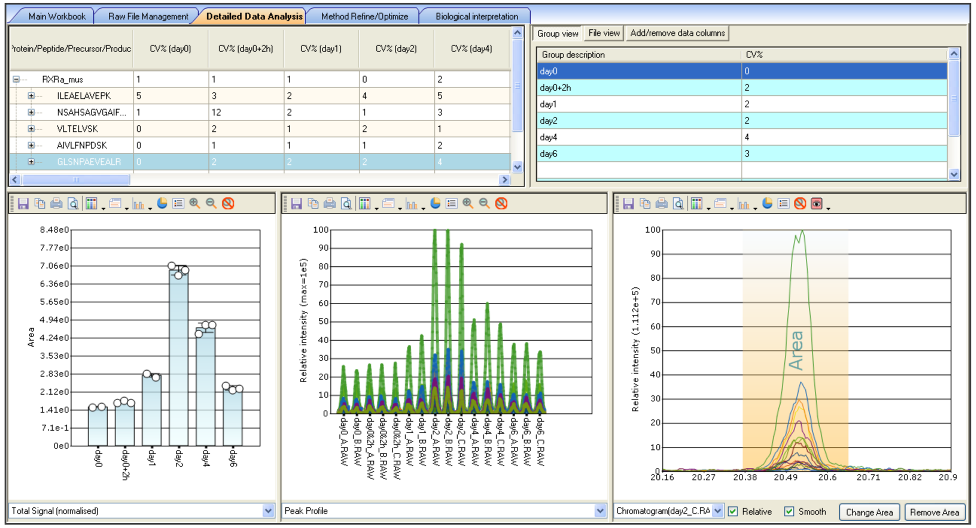
h



day0_A.RAW	10.6 fmol
day0_B.RAW	10.5 fmol
day0&2h_A.RAW	11.48 fmol
day0&2h_B.RAW	11.42 fmol
day0&2h_C.RAW	11.79 fmol
day1_A.RAW	18.52 fmol
day1_B.RAW	18.1 fmol
day2_A.RAW	49.04 fmol
day2_B.RAW	48.78 fmol
day2_C.RAW	48.34 fmol
day4_A.RAW	26.41 fmol
day4_B.RAW	26.62 fmol
day4_C.RAW	25.45 fmol
day6_A.RAW	18.08 fmol
day6_B.RAW	16.76 fmol
day6_C.RAW	17.26 fmol

i

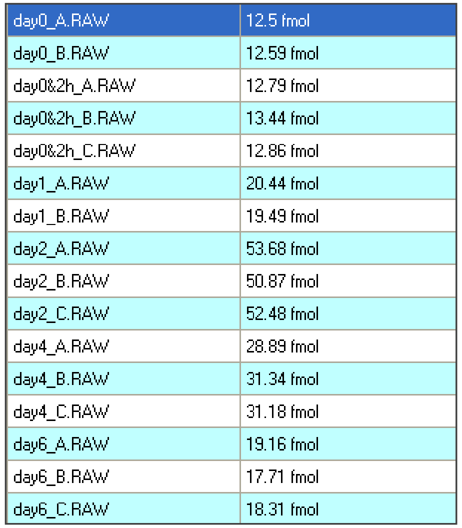
RXRα peptide: GLSNPAEVEALR



›rotein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
⊟ RXRa_mus	1	1	1	0	2
⊞ ILEAELAVEPK	5	3	2	4	5
⊞ NSAHSAGVGAIF...	1	12	2	1	3
⊞ VLTELVSK	0	2	1	2	1
⊞ AIVLFNPDSK	0	1	1	1	2
⊞ GLSNPAEVEALR	0	2	2	2	4

Group description	CV%
day0	0
day0+2h	2
day1	2
day2	2
day4	4
day6	3

j

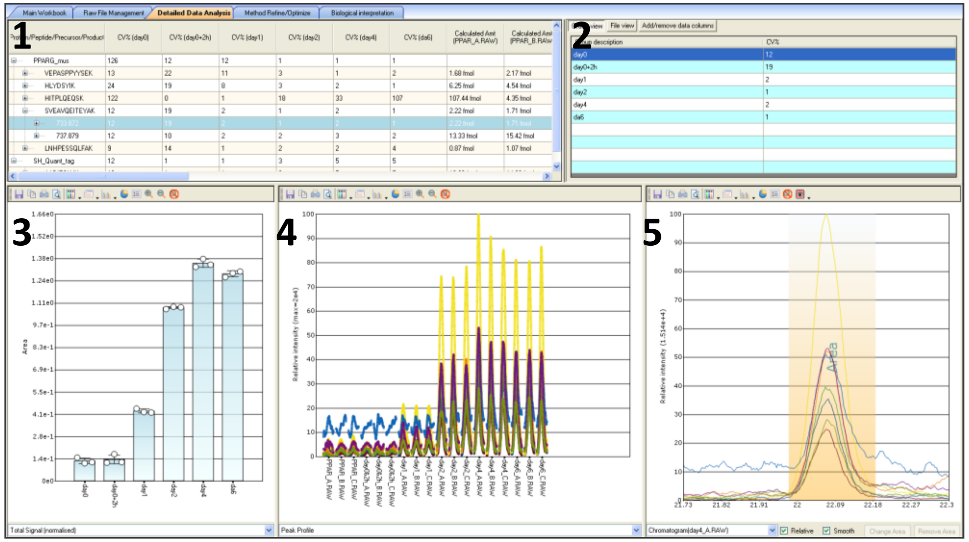


day0_A.RAW	12.5 fmol
day0_B.RAW	12.59 fmol
day0&2h_A.RAW	12.79 fmol
day0&2h_B.RAW	13.44 fmol
day0&2h_C.RAW	12.86 fmol
day1_A.RAW	20.44 fmol
day1_B.RAW	19.49 fmol
day2_A.RAW	53.68 fmol
day2_B.RAW	50.87 fmol
day2_C.RAW	52.48 fmol
day4_A.RAW	28.89 fmol
day4_B.RAW	31.34 fmol
day4_C.RAW	31.18 fmol
day6_A.RAW	19.16 fmol
day6_B.RAW	17.71 fmol
day6_C.RAW	18.31 fmol

Supplementary Figure 4. SRM monitoring of the five RXRα peptides in 3T3-L1 nuclear extract.
(a,c,e,g,i) Screen prints of all five RXRα peptides monitored by SRM. The Pinpoint window provides a summary of the evaluation of the peptide of interest. Please see Supplementary Figure 3 for a description of the subpanels. (b,d,f,h,j) Calculated amounts of endogenous levels found in nuclear extracts.

PPARγ peptide: SVEAQEITEYAK

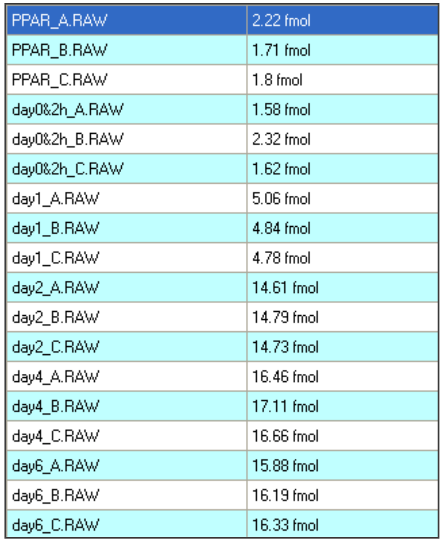
a



Protein/Peptide/Precursor/Product	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)	CV% (da6)	Calculated Amt (PPAR_A.RAW)	Calculated Amt (PPAR_B.RAW)
PPARG_mus	126	12	12	1	1	1		
VEPASPPYYSEK	13	22	11	3	1	2	1.68 fmol	2.17 fmol
HLYDSYIK	24	19	8	3	2	1	6.25 fmol	4.54 fmol
HITPLQEQSK	122	0	1	18	33	107	107.44 fmol	4.35 fmol
SVEAVQEITEYAK	12	19	2	1	2	1	2.22 fmol	1.71 fmol
733.872	12	19	2	1	2	1	2.22 fmol	1.71 fmol
737.879	12	10	2	2	3	2	13.33 fmol	15.42 fmol
LNHPESSQLFAK	9	14	1	2	2	4	0.87 fmol	1.07 fmol
SH_Quant_tag	12	1	1	3	5	5		

Group description	CV%
day0	12
day0+2h	19
day1	2
day2	1
day4	2
da6	1

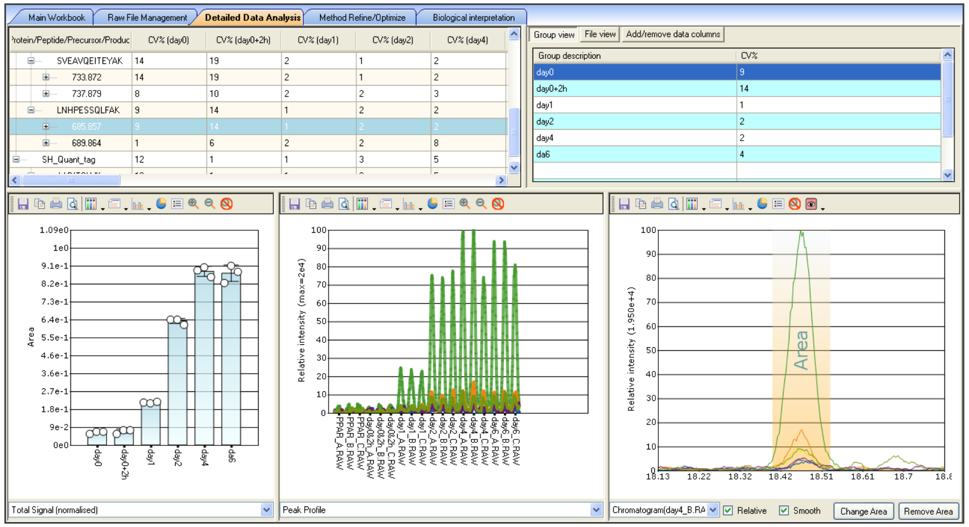
b



PPAR_A.RAW	2.22 fmol
PPAR_B.RAW	1.71 fmol
PPAR_C.RAW	1.8 fmol
day0&2h_A.RAW	1.58 fmol
day0&2h_B.RAW	2.32 fmol
day0&2h_C.RAW	1.62 fmol
day1_A.RAW	5.06 fmol
day1_B.RAW	4.84 fmol
day1_C.RAW	4.78 fmol
day2_A.RAW	14.61 fmol
day2_B.RAW	14.79 fmol
day2_C.RAW	14.73 fmol
day4_A.RAW	16.46 fmol
day4_B.RAW	17.11 fmol
day4_C.RAW	16.66 fmol
day6_A.RAW	15.88 fmol
day6_B.RAW	16.19 fmol
day6_C.RAW	16.33 fmol

PPARγ peptide: LNHPESSQLFAK

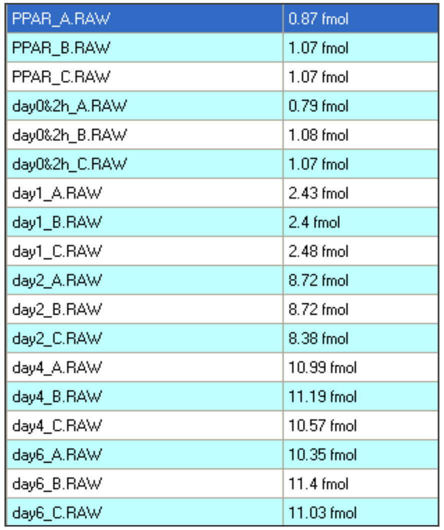
c



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
SVEAVQEITEYAK	14	19	2	1	2
733.872	14	19	2	1	2
737.879	8	10	2	2	3
LNHPESSQLFAK	9	14	1	2	2
685.857	9	14	1	2	2
689.864	1	6	2	2	8
SH_Quant_tag	12	1	1	3	5

Group description	CV%
day0	9
day0+2h	14
day1	1
day2	2
day4	2
da6	4

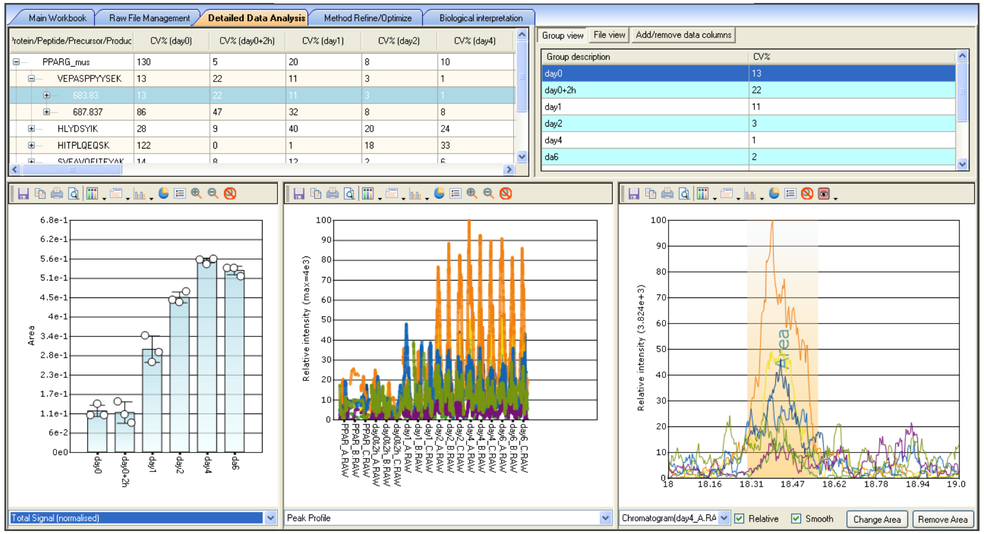
d



PPAR_A.RAW	0.87 fmol
PPAR_B.RAW	1.07 fmol
PPAR_C.RAW	1.07 fmol
day0&2h_A.RAW	0.79 fmol
day0&2h_B.RAW	1.08 fmol
day0&2h_C.RAW	1.07 fmol
day1_A.RAW	2.43 fmol
day1_B.RAW	2.4 fmol
day1_C.RAW	2.48 fmol
day2_A.RAW	8.72 fmol
day2_B.RAW	8.72 fmol
day2_C.RAW	8.38 fmol
day4_A.RAW	10.99 fmol
day4_B.RAW	11.19 fmol
day4_C.RAW	10.57 fmol
day6_A.RAW	10.35 fmol
day6_B.RAW	11.4 fmol
day6_C.RAW	11.03 fmol

PPARγ peptide: VEPASPPYYSEK

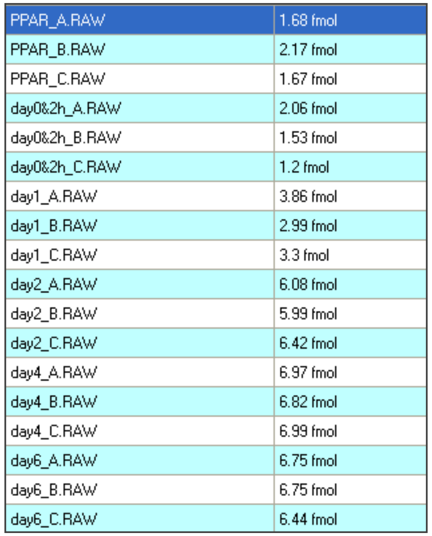
e



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
PPARG_mus	130	5	20	8	10
VEPASPPYYSEK	13	22	11	3	1
683.83	13	22	11	3	1
687.837	86	47	32	8	8
HLYDSYIK	28	9	40	20	24
HITPLQEQSK	122	0	1	18	33
SVEAVQEITEYAK	14	8	12	2	6

Group description	CV%
day0	13
day0+2h	22
day1	11
day2	3
day4	1
da6	2

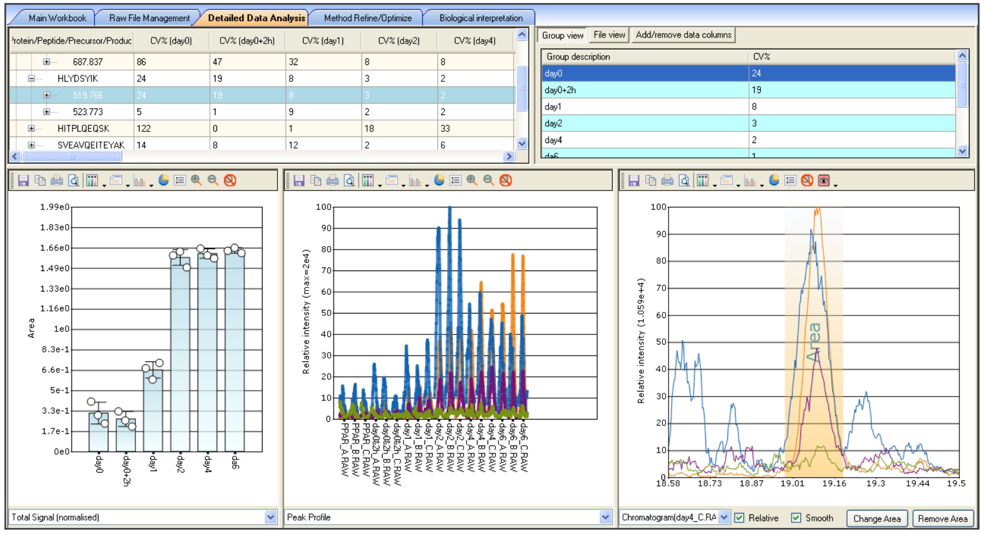
f



PPAR_A.RAW	1.68 fmol
PPAR_B.RAW	2.17 fmol
PPAR_C.RAW	1.67 fmol
day0&2h_A.RAW	2.06 fmol
day0&2h_B.RAW	1.53 fmol
day0&2h_C.RAW	1.2 fmol
day1_A.RAW	3.86 fmol
day1_B.RAW	2.99 fmol
day1_C.RAW	3.3 fmol
day2_A.RAW	6.08 fmol
day2_B.RAW	5.99 fmol
day2_C.RAW	6.42 fmol
day4_A.RAW	6.97 fmol
day4_B.RAW	6.82 fmol
day4_C.RAW	6.99 fmol
day6_A.RAW	6.75 fmol
day6_B.RAW	6.75 fmol
day6_C.RAW	6.44 fmol

PPARγ peptide: HLYDSYIK

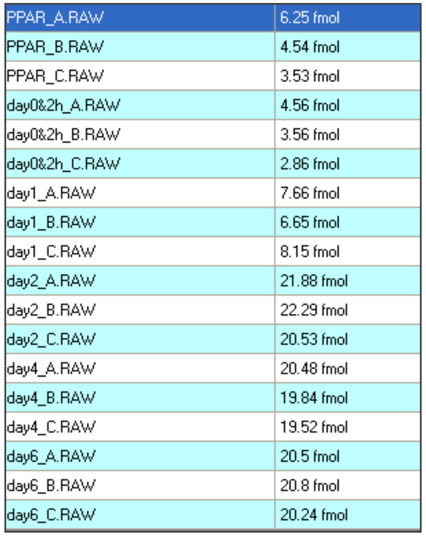
g



Protein/Peptide/Precursor/Produc	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)
687.837	86	47	32	8	8
HLYDSYIK	24	19	8	3	2
519.766	24	19	8	3	2
523.773	5	1	9	2	2
HITPLQEQSK	122	0	1	18	33
SVEAVQEITEYAK	14	8	12	2	6

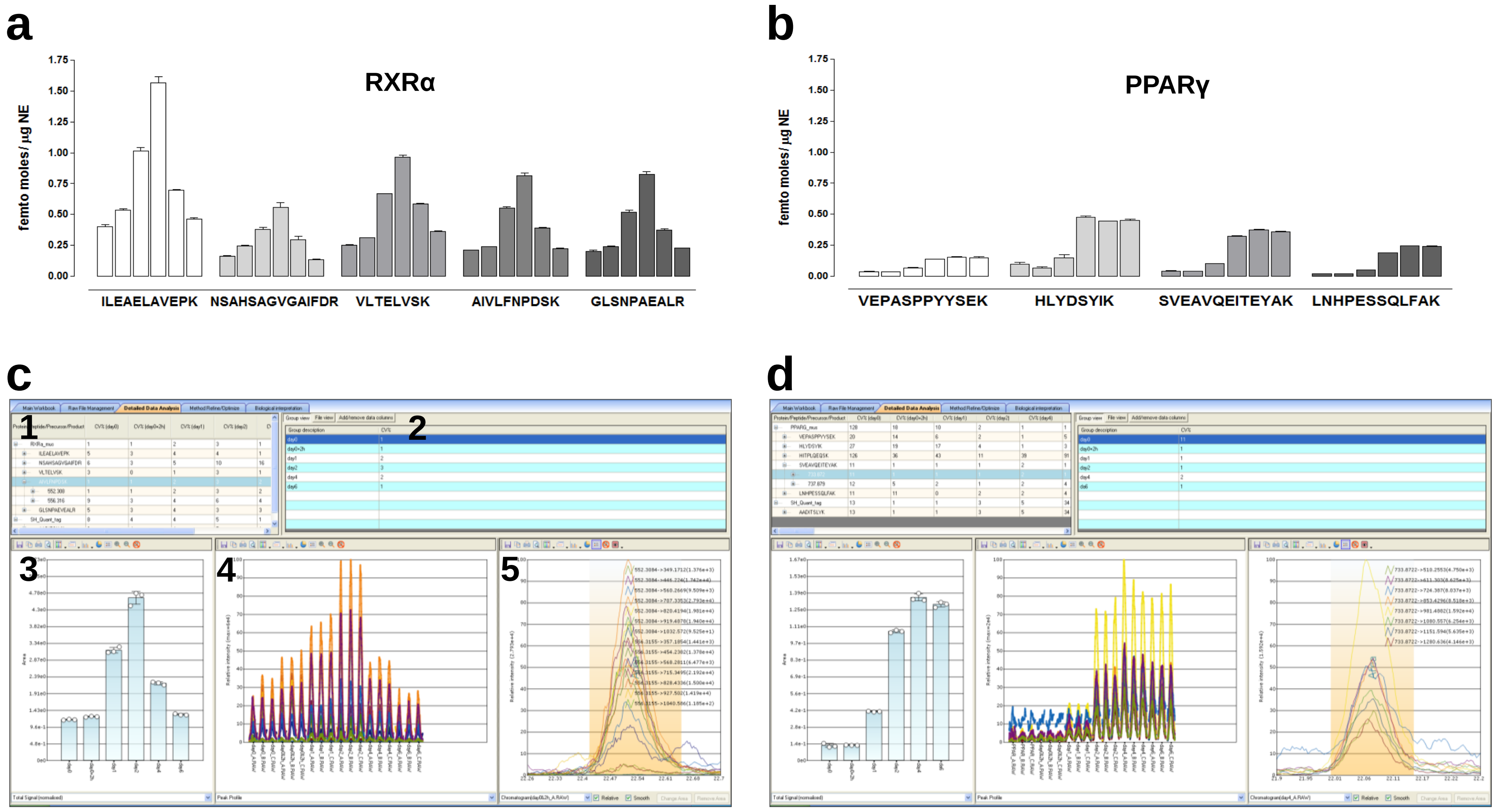
Group description	CV%
day0	24
day0+2h	19
day1	8
day2	3
day4	2
da6	1

h



PPAR_A.RAW	6.25 fmol
PPAR_B.RAW	4.54 fmol
PPAR_C.RAW	3.53 fmol
day0&2h_A.RAW	4.56 fmol
day0&2h_B.RAW	3.56 fmol
day0&2h_C.RAW	2.86 fmol
day1_A.RAW	7.66 fmol
day1_B.RAW	6.65 fmol
day1_C.RAW	8.15 fmol
day2_A.RAW	21.88 fmol
day2_B.RAW	22.29 fmol
day2_C.RAW	20.53 fmol
day4_A.RAW	20.48 fmol
day4_B.RAW	19.84 fmol
day4_C.RAW	19.52 fmol
day6_A.RAW	20.5 fmol
day6_B.RAW	20.8 fmol
day6_C.RAW	20.24 fmol

Supplementary Figure 5. SRM monitoring of the four PPARγ peptides in 3T3-L1 nuclear extract.
(a,c,e,g) Screen prints of all four PPARγ peptides monitored by SRM. The Pinpoint window provides a summary of the evaluation of the peptide of interest. Please see Supplementary Figure 3 for a description of the subpanels. (b,d,f,h) Calculated amounts of endogenous levels found in nuclear extracts.



Protein/Peptide/Precursor/Product	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	C
RXRa_mus	1	1	2	3	1
ILEAELAVEPK	5	3	4	4	1
NSAHSAGVGAIFDR	6	3	5	10	16
VLTELVSK	3	0	1	3	1
AIVLFNPDSK	1	1	2	3	2
552.308	1	1	2	3	2
556.316	9	3	4	6	4
GLSNPAEVEALR	5	3	4	3	3
SH_Quant_tag	8	4	4	5	1

Group description	CV%
day0	1
day0+2h	1
day1	2
day2	3
day4	2
day6	1

Protein/Peptide/Precursor/Product	CV% (day0)	CV% (day0+2h)	CV% (day1)	CV% (day2)	CV% (day4)	
PPARG_mus	128	18	10	2	1	1
VEPASPPYYSEK	20	14	6	2	1	5
HLYDSYIK	27	19	17	4	1	3
HITPLQEQSK	126	36	43	11	39	91
SVEAVQEITEYAK	11	1	1	1	2	1
733.872	11	1	1	1	2	1
737.879	12	5	2	1	2	4
LNHPESSQLFAK	11	11	0	2	2	4
SH_Quant_tag	13	1	1	3	5	34
AADITSLYK	13	1	1	3	5	34

Group description	CV%
day0	11
day0+2h	1
day1	1
day2	1
day4	2
da6	1

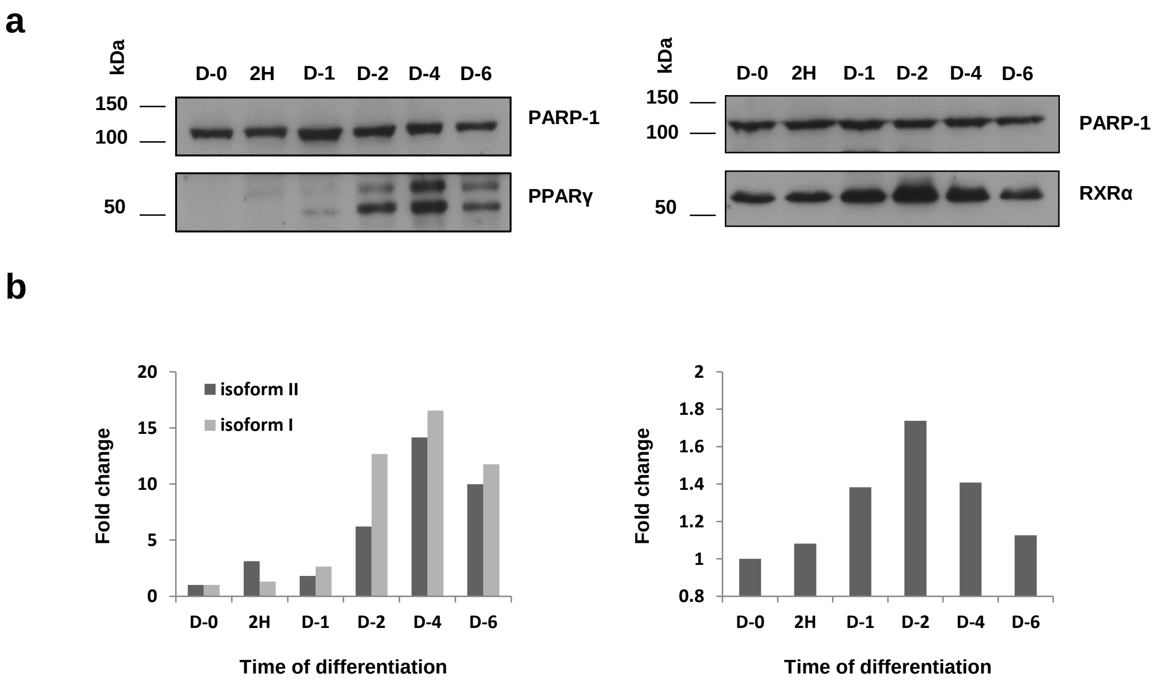
e

RXRα sequence	Parent Mass m/z $[M+2H]^{2+}$	CV of peptides analyzed at day 0 to day 6 day 0	 day 0 + 2h	 day 1	 day 2	 day 4	 day 6
ILEAELAVEPK	606.34	5%	3%	4%	4%	1%	3%
NSAHSAGVGAIFDR	701.34	6%	3%	5%	10%	16%	5%
VLTELVSK	444.77	3%	<1%	1%	3%	1%	2%
AIVLFNPDSK	552.3	1%	1%	2%	3%	2%	1%
GLSNPAEALR	628.33	5%	3%	4%	3%	3%	2%

f

PPARγ sequence	Parent Mass m/z $[M+2H]^{2+}$	CV of peptides analyzed at day 0 to day 6 day 0	 day 0 + 2h	 day 1	 day 2	 day 4	 day 6
VEPASPPYYSEK	683.83	20%	14%	6%	2%	1%	5%
HLYDSYIK	519.76	27%	19%	17%	4%	1%	3%
SVEAVQEITEYAK	733.87	11%	1%	1%	1%	2%	1%
LNHPESSQLAK	685.85	11%	11%	<1%	2%	2%	4%

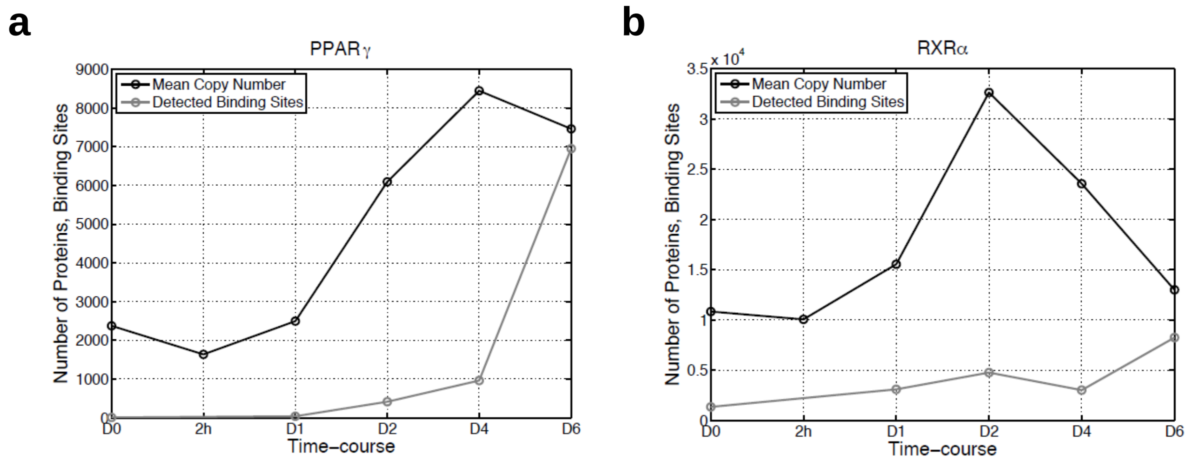
Supplementary Figure 6. Calculated coefficient of variation of all RXRα and PPARγ peptides monitored by SRM.
(a-b) Summary of all RXRα and PPARγ peptides monitored by SRM. Values are the mean ± SD of one biological sample analyzed in three technical replicates. The Pinpoint window provides a summary of the results as well as the calculated CV of RXRα (c) and PPARγ (d) peptides respectively. Please see Supplementary Figure 3 for a description of the subpanels. Tables (e-f) summarize the calculated CV values for all RXRα (n=5) and PPARγ (n=4) peptides monitored by SRM.



Supplementary Figure 7. Immunoblotting analysis of PPARγ and RXRα expression levels found in 3T3-L1 nuclear extracts.

(a) Validation of nuclear levels of PPARγ (left panel: isoform II on top at 57.6 kDa, isoform I on the bottom at 54.5 kDa) and RXRα (right panel: 51,2 kDa) expression by Western blot using TF-specific primary antibodies in 3T3-L1 cells during 6 time points of terminal differentiation. PARP-1 was used as a nuclear loading control.

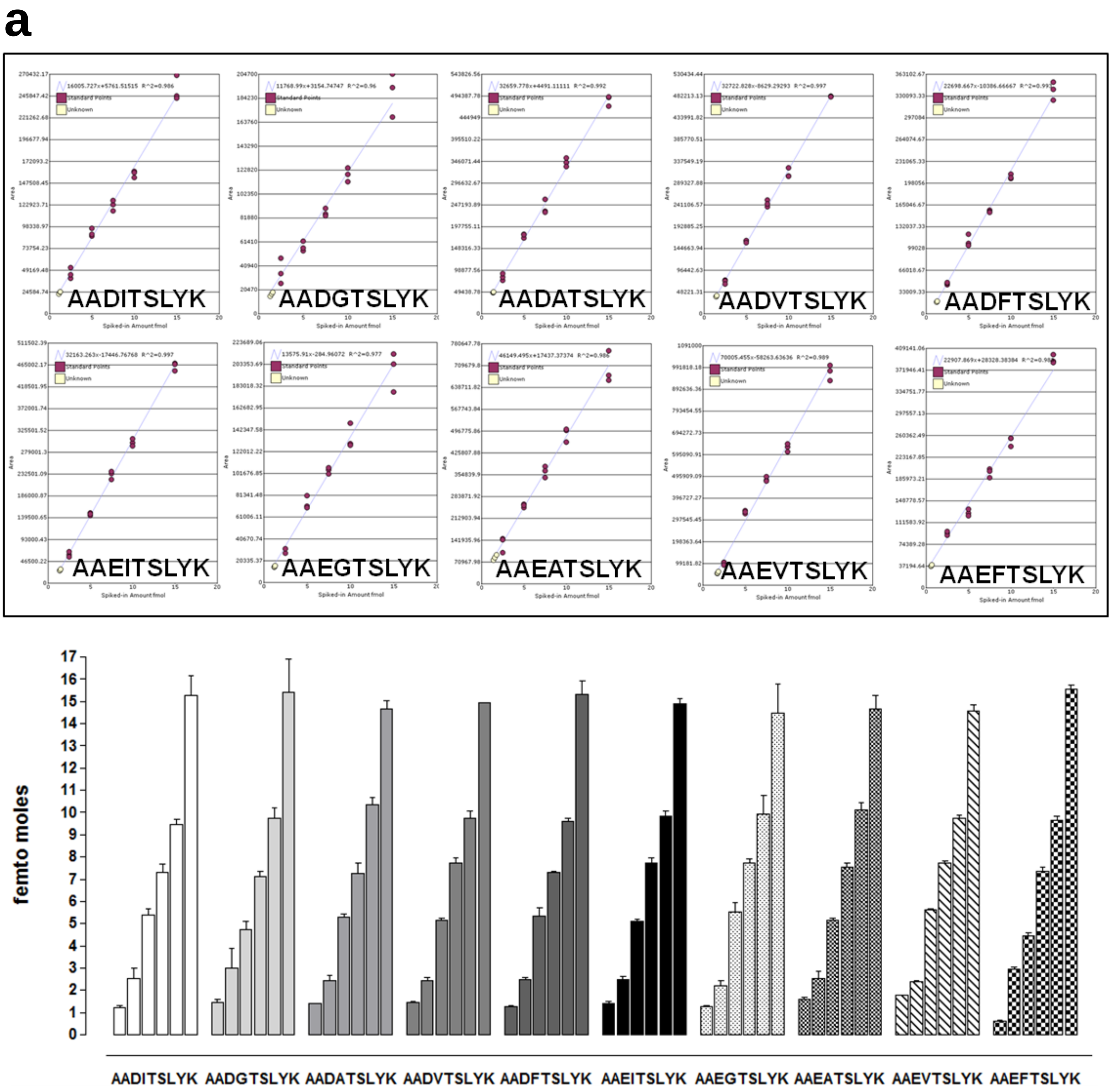
(b) Densitometric analysis was performed for both PPARγ (left) and RXRα (right). Values were normalized against PARP-1 and Day 0 was taken as reference. Values of the subsequent days are represented in terms of fold changes.



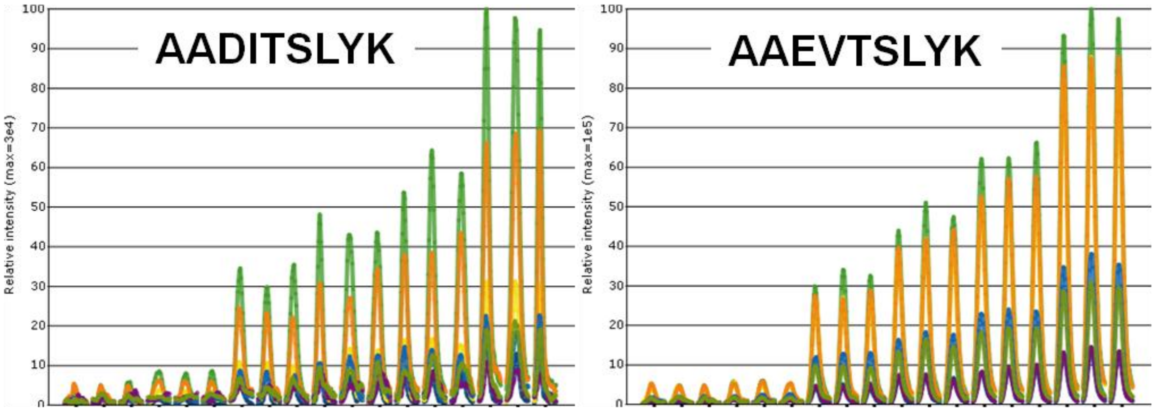
Supplementary Figure 8. PPARγ and RXRα protein copy number versus the number of detected binding sites.

(a) Graph showing the PPARγ protein copy number profile during terminal adipogenesis as detected by SRMs in relation to its binding site profile. A sharp increase in detected binding sites can be observed after Day 4, despite a saturation in PPARγ protein copy numbers.

(b) Graph showing the RXRα protein copy number profile during terminal adipogenesis as detected by SRMs in relation to its binding site profile. A sharp increase in detected binding sites can be observed after Day 4, despite a decrease in RXRα protein copy numbers.

a
16005.727x+5761.51515 R^2=0.986
11768.99x+3154.74747 R^2=0.96
32659.778x+4491.11111 R^2=0.992
32722.828x-8629.29293 R^2=0.997
22698.667x-10386.66667 R^2=0.99
32163.263x-17446.76768 R^2=0.997
13575.91x-284.96072 R^2=0.977
46149.495x+17437.37374 R^2=0.986
70005.455x-58263.63636 R^2=0.989
22907.869x+28328.38384 R^2=0.98
Standard Points
Unknown
Area
Spiked-in Amount fmol
AADITSLYK
AADGTSLYK
AADATSLYK
AADVTSLYK
AADFTSLYK
AAEITSLYK
AAEGTSLYK
AAEATSLYK
AAEVTSLYK
AAEFTSLYK
femto moles
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

b



no.	Tag	Parent Mass	CV of tags (1.25 - 15 fmole)					
	sequence	m/z $[M+2H]^{2+}$	1.25	2.5	5.0	7.5	10	15
1	AADITSLYK	491.26	5%	11%	4%	4%	2%	5%
2	AADGTSLYK	463.24	8%	29%	6%	3%	4%	8%
3	AADATSLYK	470.25	2%	9%	2%	6%	2%	2%
4	AADVTSLYK	484.25	3%	6%	1%	3%	3%	1%
5	AADFTSLYK	508.25	4%	3%	7%	1%	1%	3%
6	AAEITSLYK	498.27	5%	5%	2%	3%	2%	2%
7	AAEGTSLYK	470.24	4%	7%	7%	2%	7%	8%
8	AAEATSLYK	477.25	8%	17%	2%	4%	4%	6%
9	AAEVTSLYK	491.26	7%	7%	2%	2%	2%	3%
10	AAEFTSLYK	515.26	3%	2%	4%	3%	3%	2%

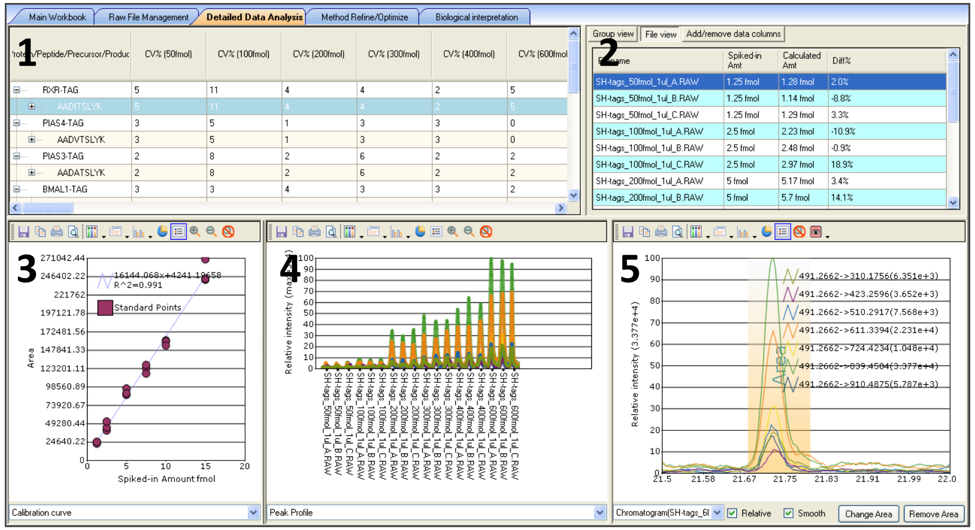
Supplementary Figure 9. Establishment of standard curves for all ten SH-Quant tags used for multiplex SRM analysis.

(a) All ten SH-Quant tags were spiked into nuclear extracts and analyzed in triplicates. Standard curves were established covering an on-column tag concentration range from 2.5 fmol to 15.0 fmol in a complex matrix. Top panels: linear regression analysis (R^2 = 0.960 – 0.997) for all ten SH-Quant tags. Bottom panel: graphical representation of the variation found within all technical replicates (2.5 - 15 fmol) as well as the calculated value of an unknown treated sample (positive control) spiked into nuclear extract at 1.25 fmol.

(b) Top panel: example of the calculated peak areas of two isobaric tags (tags 1 and 9), with slightly modified sequences, analyzed in three technical replicates at a concentration range of 1.25 – 15 fmol. Bottom panel: the table provides a summary of the calculated CV for all ten SH-Quant tags spiked at different concentrations (1.25 – 15 fmoles) into nuclear extracts. All samples were analyzed in triplicates.

AADITSLYK

a

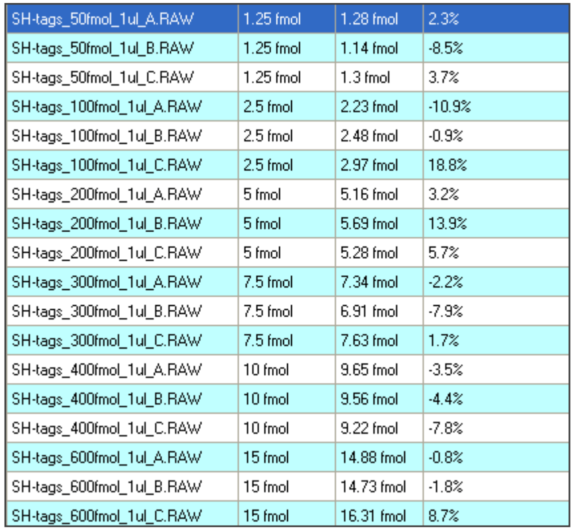


'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
⊟ RXR-TAG	5	11	4	4	2	5
⊞ AADITSLYK	5	11	4	4	2	5
⊟ PIAS4-TAG	3	5	1	3	3	0
⊞ AADVTSLYK	3	5	1	3	3	0
⊟ PIAS3-TAG	2	8	2	6	2	2
⊞ AADATSLYK	2	8	2	6	2	2
⊟ BMAL1-TAG	3	3	4	3	3	2

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.28 fmol	2.0%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.14 fmol	-8.8%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.29 fmol	3.3%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.23 fmol	-10.9%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.48 fmol	-0.9%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.97 fmol	18.9%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.17 fmol	3.4%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.7 fmol	14.1%

271042.44
246402.22
221762
197121.78
172481.56
147841.33
123201.11
98560.89
73920.67
49280.44
24640.22
0
Area
16144.068x+4241.19658
R^2=0.991
Standard Points
0
5
10
15
20
Spiked-in Amount fmol
Relative intensity (max
100
90
80
70
60
50
40
30
20
10
0
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
Relative intensity (3.377e+4)
491.2662->310.1756(6.351e+3)
491.2662->423.2596(3.652e+3)
491.2662->510.2917(7.568e+3)
491.2662->611.3394(2.231e+4)
491.2662->724.4234(1.048e+4)
491.2662->839.4504(3.377e+4)
491.2662->910.4875(5.787e+3)
Area
21.5
21.58
21.67
21.75
21.83
21.91
21.99
22.0

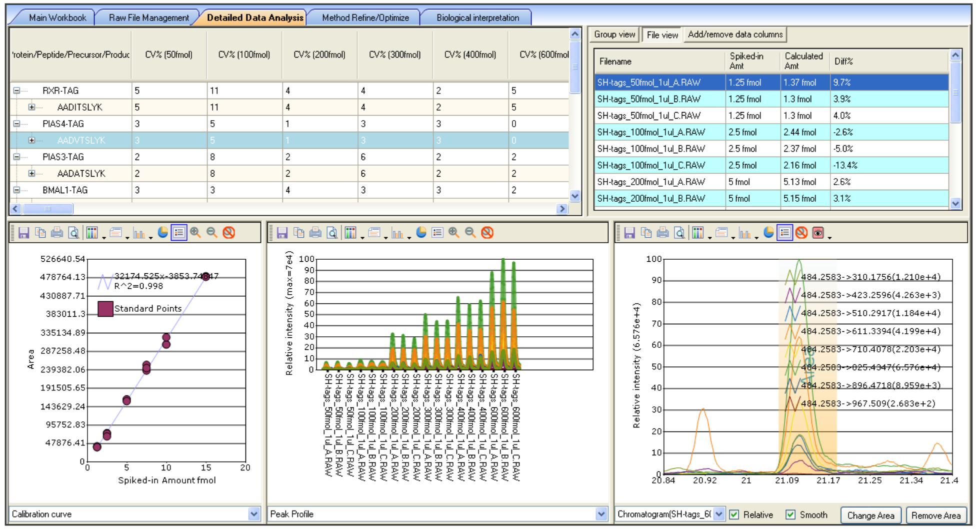
b



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.28 fmol	2.3%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.14 fmol	-8.5%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.3 fmol	3.7%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.23 fmol	-10.9%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.48 fmol	-0.9%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.97 fmol	18.8%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.16 fmol	3.2%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.69 fmol	13.9%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.28 fmol	5.7%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.34 fmol	-2.2%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	6.91 fmol	-7.9%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.63 fmol	1.7%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.65 fmol	-3.5%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.56 fmol	-4.4%
SH-tags_400fmol_1ul_C.RAW	10 fmol	9.22 fmol	-7.8%
SH-tags_600fmol_1ul_A.RAW	15 fmol	14.88 fmol	-0.8%
SH-tags_600fmol_1ul_B.RAW	15 fmol	14.73 fmol	-1.8%
SH-tags_600fmol_1ul_C.RAW	15 fmol	16.31 fmol	8.7%

AADVTSLYK

c

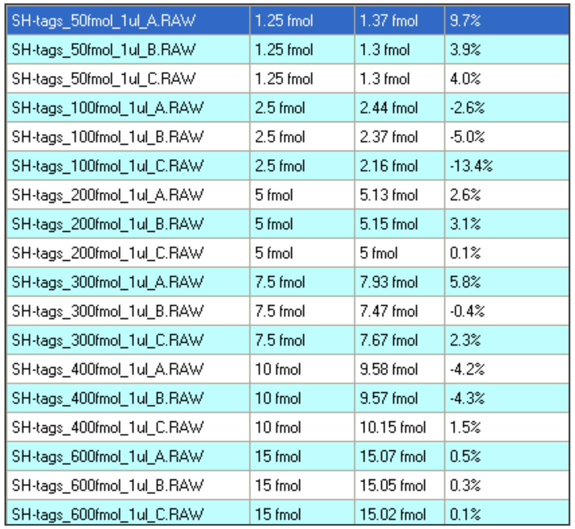


'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
⊟ RXR-TAG	5	11	4	4	2	5
⊞ AADITSLYK	5	11	4	4	2	5
⊟ PIAS4-TAG	3	5	1	3	3	0
⊞ AADVTSLYK	3	5	1	3	3	0
⊟ PIAS3-TAG	2	8	2	6	2	2
⊞ AADATSLYK	2	8	2	6	2	2
⊟ BMAL1-TAG	3	3	4	3	3	2

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.37 fmol	9.7%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.3 fmol	3.9%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.3 fmol	4.0%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.44 fmol	-2.6%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.37 fmol	-5.0%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.16 fmol	-13.4%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.13 fmol	2.6%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.15 fmol	3.1%

526640.54
478764.13
430887.71
383011.3
335134.89
287258.48
239382.06
191505.65
143629.24
95752.83
47876.41
0
Area
32174.525x-3853.74847
R^2=0.998
Standard Points
0
5
10
15
20
Spiked-in Amount fmol
Relative intensity (max=7e4)
100
90
80
70
60
50
40
30
20
10
0
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
Relative intensity (6.576e+4)
484.2583->310.1756(1.210e+4)
484.2583->423.2596(4.263e+3)
484.2583->510.2917(1.184e+4)
484.2583->611.3394(4.199e+4)
484.2583->710.4078(2.203e+4)
484.2583->825.4347(6.576e+4)
484.2583->896.4718(8.959e+3)
484.2583->967.509(2.683e+2)
Area
20.84
20.92
21
21.09
21.17
21.25
21.34
21.4

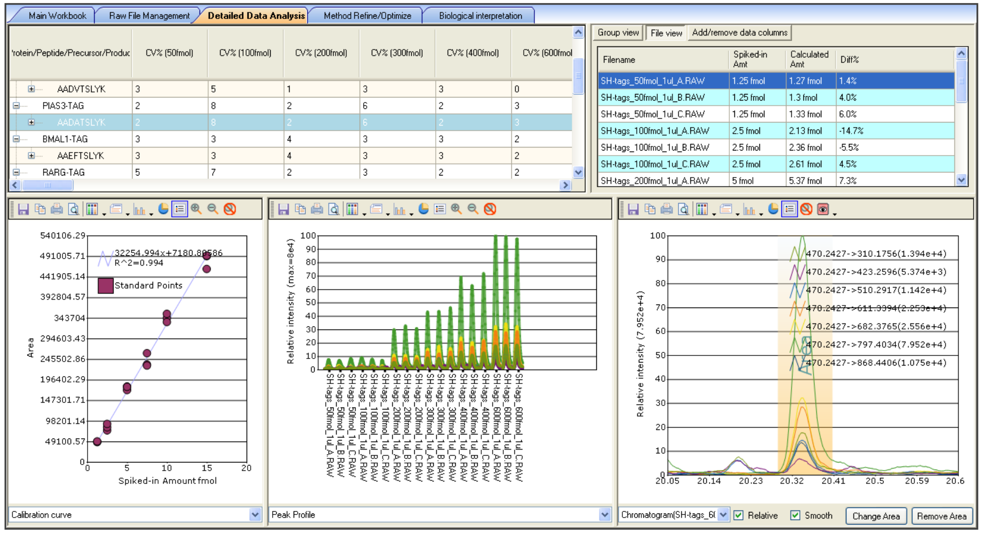
d



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.37 fmol	9.7%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.3 fmol	3.9%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.3 fmol	4.0%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.44 fmol	-2.6%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.37 fmol	-5.0%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.16 fmol	-13.4%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.13 fmol	2.6%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.15 fmol	3.1%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5 fmol	0.1%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.93 fmol	5.8%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.47 fmol	-0.4%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.67 fmol	2.3%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.58 fmol	-4.2%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.57 fmol	-4.3%
SH-tags_400fmol_1ul_C.RAW	10 fmol	10.15 fmol	1.5%
SH-tags_600fmol_1ul_A.RAW	15 fmol	15.07 fmol	0.5%
SH-tags_600fmol_1ul_B.RAW	15 fmol	15.05 fmol	0.3%
SH-tags_600fmol_1ul_C.RAW	15 fmol	15.02 fmol	0.1%

e

AADATSLYK

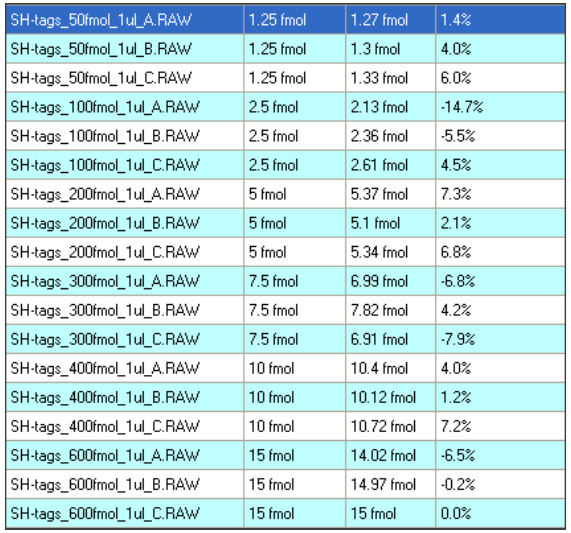


'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
AADVTSLYK	3	5	1	3	3	0
PIAS3-TAG	2	8	2	6	2	3
AADATSLYK	2	8	2	6	2	3
BMAL1-TAG	3	3	4	3	3	2
AAEFTSLYK	3	3	4	3	3	2
RARG-TAG	5	7	2	3	2	2

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.27 fmol	1.4%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.3 fmol	4.0%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.33 fmol	6.0%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.13 fmol	-14.7%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.36 fmol	-5.5%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.61 fmol	4.5%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.37 fmol	7.3%

540106.29
491005.71
441905.14
392804.57
343704
294603.43
245502.86
196402.29
147301.71
98201.14
49100.57
0
Area
32254.994x+7180.80586
R^2=0.994
Standard Points
0
5
10
15
20
Spiked-in Amount fmol
Calibration curve
Relative intensity (max=8e4)
100
90
80
70
60
50
40
30
20
10
0
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
Peak Profile
Relative intensity (7.952e+4)
470.2427->310.1756(1.394e+4)
470.2427->423.2596(5.374e+3)
470.2427->510.2917(1.142e+4)
470.2427->611.3394(2.253e+4)
470.2427->682.3765(2.556e+4)
470.2427->797.4034(7.952e+4)
470.2427->868.4406(1.075e+4)
20.05
20.14
20.23
20.32
20.41
20.5
20.59
20.6
Chromatogram(SH-tags_6I
☑ Relative
☑ Smooth
Change Area
Remove Area

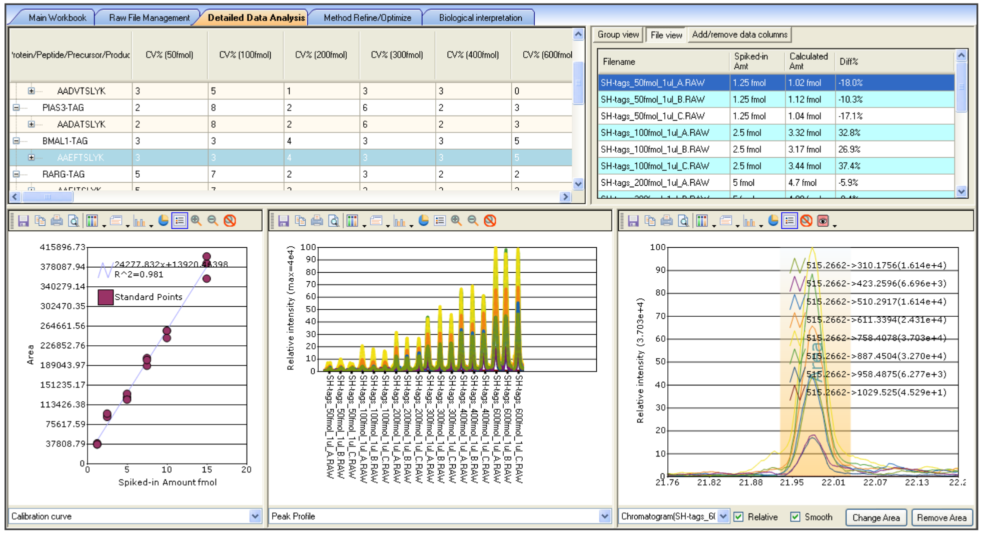
f



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.27 fmol	1.4%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.3 fmol	4.0%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.33 fmol	6.0%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.13 fmol	-14.7%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.36 fmol	-5.5%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.61 fmol	4.5%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.37 fmol	7.3%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.1 fmol	2.1%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.34 fmol	6.8%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	6.99 fmol	-6.8%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.82 fmol	4.2%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	6.91 fmol	-7.9%
SH-tags_400fmol_1ul_A.RAW	10 fmol	10.4 fmol	4.0%
SH-tags_400fmol_1ul_B.RAW	10 fmol	10.12 fmol	1.2%
SH-tags_400fmol_1ul_C.RAW	10 fmol	10.72 fmol	7.2%
SH-tags_600fmol_1ul_A.RAW	15 fmol	14.02 fmol	-6.5%
SH-tags_600fmol_1ul_B.RAW	15 fmol	14.97 fmol	-0.2%
SH-tags_600fmol_1ul_C.RAW	15 fmol	15 fmol	0.0%

g

AAEFTSLYK

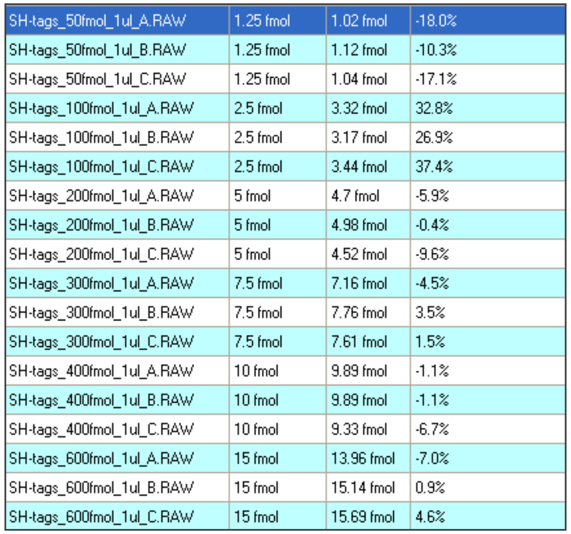


'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
AADVTSLYK	3	5	1	3	3	0
PIAS3-TAG	2	8	2	6	2	3
AADATSLYK	2	8	2	6	2	3
BMAL1-TAG	3	3	4	3	3	5
AAEFTSLYK	3	3	4	3	3	5
RARG-TAG	5	7	2	3	2	2

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.02 fmol	-18.0%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.12 fmol	-10.3%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.04 fmol	-17.1%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	3.32 fmol	32.8%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	3.17 fmol	26.9%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	3.44 fmol	37.4%
SH-tags_200fmol_1ul_A.RAW	5 fmol	4.7 fmol	-5.9%

415896.73
378087.94
340279.14
302470.35
264661.56
226852.76
189043.97
151235.17
113426.38
75617.59
37808.79
0
Area
24277.832x+13920.46398
R^2=0.981
Standard Points
0
5
10
15
20
Spiked-in Amount fmol
Calibration curve
Relative intensity (max=4e4)
100
90
80
70
60
50
40
30
20
10
0
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
Peak Profile
Relative intensity (3.703e+4)
515.2662->310.1756(1.614e+4)
515.2662->423.2596(6.696e+3)
515.2662->510.2917(1.614e+4)
515.2662->611.3394(2.431e+4)
515.2662->758.4078(3.703e+4)
515.2662->887.4504(3.270e+4)
515.2662->958.4875(6.277e+3)
515.2662->1029.525(4.529e+1)
21.76
21.82
21.88
21.95
22.01
22.07
22.13
22.2
Chromatogram(SH-tags_6I
☑ Relative
☑ Smooth
Change Area
Remove Area

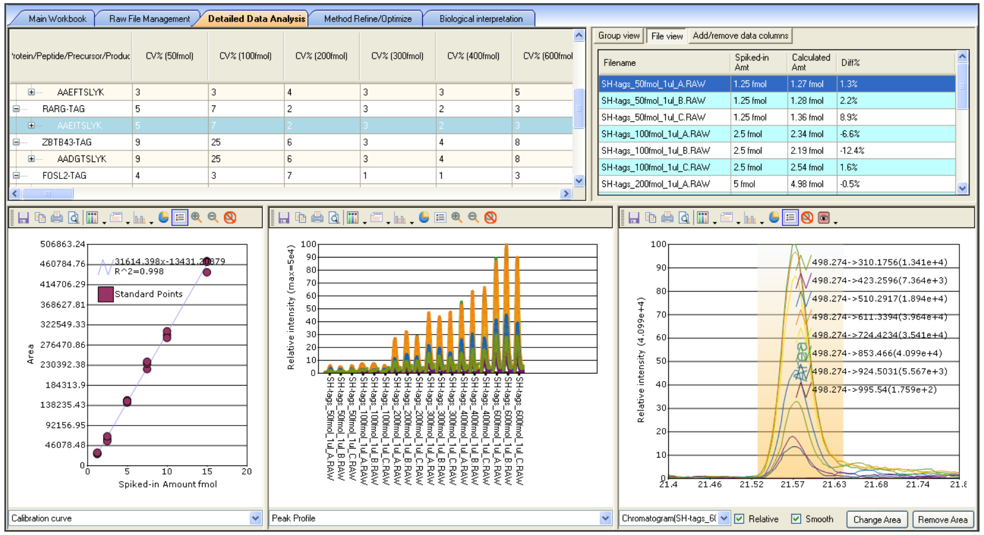
h



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.02 fmol	-18.0%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.12 fmol	-10.3%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.04 fmol	-17.1%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	3.32 fmol	32.8%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	3.17 fmol	26.9%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	3.44 fmol	37.4%
SH-tags_200fmol_1ul_A.RAW	5 fmol	4.7 fmol	-5.9%
SH-tags_200fmol_1ul_B.RAW	5 fmol	4.98 fmol	-0.4%
SH-tags_200fmol_1ul_C.RAW	5 fmol	4.52 fmol	-9.6%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.16 fmol	-4.5%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.76 fmol	3.5%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.61 fmol	1.5%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.89 fmol	-1.1%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.89 fmol	-1.1%
SH-tags_400fmol_1ul_C.RAW	10 fmol	9.33 fmol	-6.7%
SH-tags_600fmol_1ul_A.RAW	15 fmol	13.96 fmol	-7.0%
SH-tags_600fmol_1ul_B.RAW	15 fmol	15.14 fmol	0.9%
SH-tags_600fmol_1ul_C.RAW	15 fmol	15.69 fmol	4.6%

i

AAEITSLYK

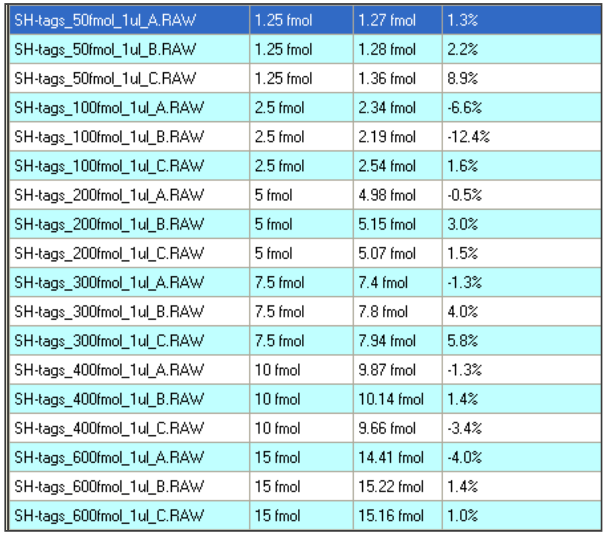


Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
Group view
File view
Add/remove data columns
Calibration curve
Peak Profile
Chromatogram(SH-tags_5l
☑ Relative
☑ Smooth
Change Area
Remove Area
31614.398x-13431.20879
R^2=0.998
Standard Points
Area
506863.24
460784.76
414706.29
368627.81
322549.33
276470.86
230392.38
184313.9
138235.43
92156.95
46078.48
0
0
5
10
15
20
Spiked-in Amount fmol
Relative intensity (max=5e4)
100
90
80
70
60
50
40
30
20
10
0
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
Relative intensity (4.099e+4)
100
90
80
70
60
50
40
30
20
10
0
498.274->310.1756(1.341e+4)
498.274->423.2596(7.364e+3)
498.274->510.2917(1.894e+4)
498.274->611.3394(3.964e+4)
498.274->724.4234(3.541e+4)
498.274->853.466(4.099e+4)
498.274->924.5031(5.567e+3)
498.274->995.54(1.759e+2)
21.4
21.46
21.52
21.57
21.63
21.68
21.74
21.8

'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
AAEFTSLYK	3	3	4	3	3	5
RARG-TAG	5	7	2	3	2	3
AAEITSLYK	5	7	2	3	2	3
ZBTB43-TAG	9	25	6	3	4	8
AADGTSLYK	9	25	6	3	4	8
FOSL2-TAG	4	3	7	1	1	3

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.27 fmol	1.3%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.28 fmol	2.2%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.36 fmol	8.9%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.34 fmol	-6.6%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.19 fmol	-12.4%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.54 fmol	1.6%
SH-tags_200fmol_1ul_A.RAW	5 fmol	4.98 fmol	-0.5%

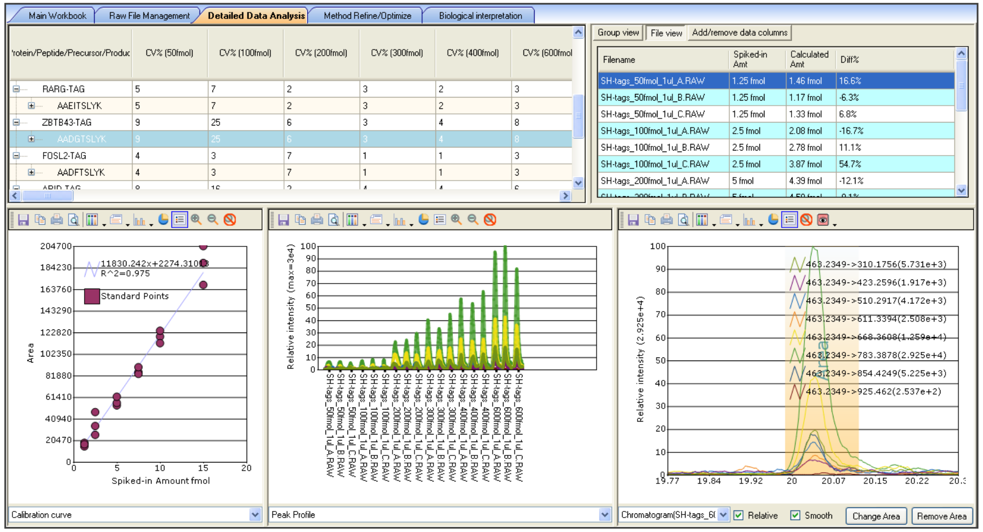
j



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.27 fmol	1.3%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.28 fmol	2.2%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.36 fmol	8.9%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.34 fmol	-6.6%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.19 fmol	-12.4%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.54 fmol	1.6%
SH-tags_200fmol_1ul_A.RAW	5 fmol	4.98 fmol	-0.5%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.15 fmol	3.0%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.07 fmol	1.5%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.4 fmol	-1.3%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.8 fmol	4.0%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.94 fmol	5.8%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.87 fmol	-1.3%
SH-tags_400fmol_1ul_B.RAW	10 fmol	10.14 fmol	1.4%
SH-tags_400fmol_1ul_C.RAW	10 fmol	9.66 fmol	-3.4%
SH-tags_600fmol_1ul_A.RAW	15 fmol	14.41 fmol	-4.0%
SH-tags_600fmol_1ul_B.RAW	15 fmol	15.22 fmol	1.4%
SH-tags_600fmol_1ul_C.RAW	15 fmol	15.16 fmol	1.0%

k

AADGTSLYK

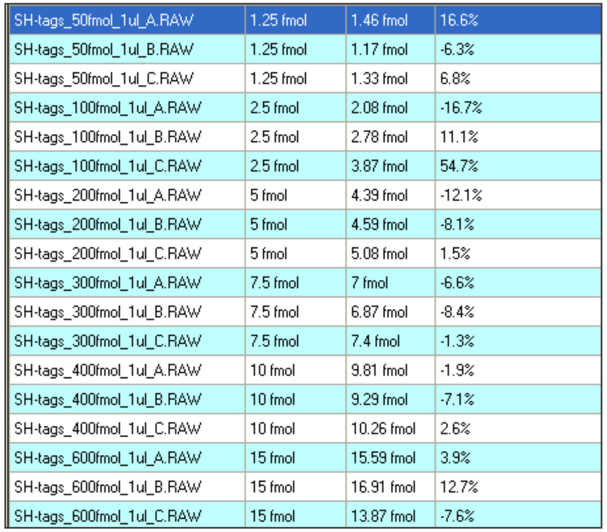


Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
Group view
File view
Add/remove data columns
Calibration curve
Peak Profile
Chromatogram(SH-tags_5l
☑ Relative
☑ Smooth
Change Area
Remove Area
11830.242x+2274.31003
R^2=0.975
Standard Points
Area
204700
184230
163760
143290
122820
102350
81880
61410
40940
20470
0
0
5
10
15
20
Spiked-in Amount fmol
Relative intensity (max=3e4)
100
90
80
70
60
50
40
30
20
10
0
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
Relative intensity (2.925e+4)
100
90
80
70
60
50
40
30
20
10
0
463.2349->310.1756(5.731e+3)
463.2349->423.2596(1.917e+3)
463.2349->510.2917(4.172e+3)
463.2349->611.3394(2.508e+3)
463.2349->668.3608(1.259e+4)
463.2349->783.3878(2.925e+4)
463.2349->854.4249(5.225e+3)
463.2349->925.462(2.537e+2)
19.77
19.84
19.92
20
20.07
20.15
20.22
20.3

'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
RARG-TAG	5	7	2	3	2	3
AAEITSLYK	5	7	2	3	2	3
ZBTB43-TAG	9	25	6	3	4	8
AADGTSLYK	9	25	6	3	4	8
FOSL2-TAG	4	3	7	1	1	3
AADFTSLYK	4	3	7	1	1	3

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.46 fmol	16.6%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.17 fmol	-6.3%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.33 fmol	6.8%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.08 fmol	-16.7%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.78 fmol	11.1%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	3.87 fmol	54.7%
SH-tags_200fmol_1ul_A.RAW	5 fmol	4.39 fmol	-12.1%

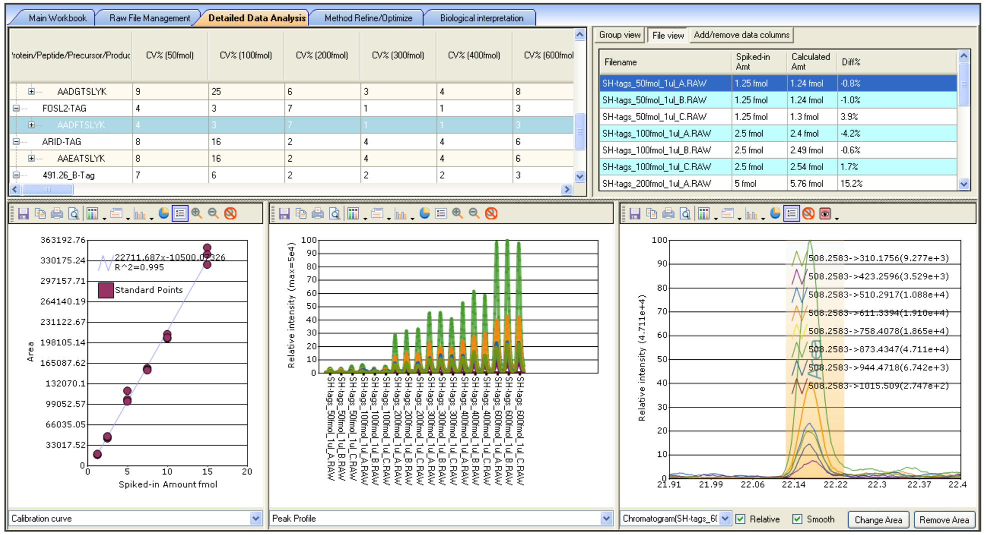
l



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.46 fmol	16.6%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.17 fmol	-6.3%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.33 fmol	6.8%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.08 fmol	-16.7%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.78 fmol	11.1%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	3.87 fmol	54.7%
SH-tags_200fmol_1ul_A.RAW	5 fmol	4.39 fmol	-12.1%
SH-tags_200fmol_1ul_B.RAW	5 fmol	4.59 fmol	-8.1%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.08 fmol	1.5%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7 fmol	-6.6%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	6.87 fmol	-8.4%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.4 fmol	-1.3%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.81 fmol	-1.9%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.29 fmol	-7.1%
SH-tags_400fmol_1ul_C.RAW	10 fmol	10.26 fmol	2.6%
SH-tags_600fmol_1ul_A.RAW	15 fmol	15.59 fmol	3.9%
SH-tags_600fmol_1ul_B.RAW	15 fmol	16.91 fmol	12.7%
SH-tags_600fmol_1ul_C.RAW	15 fmol	13.87 fmol	-7.6%

AADFTSLYK

m

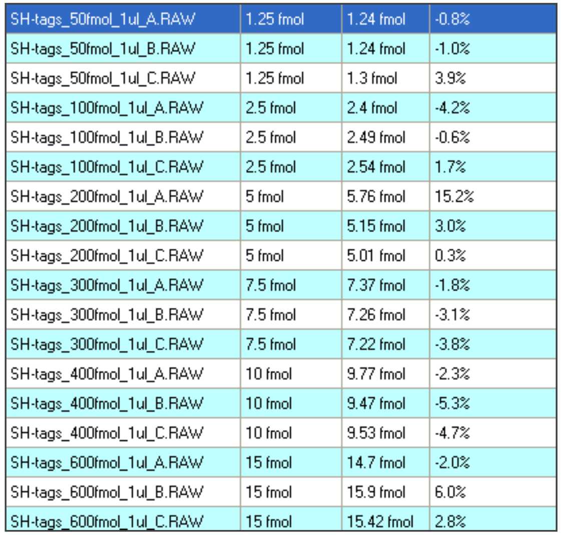


Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
Group view
File view
Add/remove data columns
Calibration curve
Peak Profile
Chromatogram(SH-tags_6(
Relative
Smooth
Change Area
Remove Area
22711.687x-10500.07326
R^2=0.995
Standard Points
Area
Spiked-in Amount fmol
Relative intensity (max=5e4)
Relative intensity (4.711e+4)
508.2583->310.1756(9.277e+3)
508.2583->423.2596(3.529e+3)
508.2583->510.2917(1.088e+4)
508.2583->611.3394(1.910e+4)
508.2583->758.4078(1.865e+4)
508.2583->873.4347(4.711e+4)
508.2583->944.4718(6.742e+3)
508.2583->1015.509(2.747e+2)

'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
⊞ AADGTSLYK	9	25	6	3	4	8
⊟ FOSL2-TAG	4	3	7	1	1	3
⊞ AADFTSLYK	4	3	7	1	1	3
⊟ ARID-TAG	8	16	2	4	4	6
⊞ AAEATSLYK	8	16	2	4	4	6
⊟ 491.26_B-Tag	7	6	2	2	2	3

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.24 fmol	-0.8%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.24 fmol	-1.0%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.3 fmol	3.9%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.4 fmol	-4.2%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.49 fmol	-0.6%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.54 fmol	1.7%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.76 fmol	15.2%

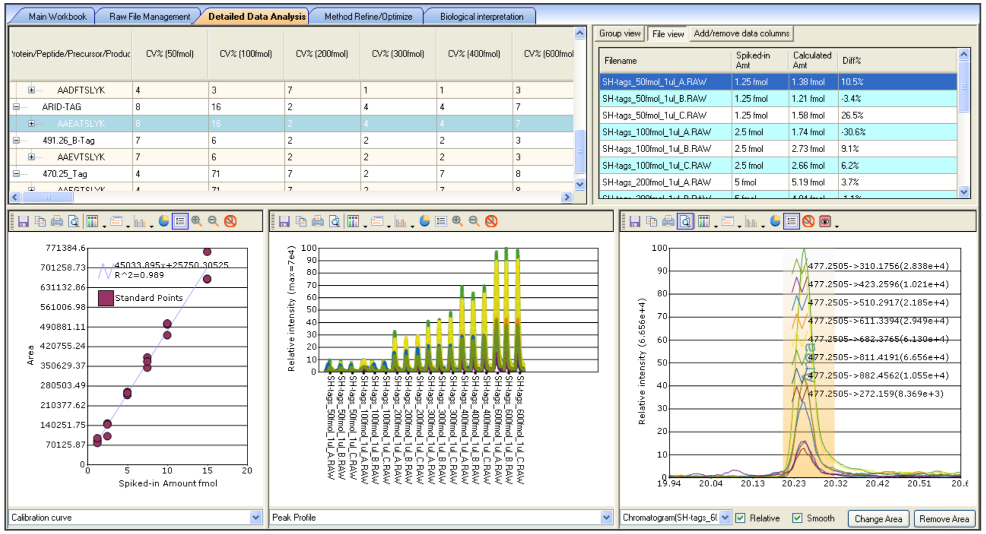
n



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.24 fmol	-0.8%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.24 fmol	-1.0%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.3 fmol	3.9%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.4 fmol	-4.2%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.49 fmol	-0.6%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.54 fmol	1.7%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.76 fmol	15.2%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.15 fmol	3.0%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.01 fmol	0.3%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.37 fmol	-1.8%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.26 fmol	-3.1%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.22 fmol	-3.8%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.77 fmol	-2.3%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.47 fmol	-5.3%
SH-tags_400fmol_1ul_C.RAW	10 fmol	9.53 fmol	-4.7%
SH-tags_600fmol_1ul_A.RAW	15 fmol	14.7 fmol	-2.0%
SH-tags_600fmol_1ul_B.RAW	15 fmol	15.9 fmol	6.0%
SH-tags_600fmol_1ul_C.RAW	15 fmol	15.42 fmol	2.8%

AAEATSLYK

o

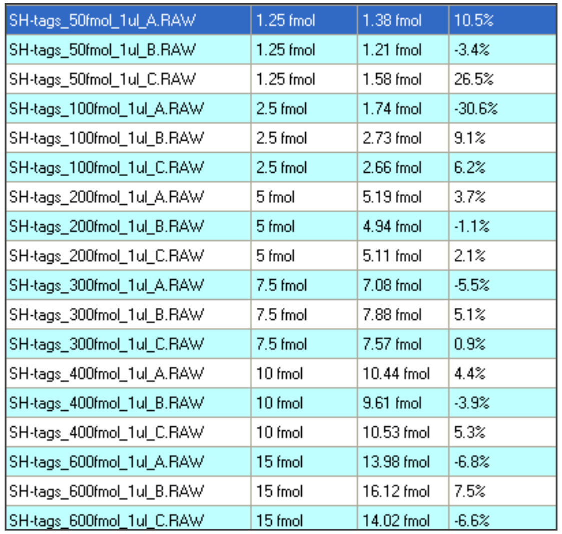


Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
Group view
File view
Add/remove data columns
Calibration curve
Peak Profile
Chromatogram(SH-tags_6(
Relative
Smooth
Change Area
Remove Area
45033.895x+25750.30525
R^2=0.989
Standard Points
Area
Spiked-in Amount fmol
Relative intensity (max=7e4)
Relative intensity (6.656e+4)
477.2505->310.1756(2.838e+4)
477.2505->423.2596(1.021e+4)
477.2505->510.2917(2.185e+4)
477.2505->611.3394(2.949e+4)
477.2505->682.3765(6.130e+4)
477.2505->811.4191(6.656e+4)
477.2505->882.4562(1.055e+4)
477.2505->272.159(8.369e+3)

'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
⊞ AADFTSLYK	4	3	7	1	1	3
⊟ ARID-TAG	8	16	2	4	4	7
⊞ AAEATSLYK	8	16	2	4	4	7
⊟ 491.26_B-Tag	7	6	2	2	2	3
⊞ AAEVTSLYK	7	6	2	2	2	3
⊟ 470.25_Tag	4	71	7	2	7	8

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.38 fmol	10.5%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.21 fmol	-3.4%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.58 fmol	26.5%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	1.74 fmol	-30.6%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.73 fmol	9.1%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.66 fmol	6.2%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.19 fmol	3.7%

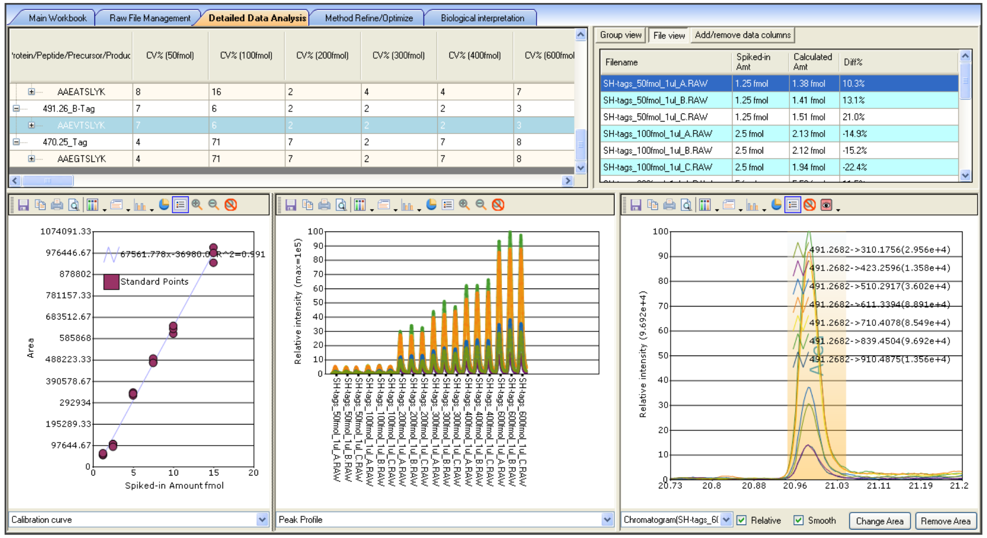
p



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.38 fmol	10.5%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.21 fmol	-3.4%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.58 fmol	26.5%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	1.74 fmol	-30.6%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.73 fmol	9.1%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	2.66 fmol	6.2%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.19 fmol	3.7%
SH-tags_200fmol_1ul_B.RAW	5 fmol	4.94 fmol	-1.1%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.11 fmol	2.1%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.08 fmol	-5.5%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.88 fmol	5.1%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.57 fmol	0.9%
SH-tags_400fmol_1ul_A.RAW	10 fmol	10.44 fmol	4.4%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.61 fmol	-3.9%
SH-tags_400fmol_1ul_C.RAW	10 fmol	10.53 fmol	5.3%
SH-tags_600fmol_1ul_A.RAW	15 fmol	13.98 fmol	-6.8%
SH-tags_600fmol_1ul_B.RAW	15 fmol	16.12 fmol	7.5%
SH-tags_600fmol_1ul_C.RAW	15 fmol	14.02 fmol	-6.6%

AAEVTSLYK

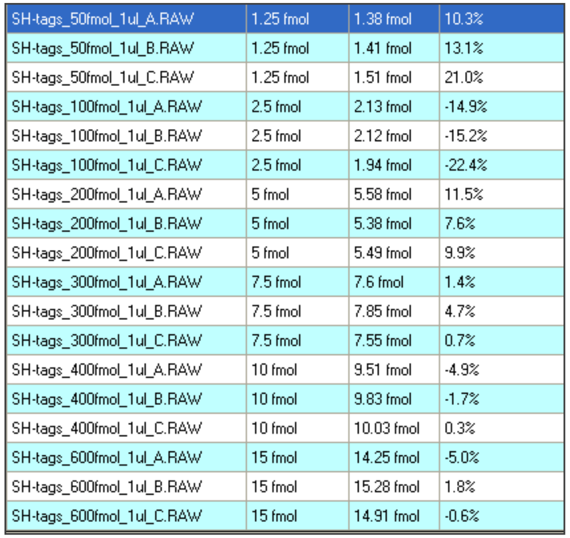
q


Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
Group view
File view
Add/remove data columns
67561.778x-36980.0 R^2=0.991
Standard Points
Area
Spiked-in Amount fmol
Relative intensity (max=1e5)
Relative intensity (9.692e+4)
491.2682->310.1756(2.956e+4)
491.2682->423.2596(1.358e+4)
491.2682->510.2917(3.602e+4)
491.2682->611.3394(8.891e+4)
491.2682->710.4078(8.549e+4)
491.2682->839.4504(9.692e+4)
491.2682->910.4875(1.356e+4)
Calibration curve
Peak Profile
Chromatogram(SH-tags_6
☑ Relative
☑ Smooth
Change Area
Remove Area

'rotein/Peptide/Precursor/Produc	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol
⊞ AAEATSLYK	8	16	2	4	4	7
⊟ 491.26_B-Tag	7	6	2	2	2	3
⊞ AAEVTSLYK	7	6	2	2	2	3
⊟ 470.25_Tag	4	71	7	2	7	8
⊞ AAEGTSLYK	4	71	7	2	7	8

Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.38 fmol	10.3%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.41 fmol	13.1%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.51 fmol	21.0%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.13 fmol	-14.9%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.12 fmol	-15.2%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	1.94 fmol	-22.4%

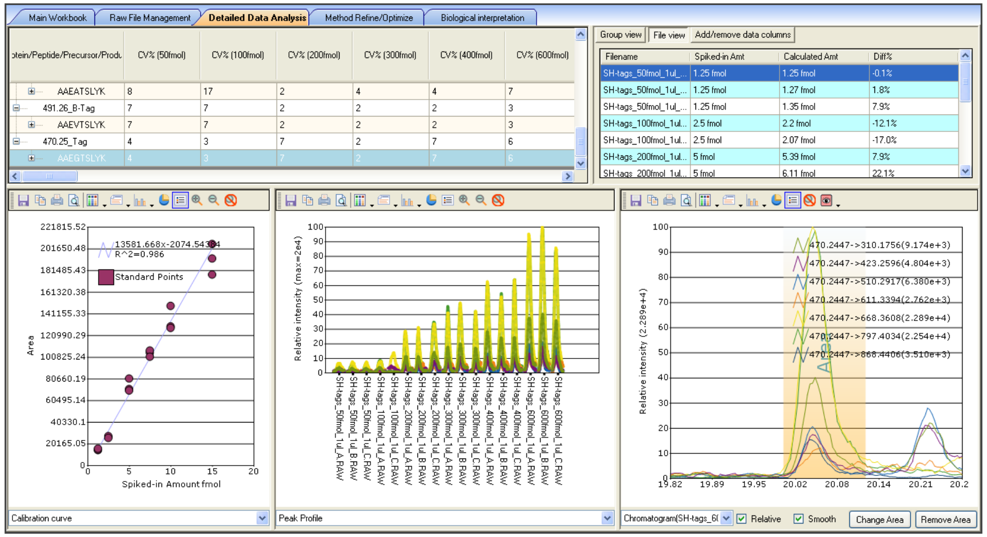
r



SH-tags_50fmol_1ul_A.RAW	1.25 fmol	1.38 fmol	10.3%
SH-tags_50fmol_1ul_B.RAW	1.25 fmol	1.41 fmol	13.1%
SH-tags_50fmol_1ul_C.RAW	1.25 fmol	1.51 fmol	21.0%
SH-tags_100fmol_1ul_A.RAW	2.5 fmol	2.13 fmol	-14.9%
SH-tags_100fmol_1ul_B.RAW	2.5 fmol	2.12 fmol	-15.2%
SH-tags_100fmol_1ul_C.RAW	2.5 fmol	1.94 fmol	-22.4%
SH-tags_200fmol_1ul_A.RAW	5 fmol	5.58 fmol	11.5%
SH-tags_200fmol_1ul_B.RAW	5 fmol	5.38 fmol	7.6%
SH-tags_200fmol_1ul_C.RAW	5 fmol	5.49 fmol	9.9%
SH-tags_300fmol_1ul_A.RAW	7.5 fmol	7.6 fmol	1.4%
SH-tags_300fmol_1ul_B.RAW	7.5 fmol	7.85 fmol	4.7%
SH-tags_300fmol_1ul_C.RAW	7.5 fmol	7.55 fmol	0.7%
SH-tags_400fmol_1ul_A.RAW	10 fmol	9.51 fmol	-4.9%
SH-tags_400fmol_1ul_B.RAW	10 fmol	9.83 fmol	-1.7%
SH-tags_400fmol_1ul_C.RAW	10 fmol	10.03 fmol	0.3%
SH-tags_600fmol_1ul_A.RAW	15 fmol	14.25 fmol	-5.0%
SH-tags_600fmol_1ul_B.RAW	15 fmol	15.28 fmol	1.8%
SH-tags_600fmol_1ul_C.RAW	15 fmol	14.91 fmol	-0.6%

AAEGTSLYK

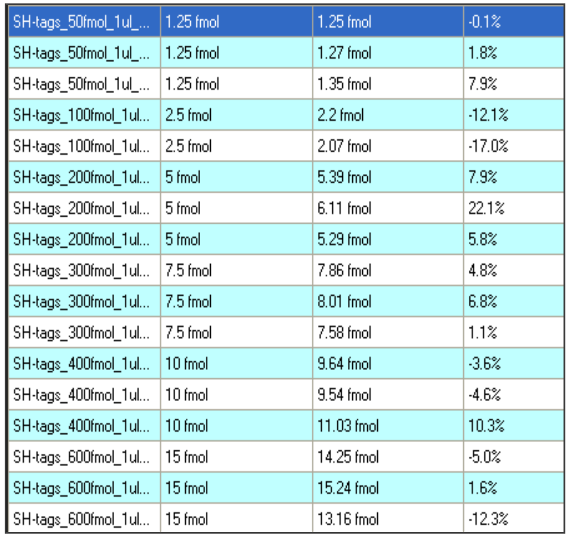
s



otein/Peptide/Precursor/Produ	CV% (50fmol)	CV% (100fmol)	CV% (200fmol)	CV% (300fmol)	CV% (400fmol)	CV% (600fmol)
⊞ AAEATSLYK	8	17	2	4	4	7
⊟ 491.26_B-Tag	7	7	2	2	2	3
⊞ AAEVTSLYK	7	7	2	2	2	3
⊟ 470.25_Tag	4	3	7	2	7	6
⊞ AAEGTSLYK	4	3	7	2	7	6

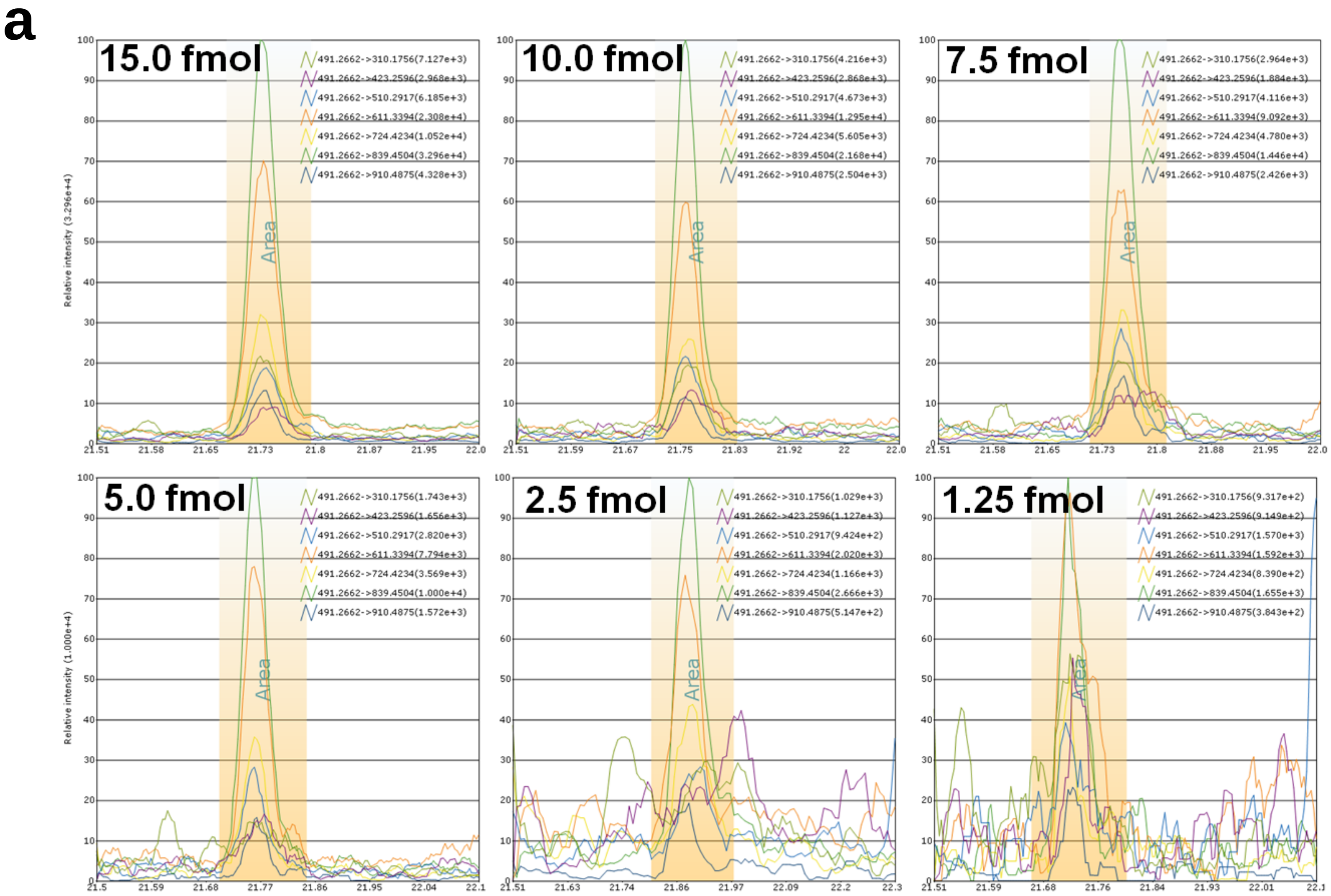
Filename	Spiked-in Amt	Calculated Amt	Diff%
SH-tags_50fmol_1ul_...	1.25 fmol	1.25 fmol	-0.1%
SH-tags_50fmol_1ul_...	1.25 fmol	1.27 fmol	1.8%
SH-tags_50fmol_1ul_...	1.25 fmol	1.35 fmol	7.9%
SH-tags_100fmol_1ul...	2.5 fmol	2.2 fmol	-12.1%
SH-tags_100fmol_1ul...	2.5 fmol	2.07 fmol	-17.0%
SH-tags_200fmol_1ul...	5 fmol	5.39 fmol	7.9%
SH-tags_200fmol_1ul...	5 fmol	6.11 fmol	22.1%

t

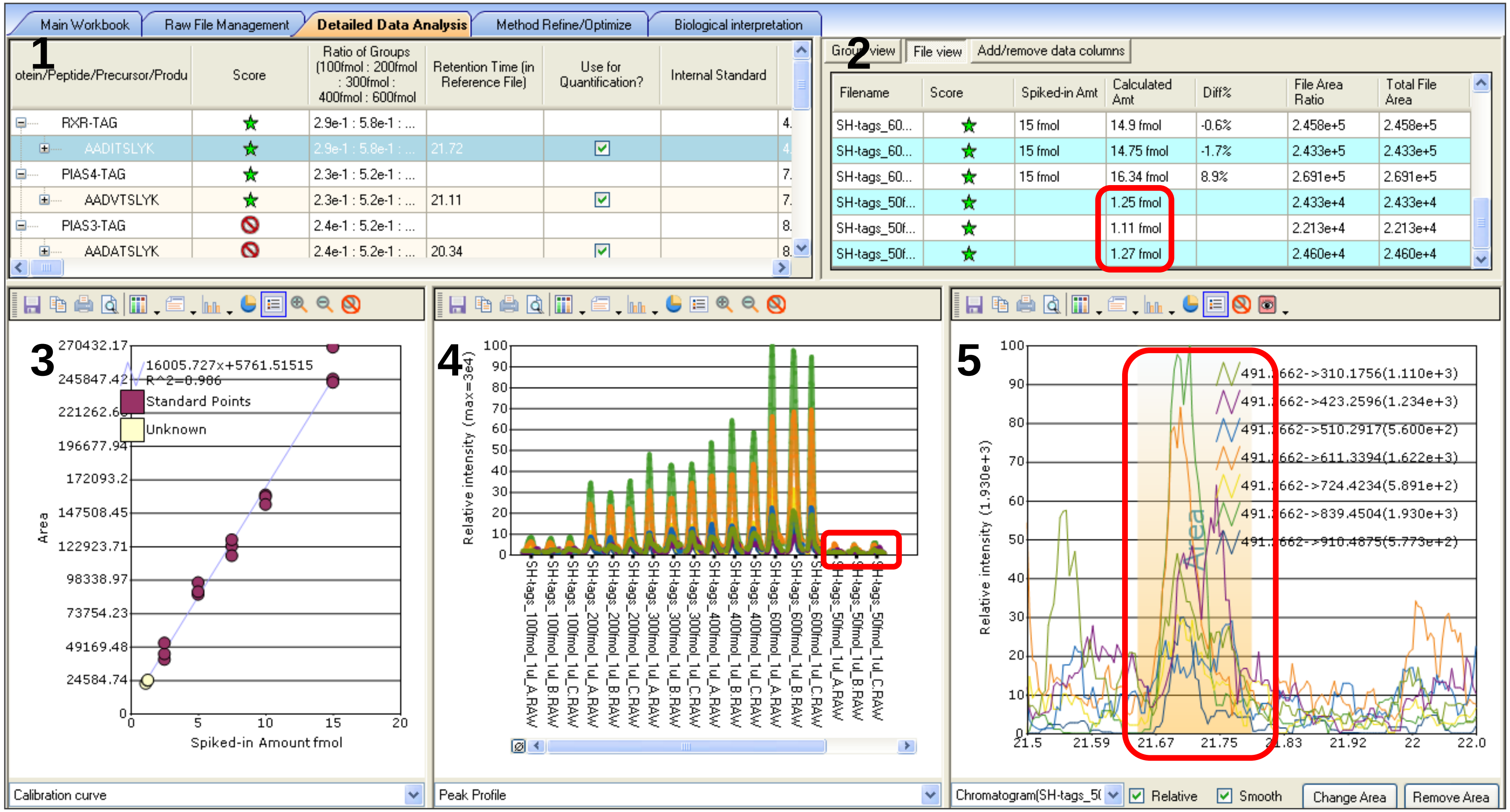


SH-tags_50fmol_1ul_...	1.25 fmol	1.25 fmol	-0.1%
SH-tags_50fmol_1ul_...	1.25 fmol	1.27 fmol	1.8%
SH-tags_50fmol_1ul_...	1.25 fmol	1.35 fmol	7.9%
SH-tags_100fmol_1ul...	2.5 fmol	2.2 fmol	-12.1%
SH-tags_100fmol_1ul...	2.5 fmol	2.07 fmol	-17.0%
SH-tags_200fmol_1ul...	5 fmol	5.39 fmol	7.9%
SH-tags_200fmol_1ul...	5 fmol	6.11 fmol	22.1%
SH-tags_200fmol_1ul...	5 fmol	5.29 fmol	5.8%
SH-tags_300fmol_1ul...	7.5 fmol	7.86 fmol	4.8%
SH-tags_300fmol_1ul...	7.5 fmol	8.01 fmol	6.8%
SH-tags_300fmol_1ul...	7.5 fmol	7.58 fmol	1.1%
SH-tags_400fmol_1ul...	10 fmol	9.64 fmol	-3.6%
SH-tags_400fmol_1ul...	10 fmol	9.54 fmol	-4.6%
SH-tags_400fmol_1ul...	10 fmol	11.03 fmol	10.3%
SH-tags_600fmol_1ul...	15 fmol	14.25 fmol	-5.0%
SH-tags_600fmol_1ul...	15 fmol	15.24 fmol	1.6%
SH-tags_600fmol_1ul...	15 fmol	13.16 fmol	-12.3%

Supplementary Figure 10. Standard curves obtained with the ten SH-Quant tags in 3T3-L1 nuclear extract.
(a,c,e,g,i,k,m,o,q,s) Standard curves of all ten SH-Quant tags. The Pinpoint window provides a summary of the evaluation of the peptide-tags of interest. Subpanels 1 and 2 show an overview of the coefficients of variation per peptide-tag per time-point. Subpanel 3 shows the calculated standard curve regression line. Subpanel 4 shows the peak profiles of all technical replicates. Subpanel 5 shows the calculated peak area and chromatogram of peptide-tags transitions. Panel legend and numbering applies to all figures. (b,d,f,h,j,l,n,p,r,t) Calculated amounts of spiked samples.


a
15.0 fmol
491.2662->310.1756(7.127e+3)
491.2662->423.2596(2.968e+3)
491.2662->510.2917(6.185e+3)
491.2662->611.3394(2.308e+4)
491.2662->724.4234(1.052e+4)
491.2662->839.4504(3.296e+4)
491.2662->910.4875(4.328e+3)
Area
Relative intensity (3.296e+4)
100
90
80
70
60
50
40
30
20
10
0
21.51
21.58
21.65
21.73
21.8
21.87
21.95
22.0
10.0 fmol
491.2662->310.1756(4.216e+3)
491.2662->423.2596(2.868e+3)
491.2662->510.2917(4.673e+3)
491.2662->611.3394(1.295e+4)
491.2662->724.4234(5.605e+3)
491.2662->839.4504(2.168e+4)
491.2662->910.4875(2.504e+3)
Area
21.51
21.59
21.67
21.75
21.83
21.92
22
22.0
7.5 fmol
491.2662->310.1756(2.964e+3)
491.2662->423.2596(1.884e+3)
491.2662->510.2917(4.116e+3)
491.2662->611.3394(9.092e+3)
491.2662->724.4234(4.780e+3)
491.2662->839.4504(1.446e+4)
491.2662->910.4875(2.426e+3)
Area
21.51
21.58
21.65
21.73
21.8
21.88
21.95
22.0
5.0 fmol
491.2662->310.1756(1.743e+3)
491.2662->423.2596(1.656e+3)
491.2662->510.2917(2.820e+3)
491.2662->611.3394(7.794e+3)
491.2662->724.4234(3.569e+3)
491.2662->839.4504(1.000e+4)
491.2662->910.4875(1.572e+3)
Area
Relative intensity (1.000e+4)
21.5
21.59
21.68
21.77
21.86
21.95
22.04
22.1
2.5 fmol
491.2662->310.1756(1.029e+3)
491.2662->423.2596(1.127e+3)
491.2662->510.2917(9.424e+2)
491.2662->611.3394(2.020e+3)
491.2662->724.4234(1.166e+3)
491.2662->839.4504(2.666e+3)
491.2662->910.4875(5.147e+2)
Area
21.51
21.63
21.74
21.86
21.97
22.09
22.2
22.3
1.25 fmol
491.2662->310.1756(9.317e+2)
491.2662->423.2596(9.149e+2)
491.2662->510.2917(1.570e+3)
491.2662->611.3394(1.592e+3)
491.2662->724.4234(8.390e+2)
491.2662->839.4504(1.655e+3)
491.2662->910.4875(3.843e+2)
Area
21.51
21.59
21.68
21.76
21.84
21.93
22.01
22.1

b


Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
1
2
Group view
File view
Add/remove data columns
3
270432.17
245847.42
221262.6
196677.94
172093.2
147508.45
122923.71
98338.97
73754.23
49169.48
24584.74
0
Area
16005.727x+5761.51515
R^2=0.986
Standard Points
Unknown
0
5
10
15
20
Spiked-in Amount fmol
Calibration curve
4
Relative intensity (max=3e4)
100
90
80
70
60
50
40
30
20
10
0
SH-tags_100fmol_1ul_A.RAW
SH-tags_100fmol_1ul_B.RAW
SH-tags_100fmol_1ul_C.RAW
SH-tags_200fmol_1ul_A.RAW
SH-tags_200fmol_1ul_B.RAW
SH-tags_200fmol_1ul_C.RAW
SH-tags_300fmol_1ul_A.RAW
SH-tags_300fmol_1ul_B.RAW
SH-tags_300fmol_1ul_C.RAW
SH-tags_400fmol_1ul_A.RAW
SH-tags_400fmol_1ul_B.RAW
SH-tags_400fmol_1ul_C.RAW
SH-tags_600fmol_1ul_A.RAW
SH-tags_600fmol_1ul_B.RAW
SH-tags_600fmol_1ul_C.RAW
SH-tags_50fmol_1ul_A.RAW
SH-tags_50fmol_1ul_B.RAW
SH-tags_50fmol_1ul_C.RAW
Peak Profile
5
Relative intensity (1.930e+3)
100
90
80
70
60
50
40
30
20
10
0
21.5
21.59
21.67
21.75
21.83
21.92
22
22.0
Area
491. 662->310.1756(1.110e+3)
491. 662->423.2596(1.234e+3)
491. 662->510.2917(5.600e+2)
491. 662->611.3394(1.622e+3)
491. 662->724.4234(5.891e+2)
491. 662->839.4504(1.930e+3)
491. 662->910.4875(5.773e+2)
Chromatogram(SH-tags_5(
☑ Relative
☑ Smooth
Change Area
Remove Area

otein/Peptide/Precursor/Produ	Score	Ratio of Groups (100fmol : 200fmol : 300fmol : 400fmol : 600fmol	Retention Time (in Reference File)	Use for Quantification?	Internal Standard	
⊟ RXR-TAG	★	2.9e-1 : 5.8e-1 : ...				4.
⊞ AADITSLYK	★	2.9e-1 : 5.8e-1 : ...	21.72	☑		4.
⊟ PIAS4-TAG	★	2.3e-1 : 5.2e-1 : ...				7.
⊞ AADVTSLYK	★	2.3e-1 : 5.2e-1 : ...	21.11	☑		7.
⊟ PIAS3-TAG	⊘	2.4e-1 : 5.2e-1 : ...				8.
⊞ AADATSLYK	⊘	2.4e-1 : 5.2e-1 : ...	20.34	☑		8.

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
SH-tags_60...	★	15 fmol	14.9 fmol	-0.6%	2.458e+5	2.458e+5
SH-tags_60...	★	15 fmol	14.75 fmol	-1.7%	2.433e+5	2.433e+5
SH-tags_60...	★	15 fmol	16.34 fmol	8.9%	2.691e+5	2.691e+5
SH-tags_50f...	★		1.25 fmol		2.433e+4	2.433e+4
SH-tags_50f...	★		1.11 fmol		2.213e+4	2.213e+4
SH-tags_50f...	★		1.27 fmol		2.460e+4	2.460e+4

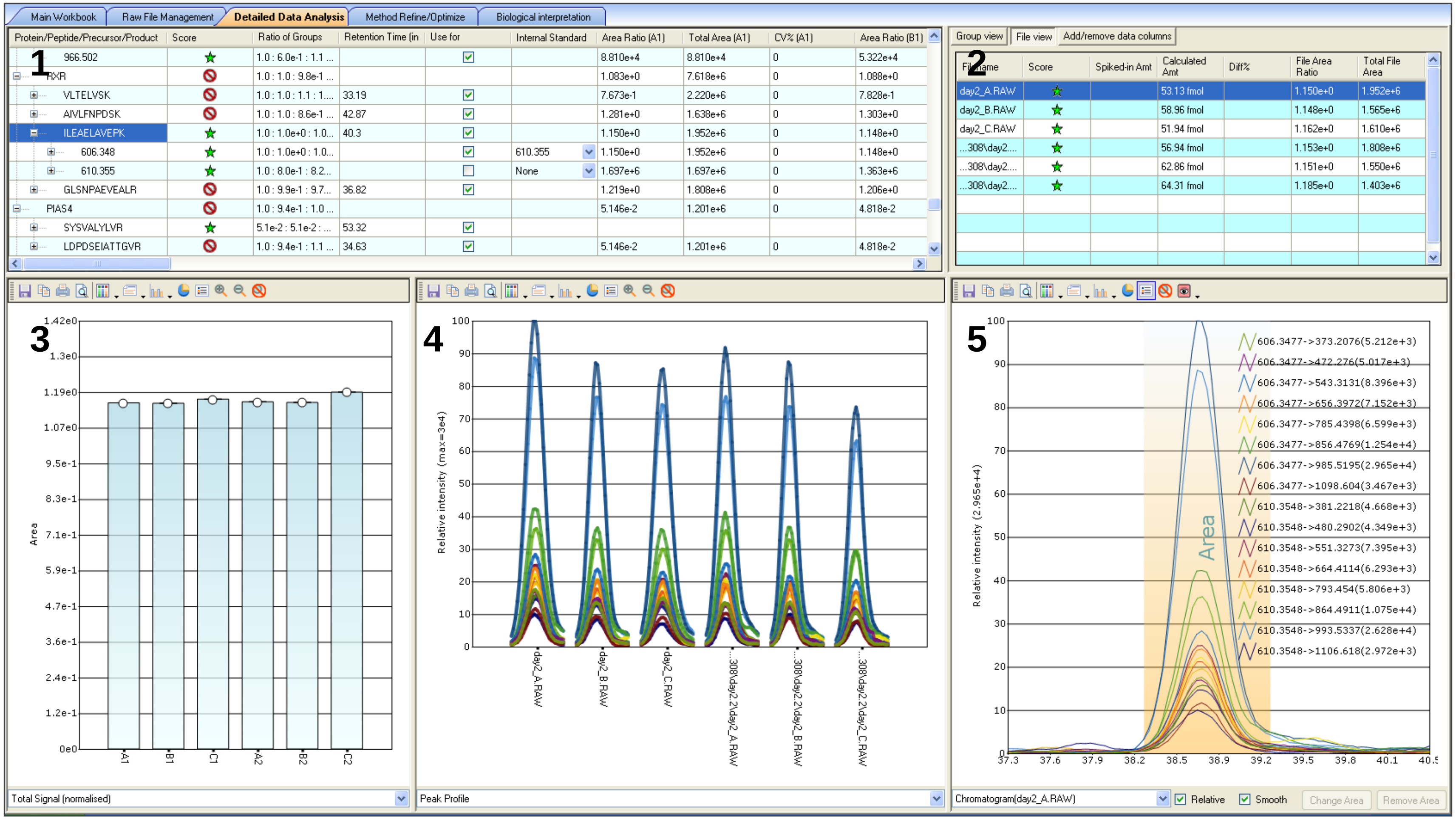
Supplementary Figure 11. Determination of the lower limit of quantitation (LLOQ) of SH-Quant tags in multiplexed samples.

(a) Example of the calculated peak area of a SH-Quant tag (AADITSLYK) spiked at concentrations of 15.0 fmol, 5.0 fmol, 2.50 fmol and 1.25 fmol (unknown treated sample). Little variation (≤ 0.15 min) in elution time was observed for all monitored concentrations.

(b) The Pinpoint window provides a summary of the evaluation of the peptide-tags of interest. Subpanel 1 presents an overview of the all peptides monitored. Subpanel 2 shows the calculated peptide-tag amounts. Subpanel 3 shows the calculated standard curve regression line. Subpanel 4 shows peak profiles of different spikes of all technical replicates. Subpanel 5 shows the calculated peak area and chromatogram of the transitions of a lower-end peptide-tag. The calculated concentrations of all data points within the calibration curve were found to lie in the acceptable range of ± 15% variation. The red boxes highlight the peak area and concentrations found for the “unknown” sample. The three technical replicates show little variation in concentrations (1.25 fmol, 1.11 fmol, 1.27 fmol, mean = 1.21 ± 0.09, variation ± 7.4%), which fulfills the acceptable range of variation as set by the FDA guidelines for the LLOQ (± 20%). A complete overview of all ten SH-Quant tags with the calculated concentrations can be found in Supplementary Figure 10.

a

RXRα



Main Workbook | Raw File Management | Detailed Data Analysis | Method Refine/Optimize | Biological interpretation

Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
966.502	★	1.0 : 6.0e-1 : 1.1 ...		☑		8.810e+4	8.810e+4	0	5.322e+4
RXR	⊘	1.0 : 1.0 : 9.8e-1 ...				1.083e+0	7.618e+6	0	1.088e+0
VLTELVSK	⊘	1.0 : 1.0 : 1.1 : 1....	33.19	☑		7.673e-1	2.220e+6	0	7.828e-1
AIVLFNPDSK	⊘	1.0 : 1.0 : 8.6e-1 ...	42.87	☑		1.281e+0	1.638e+6	0	1.303e+0
ILEAELAVEPK	★	1.0 : 1.0e+0 : 1.0...	40.3	☑		1.150e+0	1.952e+6	0	1.148e+0
606.348	★	1.0 : 1.0e+0 : 1.0...		☑	610.355	1.150e+0	1.952e+6	0	1.148e+0
610.355	★	1.0 : 8.0e-1 : 8.2...		□	None	1.697e+6	1.697e+6	0	1.363e+6
GLSNPAEVEALR	⊘	1.0 : 9.9e-1 : 9.7...	36.82	☑		1.219e+0	1.808e+6	0	1.206e+0
PIAS4	⊘	1.0 : 9.4e-1 : 1.0 ...				5.146e-2	1.201e+6	0	4.818e-2
SYSVALYLVR	★	5.1e-2 : 5.1e-2 : ...	53.32	☑					
LDPDSEIATTGVR	⊘	1.0 : 9.4e-1 : 1.1 ...	34.63	☑		5.146e-2	1.201e+6	0	4.818e-2

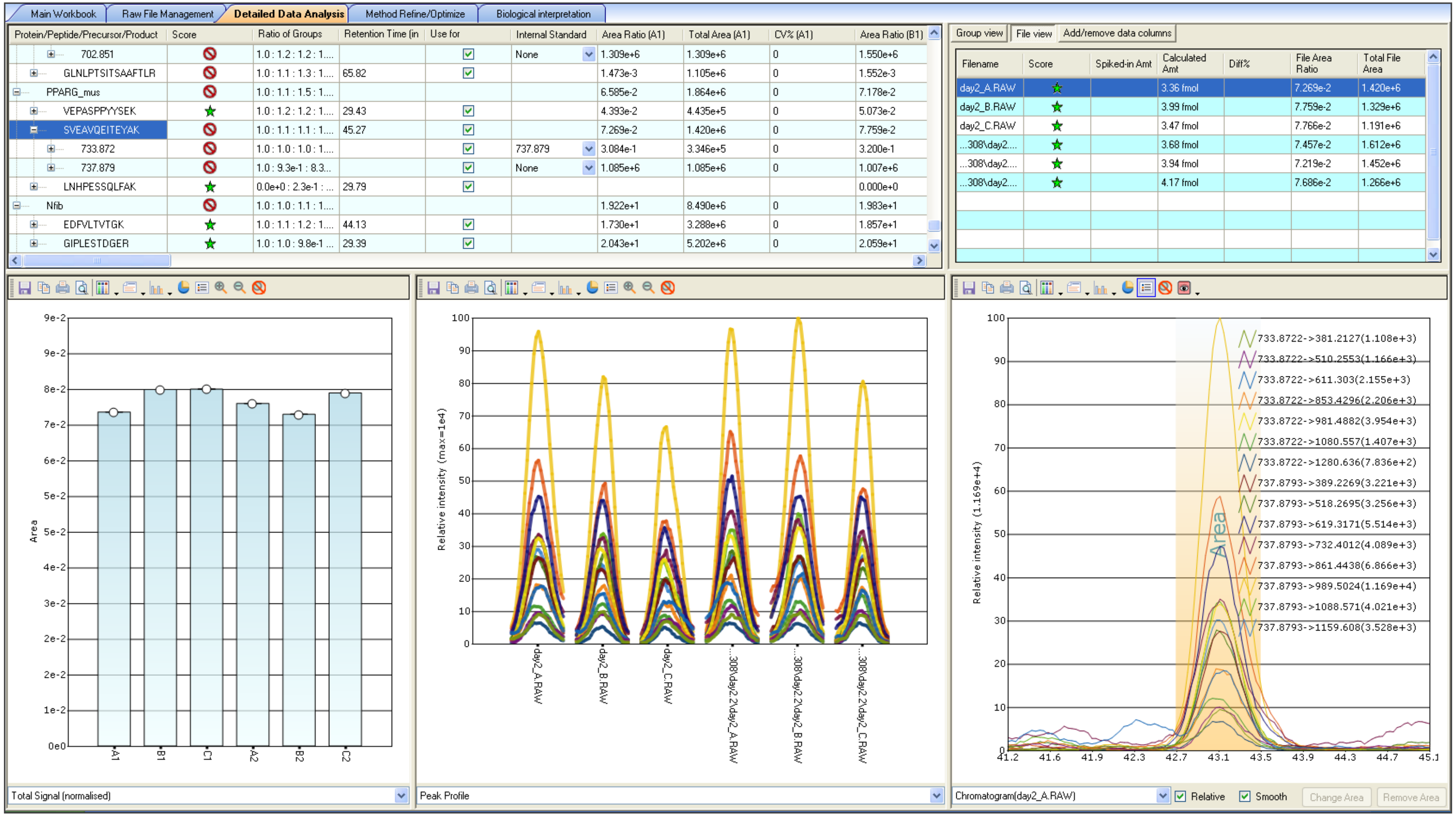
Group view | File view | Add/remove data columns

File name	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	★		53.13 fmol		1.150e+0	1.952e+6
day2_B.RAW	★		58.96 fmol		1.148e+0	1.565e+6
day2_C.RAW	★		51.94 fmol		1.162e+0	1.610e+6
...308\day2....	★		56.94 fmol		1.153e+0	1.808e+6
...308\day2....	★		62.86 fmol		1.151e+0	1.550e+6
...308\day2....	★		64.31 fmol		1.185e+0	1.403e+6

1
2
3
Area
1.42e0
1.3e0
1.19e0
1.07e0
9.5e-1
8.3e-1
7.1e-1
5.9e-1
4.7e-1
3.6e-1
2.4e-1
1.2e-1
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
4
Relative intensity (max=3e4)
100
90
80
70
60
50
40
30
20
10
0
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
5
Relative intensity (2.965e+4)
Area
606.3477->373.2076(5.212e+3)
606.3477->472.276(5.017e+3)
606.3477->543.3131(8.396e+3)
606.3477->656.3972(7.152e+3)
606.3477->785.4398(6.599e+3)
606.3477->856.4769(1.254e+4)
606.3477->985.5195(2.965e+4)
606.3477->1098.604(3.467e+3)
610.3548->381.2218(4.668e+3)
610.3548->480.2902(4.349e+3)
610.3548->551.3273(7.395e+3)
610.3548->664.4114(6.293e+3)
610.3548->793.454(5.806e+3)
610.3548->864.4911(1.075e+4)
610.3548->993.5337(2.628e+4)
610.3548->1106.618(2.972e+3)
37.3
37.6
37.9
38.2
38.5
38.9
39.2
39.5
39.8
40.1
40.5
Chromatogram(day2_A.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

b

PPARγ



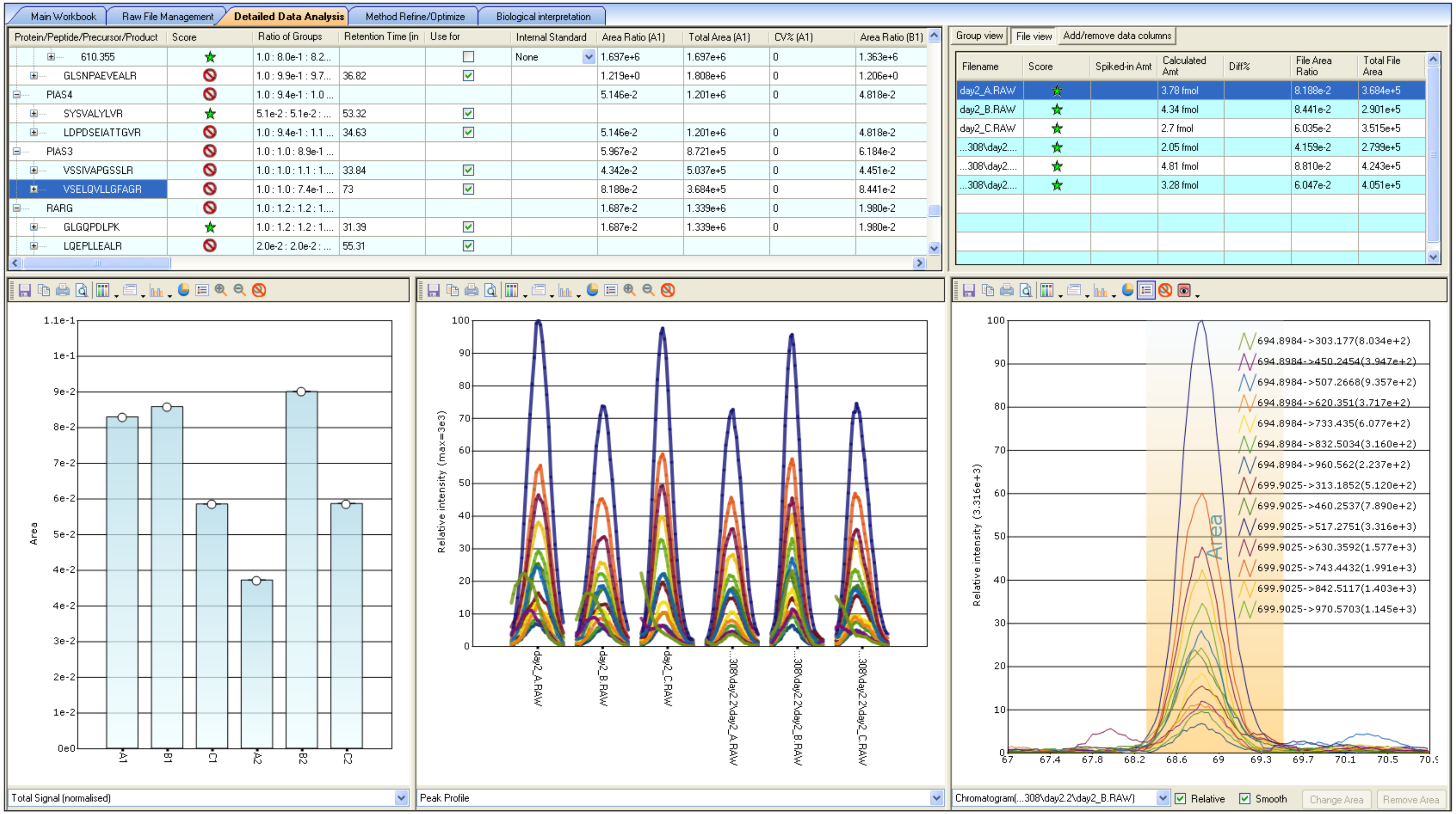
Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
702.851	🚫	1.0 : 1.2 : 1.2 : 1....		☑	None	1.309e+6	1.309e+6	0	1.550e+6
GLNLPTSITSAAFTLR	🚫	1.0 : 1.1 : 1.3 : 1....	65.82	☑		1.473e-3	1.105e+6	0	1.552e-3
PPARG_mus	🚫	1.0 : 1.1 : 1.5 : 1....				6.585e-2	1.864e+6	0	7.178e-2
VEPASPPYYSEK	★	1.0 : 1.2 : 1.2 : 1....	29.43	☑		4.393e-2	4.435e+5	0	5.073e-2
SVEAVQEITEYAK	🚫	1.0 : 1.1 : 1.1 : 1....	45.27	☑		7.269e-2	1.420e+6	0	7.759e-2
733.872	🚫	1.0 : 1.0 : 1.0 : 1....		☑	737.879	3.084e-1	3.346e+5	0	3.200e-1
737.879	🚫	1.0 : 9.3e-1 : 8.3...		☑	None	1.085e+6	1.085e+6	0	1.007e+6
LNHPESSQLFAK	★	0.0e+0 : 2.3e-1 : ...	29.79	☑					0.000e+0
Nfib	🚫	1.0 : 1.0 : 1.1 : 1....				1.922e+1	8.490e+6	0	1.983e+1
EDFVLTVTGK	★	1.0 : 1.1 : 1.2 : 1....	44.13	☑		1.730e+1	3.288e+6	0	1.857e+1
GIPLESTDGER	★	1.0 : 1.0 : 9.8e-1 ...	29.39	☑		2.043e+1	5.202e+6	0	2.059e+1

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	★		3.36 fmol		7.269e-2	1.420e+6
day2_B.RAW	★		3.99 fmol		7.759e-2	1.329e+6
day2_C.RAW	★		3.47 fmol		7.766e-2	1.191e+6
...308\day2....	★		3.68 fmol		7.457e-2	1.612e+6
...308\day2....	★		3.94 fmol		7.219e-2	1.452e+6
...308\day2....	★		4.17 fmol		7.686e-2	1.266e+6

Area
9e-2
9e-2
8e-2
7e-2
6e-2
5e-2
5e-2
4e-2
3e-2
2e-2
2e-2
1e-2
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
Relative intensity (max=1e4)
100
90
80
70
60
50
40
30
20
10
0
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (1.169e+4)
Area
733.8722->381.2127(1.108e+3)
733.8722->510.2553(1.166e+3)
733.8722->611.303(2.155e+3)
733.8722->853.4296(2.206e+3)
733.8722->981.4882(3.954e+3)
733.8722->1080.557(1.407e+3)
733.8722->1280.636(7.836e+2)
737.8793->389.2269(3.221e+3)
737.8793->518.2695(3.256e+3)
737.8793->619.3171(5.514e+3)
737.8793->732.4012(4.089e+3)
737.8793->861.4438(6.866e+3)
737.8793->989.5024(1.169e+4)
737.8793->1088.571(4.021e+3)
737.8793->1159.608(3.528e+3)
41.2
41.6
41.9
42.3
42.7
43.1
43.5
43.9
44.3
44.7
45.1
Chromatogram(day2_A.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

C

PIAS3

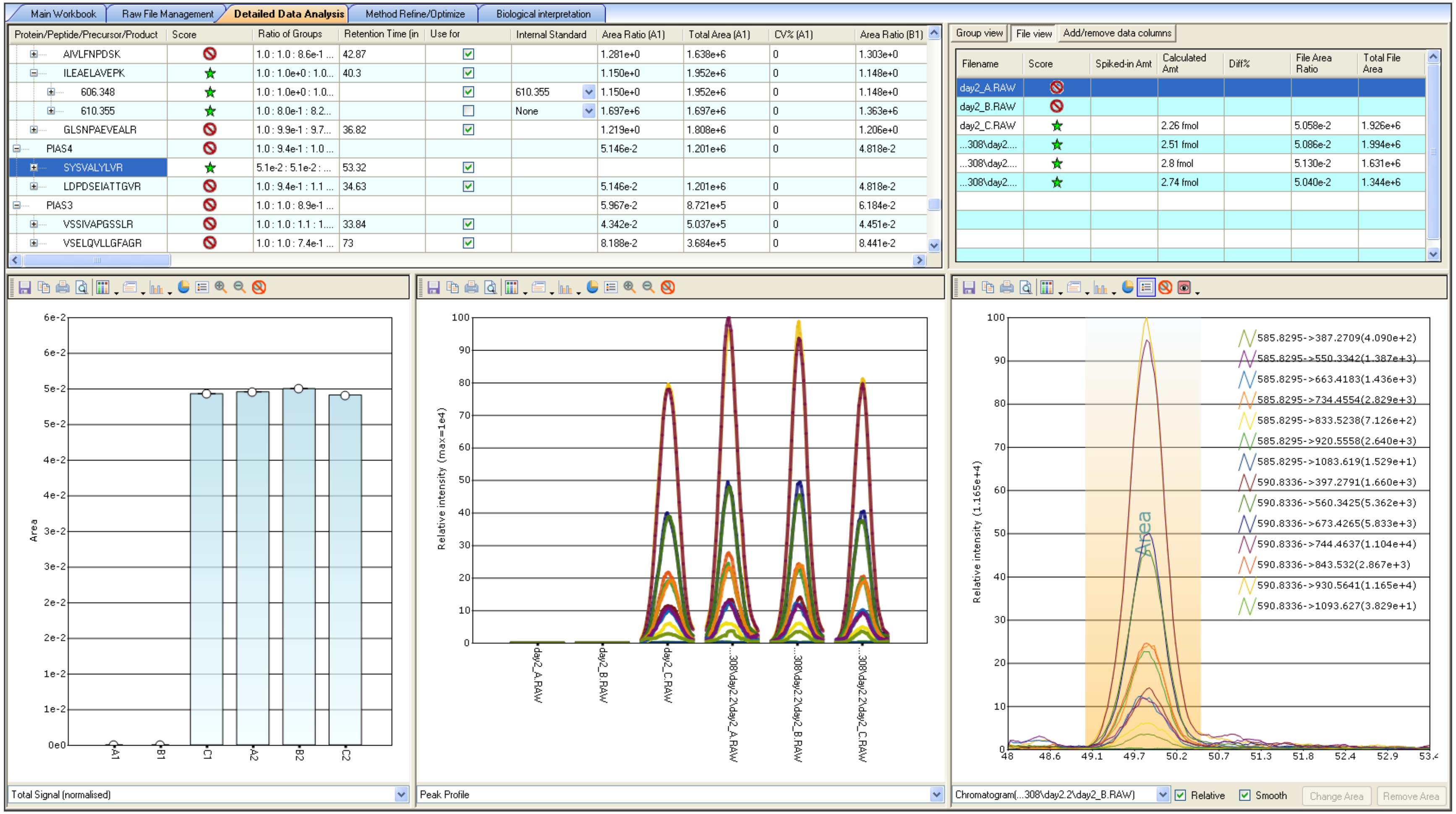

Main Workbook
Raw File Management
Detailed Data Analysis
Method Refine/Optimize
Biological interpretation
Group view
File view
Add/remove data columns
1.1e-1
1e-1
9e-2
8e-2
7e-2
6e-2
5e-2
4e-2
4e-2
3e-2
2e-2
1e-2
0e0
Area
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
100
90
80
70
60
50
40
30
20
10
0
Relative intensity (max=3e3)
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
100
90
80
70
60
50
40
30
20
10
0
Relative intensity (3.316e+3)
67
67.4
67.8
68.2
68.6
69
69.3
69.7
70.1
70.5
70.9
Area
694.8984->303.177(8.034e+2)
694.8984->450.2454(3.947e+2)
694.8984->507.2668(9.357e+2)
694.8984->620.351(3.717e+2)
694.8984->733.435(6.077e+2)
694.8984->832.5034(3.160e+2)
694.8984->960.562(2.237e+2)
699.9025->313.1852(5.120e+2)
699.9025->460.2537(7.890e+2)
699.9025->517.2751(3.316e+3)
699.9025->630.3592(1.577e+3)
699.9025->743.4432(1.991e+3)
699.9025->842.5117(1.403e+3)
699.9025->970.5703(1.145e+3)
Chromatogram(...308\day2.2\day2_B.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
610.355	★	1.0 : 8.0e-1 : 8.2...		□	None	1.697e+6	1.697e+6	0	1.363e+6
GLSNPAEVEALR	⊘	1.0 : 9.9e-1 : 9.7...	36.82	☑		1.219e+0	1.808e+6	0	1.206e+0
PIAS4	⊘	1.0 : 9.4e-1 : 1.0 ...				5.146e-2	1.201e+6	0	4.818e-2
SYSVALYLVR	★	5.1e-2 : 5.1e-2 : ...	53.32	☑					
LDPDSEIATTGVR	⊘	1.0 : 9.4e-1 : 1.1 ...	34.63	☑		5.146e-2	1.201e+6	0	4.818e-2
PIAS3	⊘	1.0 : 1.0 : 8.9e-1 ...				5.967e-2	8.721e+5	0	6.184e-2
VSSIVAPGSSLR	⊘	1.0 : 1.0 : 1.1 : 1....	33.84	☑		4.342e-2	5.037e+5	0	4.451e-2
VSELQVLLGFAGR	⊘	1.0 : 1.0 : 7.4e-1 ...	73	☑		8.188e-2	3.684e+5	0	8.441e-2
RARG	⊘	1.0 : 1.2 : 1.2 : 1....				1.687e-2	1.339e+6	0	1.980e-2
GLGQPDLPK	★	1.0 : 1.2 : 1.2 : 1....	31.39	☑		1.687e-2	1.339e+6	0	1.980e-2
LQEPLLEALR	⊘	2.0e-2 : 2.0e-2 : ...	55.31	☑					

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	★		3.78 fmol		8.188e-2	3.684e+5
day2_B.RAW	★		4.34 fmol		8.441e-2	2.901e+5
day2_C.RAW	★		2.7 fmol		6.035e-2	3.515e+5
...308\day2....	★		2.05 fmol		4.159e-2	2.799e+5
...308\day2....	★		4.81 fmol		8.810e-2	4.243e+5
...308\day2....	★		3.28 fmol		6.047e-2	4.051e+5

d

PIAS4



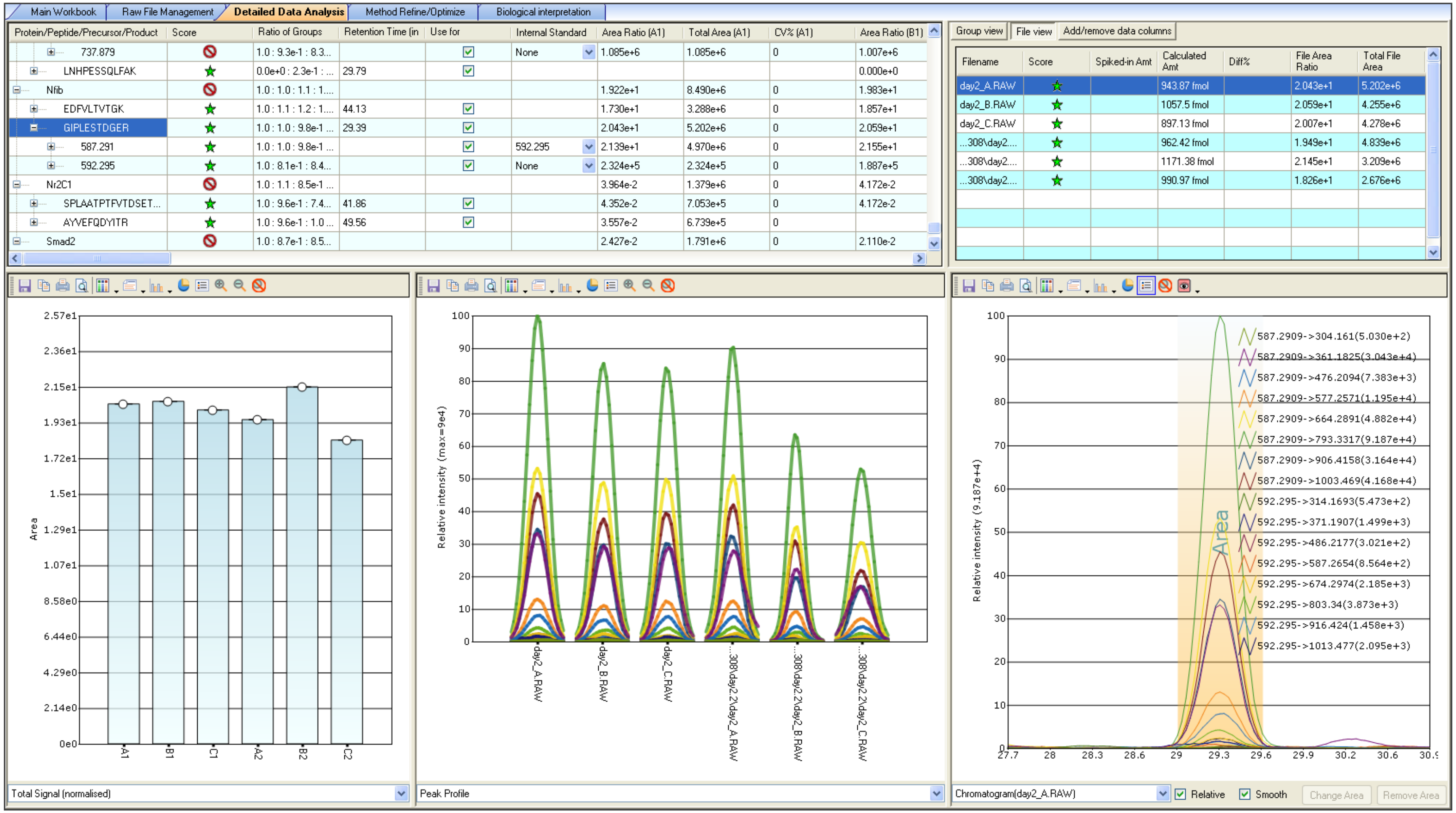
Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
⊞ AIVLFNPDSK	⊘	1.0 : 1.0 : 8.6e-1 ...	42.87	☑		1.281e+0	1.638e+6	0	1.303e+0
⊟ ILEAELAVEPK	★	1.0 : 1.0e+0 : 1.0...	40.3	☑		1.150e+0	1.952e+6	0	1.148e+0
⊞ 606.348	★	1.0 : 1.0e+0 : 1.0...		☑	610.355	1.150e+0	1.952e+6	0	1.148e+0
⊞ 610.355	★	1.0 : 8.0e-1 : 8.2...		☐	None	1.697e+6	1.697e+6	0	1.363e+6
⊞ GLSNPAEVEALR	⊘	1.0 : 9.9e-1 : 9.7...	36.82	☑		1.219e+0	1.808e+6	0	1.206e+0
⊟ PIAS4	⊘	1.0 : 9.4e-1 : 1.0 ...				5.146e-2	1.201e+6	0	4.818e-2
⊞ SYSVALYLVR	★	5.1e-2 : 5.1e-2 : ...	53.32	☑					
⊞ LDPDSEIATTGVR	⊘	1.0 : 9.4e-1 : 1.1 ...	34.63	☑		5.146e-2	1.201e+6	0	4.818e-2
⊟ PIAS3	⊘	1.0 : 1.0 : 8.9e-1 ...				5.967e-2	8.721e+5	0	6.184e-2
⊞ VSSIVAPGSSLR	⊘	1.0 : 1.0 : 1.1 : 1....	33.84	☑		4.342e-2	5.037e+5	0	4.451e-2
⊞ VSELQVLLGFAGR	⊘	1.0 : 1.0 : 7.4e-1 ...	73	☑		8.188e-2	3.684e+5	0	8.441e-2

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	⊘					
day2_B.RAW	⊘					
day2_C.RAW	★		2.26 fmol		5.058e-2	1.926e+6
...308\day2....	★		2.51 fmol		5.086e-2	1.994e+6
...308\day2....	★		2.8 fmol		5.130e-2	1.631e+6
...308\day2....	★		2.74 fmol		5.040e-2	1.344e+6

Area
6e-2
6e-2
5e-2
5e-2
4e-2
4e-2
3e-2
3e-2
2e-2
2e-2
1e-2
1e-2
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
Relative intensity (max=1e4)
100
90
80
70
60
50
40
30
20
10
0
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (1.165e+4)
Area
48
48.6
49.1
49.7
50.2
50.7
51.3
51.8
52.4
52.9
53.4
585.8295->387.2709(4.090e+2)
585.8295->550.3342(1.387e+3)
585.8295->663.4183(1.436e+3)
585.8295->734.4554(2.829e+3)
585.8295->833.5238(7.126e+2)
585.8295->920.5558(2.640e+3)
585.8295->1083.619(1.529e+1)
590.8336->397.2791(1.660e+3)
590.8336->560.3425(5.362e+3)
590.8336->673.4265(5.833e+3)
590.8336->744.4637(1.104e+4)
590.8336->843.532(2.867e+3)
590.8336->930.5641(1.165e+4)
590.8336->1093.627(3.829e+1)
Chromatogram(...308\day2.2\day2_B.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

e

NFIB



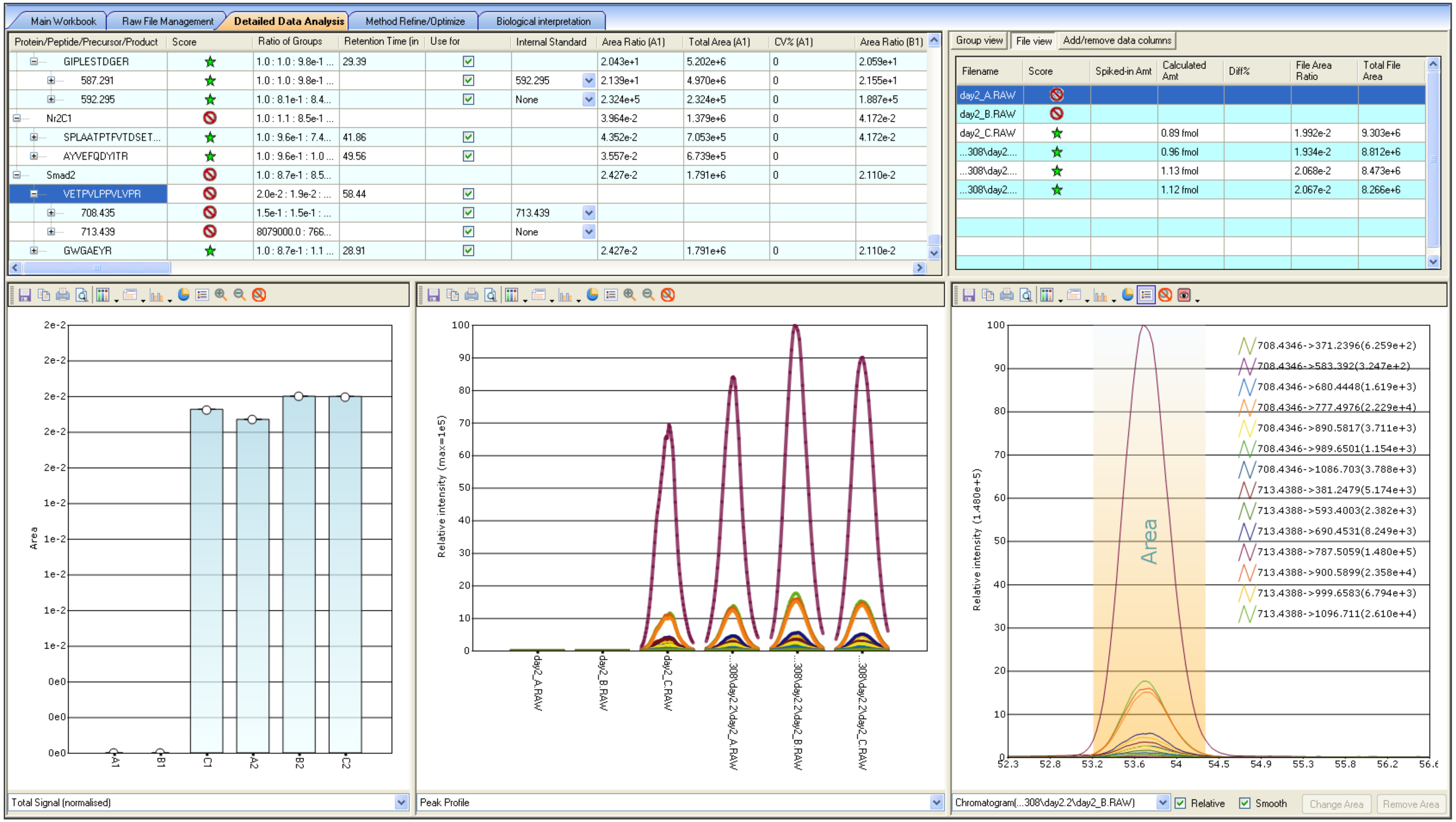
Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
⊞ 737.879	⊘	1.0 : 9.3e-1 : 8.3...		☑	None	1.085e+6	1.085e+6	0	1.007e+6
⊞ LNHPESSQLFAK	★	0.0e+0 : 2.3e-1 : ...	29.79	☑					0.000e+0
⊟ Nfib	⊘	1.0 : 1.0 : 1.1 : 1....				1.922e+1	8.490e+6	0	1.983e+1
⊞ EDFVLTVTGK	★	1.0 : 1.1 : 1.2 : 1....	44.13	☑		1.730e+1	3.288e+6	0	1.857e+1
⊟ GIPLESTDGER	★	1.0 : 1.0 : 9.8e-1 ...	29.39	☑		2.043e+1	5.202e+6	0	2.059e+1
⊞ 587.291	★	1.0 : 1.0 : 9.8e-1 ...		☑	592.295	2.139e+1	4.970e+6	0	2.155e+1
⊞ 592.295	★	1.0 : 8.1e-1 : 8.4...		☑	None	2.324e+5	2.324e+5	0	1.887e+5
⊟ Nr2C1	⊘	1.0 : 1.1 : 8.5e-1 ...				3.964e-2	1.379e+6	0	4.172e-2
⊞ SPLAATPTFVTDSET...	★	1.0 : 9.6e-1 : 7.4...	41.86	☑		4.352e-2	7.053e+5	0	4.172e-2
⊞ AYVEFQDYITR	★	1.0 : 9.6e-1 : 1.0 ...	49.56	☑		3.557e-2	6.739e+5	0	
⊟ Smad2	⊘	1.0 : 8.7e-1 : 8.5...				2.427e-2	1.791e+6	0	2.110e-2

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	★		943.87 fmol		2.043e+1	5.202e+6
day2_B.RAW	★		1057.5 fmol		2.059e+1	4.255e+6
day2_C.RAW	★		897.13 fmol		2.007e+1	4.278e+6
...308\day2....	★		962.42 fmol		1.949e+1	4.839e+6
...308\day2....	★		1171.38 fmol		2.145e+1	3.209e+6
...308\day2....	★		990.97 fmol		1.826e+1	2.676e+6

2.57e1
2.36e1
2.15e1
1.93e1
1.72e1
1.5e1
1.29e1
1.07e1
8.58e0
6.44e0
4.29e0
2.14e0
0e0
Area
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
100
90
80
70
60
50
40
30
20
10
0
Relative intensity (max=9e4)
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (9.187e+4)
27.7
28
28.3
28.6
29
29.3
29.6
29.9
30.2
30.6
30.9
Area
587.2909->304.161(5.030e+2)
587.2909->361.1825(3.043e+4)
587.2909->476.2094(7.383e+3)
587.2909->577.2571(1.195e+4)
587.2909->664.2891(4.882e+4)
587.2909->793.3317(9.187e+4)
587.2909->906.4158(3.164e+4)
587.2909->1003.469(4.168e+4)
592.295->314.1693(5.473e+2)
592.295->371.1907(1.499e+3)
592.295->486.2177(3.021e+2)
592.295->587.2654(8.564e+2)
592.295->674.2974(2.185e+3)
592.295->803.34(3.873e+3)
592.295->916.424(1.458e+3)
592.295->1013.477(2.095e+3)
Chromatogram(day2_A.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

f

SMAD2



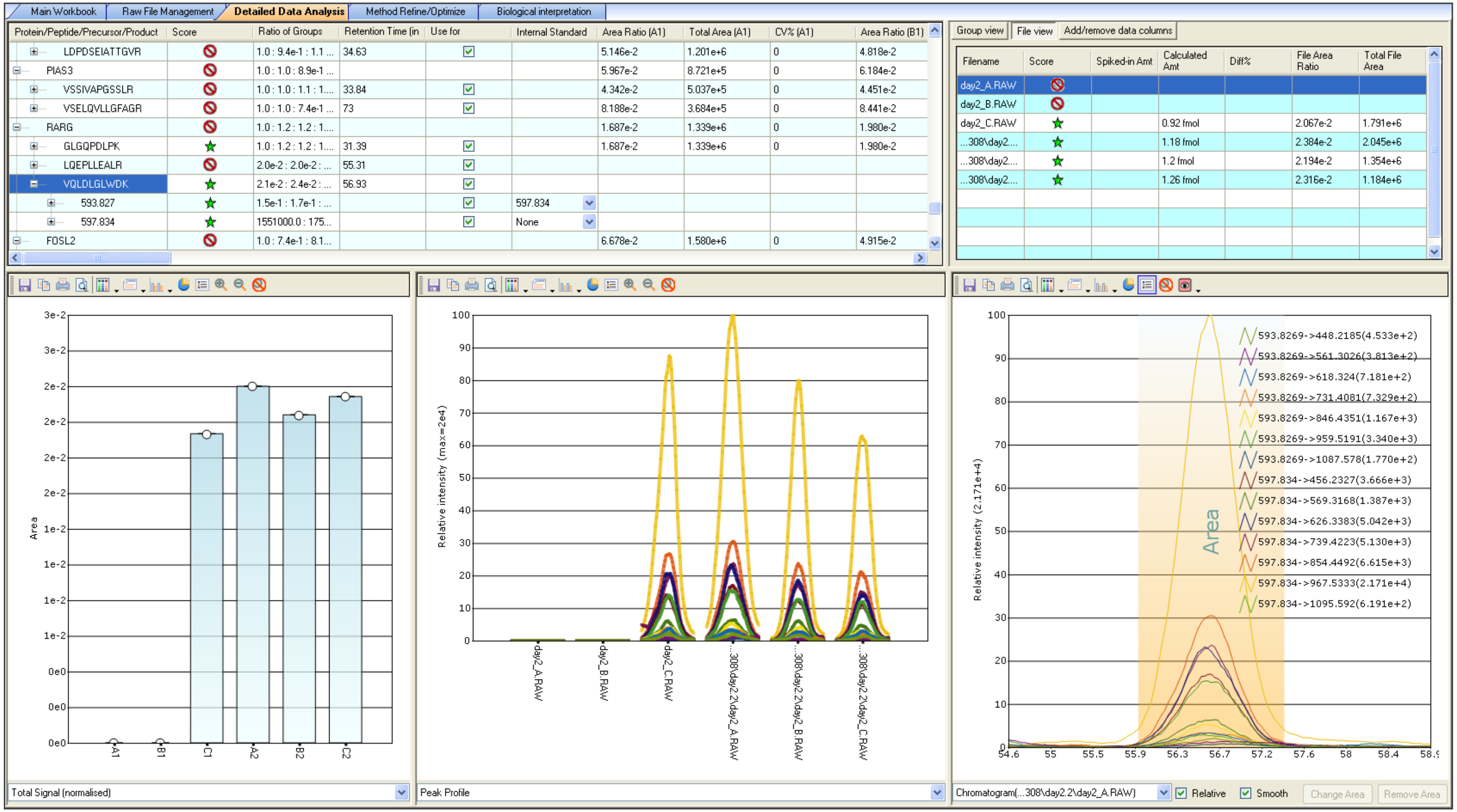
Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
⊟ GIPLESTDGER	★	1.0 : 1.0 : 9.8e-1 ...	29.39	☑		2.043e+1	5.202e+6	0	2.059e+1
⊞ 587.291	★	1.0 : 1.0 : 9.8e-1 ...		☑	592.295	2.139e+1	4.970e+6	0	2.155e+1
⊞ 592.295	★	1.0 : 8.1e-1 : 8.4...		☑	None	2.324e+5	2.324e+5	0	1.887e+5
⊟ Nr2C1	⊘	1.0 : 1.1 : 8.5e-1 ...				3.964e-2	1.379e+6	0	4.172e-2
⊞ SPLAATPTFVTDSET...	★	1.0 : 9.6e-1 : 7.4...	41.86	☑		4.352e-2	7.053e+5	0	4.172e-2
⊞ AYVEFQDYITR	★	1.0 : 9.6e-1 : 1.0 ...	49.56	☑		3.557e-2	6.739e+5	0	
⊟ Smad2	⊘	1.0 : 8.7e-1 : 8.5...				2.427e-2	1.791e+6	0	2.110e-2
⊟ VETPVLPPVLVPR	⊘	2.0e-2 : 1.9e-2 : ...	58.44	☑					
⊞ 708.435	⊘	1.5e-1 : 1.5e-1 : ...		☑	713.439				
⊞ 713.439	⊘	8079000.0 : 766...		☑	None				
⊞ GWGAEYR	★	1.0 : 8.7e-1 : 1.1 ...	28.91	☑		2.427e-2	1.791e+6	0	2.110e-2

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	⊘					
day2_B.RAW	⊘					
day2_C.RAW	★		0.89 fmol		1.992e-2	9.303e+6
...308\day2....	★		0.96 fmol		1.934e-2	8.812e+6
...308\day2....	★		1.13 fmol		2.068e-2	8.473e+6
...308\day2....	★		1.12 fmol		2.067e-2	8.266e+6

Area
2e-2
2e-2
2e-2
2e-2
2e-2
1e-2
1e-2
1e-2
1e-2
1e-2
0e0
0e0
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
Relative intensity (max=1e5)
100
90
80
70
60
50
40
30
20
10
0
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (1.480e+5)
100
90
80
70
60
50
40
30
20
10
0
52.3
52.8
53.2
53.6
54
54.5
54.9
55.3
55.8
56.2
56.6
Area
708.4346->371.2396(6.259e+2)
708.4346->583.392(3.247e+2)
708.4346->680.4448(1.619e+3)
708.4346->777.4976(2.229e+4)
708.4346->890.5817(3.711e+3)
708.4346->989.6501(1.154e+3)
708.4346->1086.703(3.788e+3)
713.4388->381.2479(5.174e+3)
713.4388->593.4003(2.382e+3)
713.4388->690.4531(8.249e+3)
713.4388->787.5059(1.480e+5)
713.4388->900.5899(2.358e+4)
713.4388->999.6583(6.794e+3)
713.4388->1096.711(2.610e+4)
Chromatogram(...308\day2.2\day2_B.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

g

RARg



Main Workbook | Raw File Management | **Detailed Data Analysis** | Method Refine/Optimize | Biological interpretation

Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
⊞ LDPDSEIATTGVR	⊘	1.0 : 9.4e-1 : 1.1 ...	34.63	☑		5.146e-2	1.201e+6	0	4.818e-2
⊟ PIAS3	⊘	1.0 : 1.0 : 8.9e-1 ...				5.967e-2	8.721e+5	0	6.184e-2
⊞ VSSIVAPGSSLR	⊘	1.0 : 1.0 : 1.1 : 1....	33.84	☑		4.342e-2	5.037e+5	0	4.451e-2
⊞ VSELQVLLGFAGR	⊘	1.0 : 1.0 : 7.4e-1 ...	73	☑		8.188e-2	3.684e+5	0	8.441e-2
⊟ RARG	⊘	1.0 : 1.2 : 1.2 : 1....				1.687e-2	1.339e+6	0	1.980e-2
⊞ GLGQPDLPK	★	1.0 : 1.2 : 1.2 : 1....	31.39	☑		1.687e-2	1.339e+6	0	1.980e-2
⊞ LQEPLLEALR	⊘	2.0e-2 : 2.0e-2 : ...	55.31	☑					
⊟ VQLDLGLWDK	★	2.1e-2 : 2.4e-2 : ...	56.93	☑					
⊞ 593.827	★	1.5e-1 : 1.7e-1 : ...		☑	597.834				
⊞ 597.834	★	1551000.0 : 175...		☑	None				
⊟ FOSL2	⊘	1.0 : 7.4e-1 : 8.1...				6.678e-2	1.580e+6	0	4.915e-2

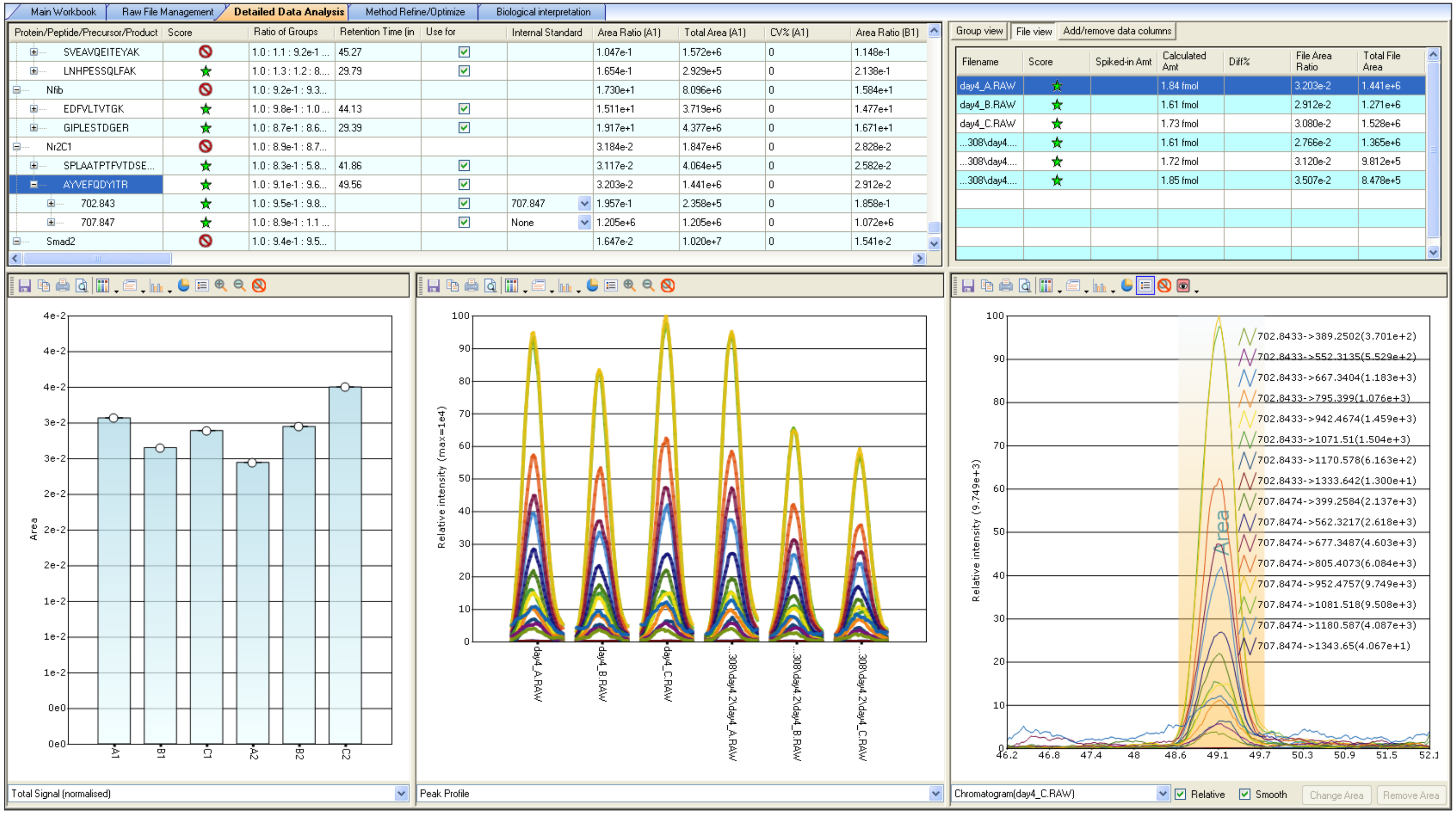
Group view | File view | Add/remove data columns

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	⊘					
day2_B.RAW	⊘					
day2_C.RAW	★		0.92 fmol		2.067e-2	1.791e+6
...308\day2....	★		1.18 fmol		2.384e-2	2.045e+6
...308\day2....	★		1.2 fmol		2.194e-2	1.354e+6
...308\day2....	★		1.26 fmol		2.316e-2	1.184e+6

Area
3e-2
3e-2
2e-2
2e-2
2e-2
2e-2
1e-2
1e-2
1e-2
1e-2
0e0
0e0
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
Relative intensity (max=2e4)
100
90
80
70
60
50
40
30
20
10
0
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (2.171e+4)
593.8269->448.2185(4.533e+2)
593.8269->561.3026(3.813e+2)
593.8269->618.324(7.181e+2)
593.8269->731.4081(7.329e+2)
593.8269->846.4351(1.167e+3)
593.8269->959.5191(3.340e+3)
593.8269->1087.578(1.770e+2)
597.834->456.2327(3.666e+3)
597.834->569.3168(1.387e+3)
597.834->626.3383(5.042e+3)
597.834->739.4223(5.130e+3)
597.834->854.4492(6.615e+3)
597.834->967.5333(2.171e+4)
597.834->1095.592(6.191e+2)
Area
54.6
55
55.5
55.9
56.3
56.7
57.2
57.6
58
58.4
58.9
Chromatogram(...308\day2.2\day2_A.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

h

NR2C1



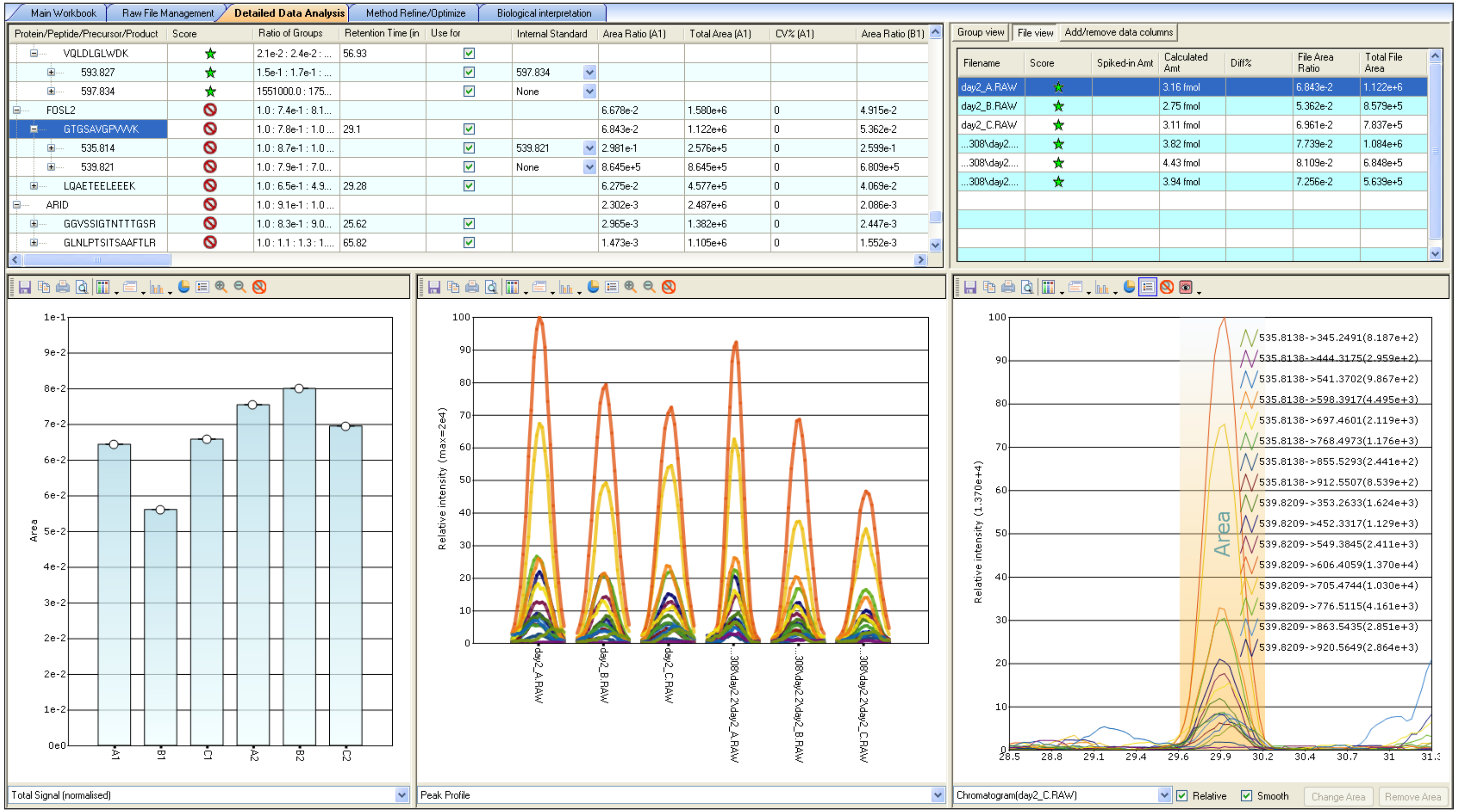
Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
SVEAVQEITEYAK	⊘	1.0 : 1.1 : 9.2e-1 ...	45.27	☑		1.047e-1	1.572e+6	0	1.148e-1
LNHPESSQLFAK	★	1.0 : 1.3 : 1.2 : 8....	29.79	☑		1.654e-1	2.929e+5	0	2.138e-1
Nfib	⊘	1.0 : 9.2e-1 : 9.3...				1.730e+1	8.096e+6	0	1.584e+1
EDFVLTVTGK	★	1.0 : 9.8e-1 : 1.0 ...	44.13	☑		1.511e+1	3.719e+6	0	1.477e+1
GIPLESTDGER	★	1.0 : 8.7e-1 : 8.6...	29.39	☑		1.917e+1	4.377e+6	0	1.671e+1
Nr2C1	⊘	1.0 : 8.9e-1 : 8.7...				3.184e-2	1.847e+6	0	2.828e-2
SPLAATPTFVTDSE...	★	1.0 : 8.3e-1 : 5.8...	41.86	☑		3.117e-2	4.064e+5	0	2.582e-2
AYVEFQDYITR	★	1.0 : 9.1e-1 : 9.6...	49.56	☑		3.203e-2	1.441e+6	0	2.912e-2
702.843	★	1.0 : 9.5e-1 : 9.8...		☑	707.847	1.957e-1	2.358e+5	0	1.858e-1
707.847	★	1.0 : 8.9e-1 : 1.1 ...		☑	None	1.205e+6	1.205e+6	0	1.072e+6
Smad2	⊘	1.0 : 9.4e-1 : 9.5...				1.647e-2	1.020e+7	0	1.541e-2

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day4_A.RAW	★		1.84 fmol		3.203e-2	1.441e+6
day4_B.RAW	★		1.61 fmol		2.912e-2	1.271e+6
day4_C.RAW	★		1.73 fmol		3.080e-2	1.528e+6
...308\day4....	★		1.61 fmol		2.766e-2	1.365e+6
...308\day4....	★		1.72 fmol		3.120e-2	9.812e+5
...308\day4....	★		1.85 fmol		3.507e-2	8.478e+5

Area
4e-2
4e-2
4e-2
3e-2
3e-2
2e-2
2e-2
2e-2
1e-2
1e-2
1e-2
0e0
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
Relative intensity (max=1e4)
100
90
80
70
60
50
40
30
20
10
0
day4_A.RAW
day4_B.RAW
day4_C.RAW
...308\day4.2\day4_A.RAW
...308\day4.2\day4_B.RAW
...308\day4.2\day4_C.RAW
Peak Profile
Relative intensity (9.749e+3)
100
90
80
70
60
50
40
30
20
10
0
46.2
46.8
47.4
48
48.6
49.1
49.7
50.3
50.9
51.5
52.1
Area
702.8433->389.2502(3.701e+2)
702.8433->552.3135(5.529e+2)
702.8433->667.3404(1.183e+3)
702.8433->795.399(1.076e+3)
702.8433->942.4674(1.459e+3)
702.8433->1071.51(1.504e+3)
702.8433->1170.578(6.163e+2)
702.8433->1333.642(1.300e+1)
707.8474->399.2584(2.137e+3)
707.8474->562.3217(2.618e+3)
707.8474->677.3487(4.603e+3)
707.8474->805.4073(6.084e+3)
707.8474->952.4757(9.749e+3)
707.8474->1081.518(9.508e+3)
707.8474->1180.587(4.087e+3)
707.8474->1343.65(4.067e+1)
Chromatogram(day4_C.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

i

FOSL2



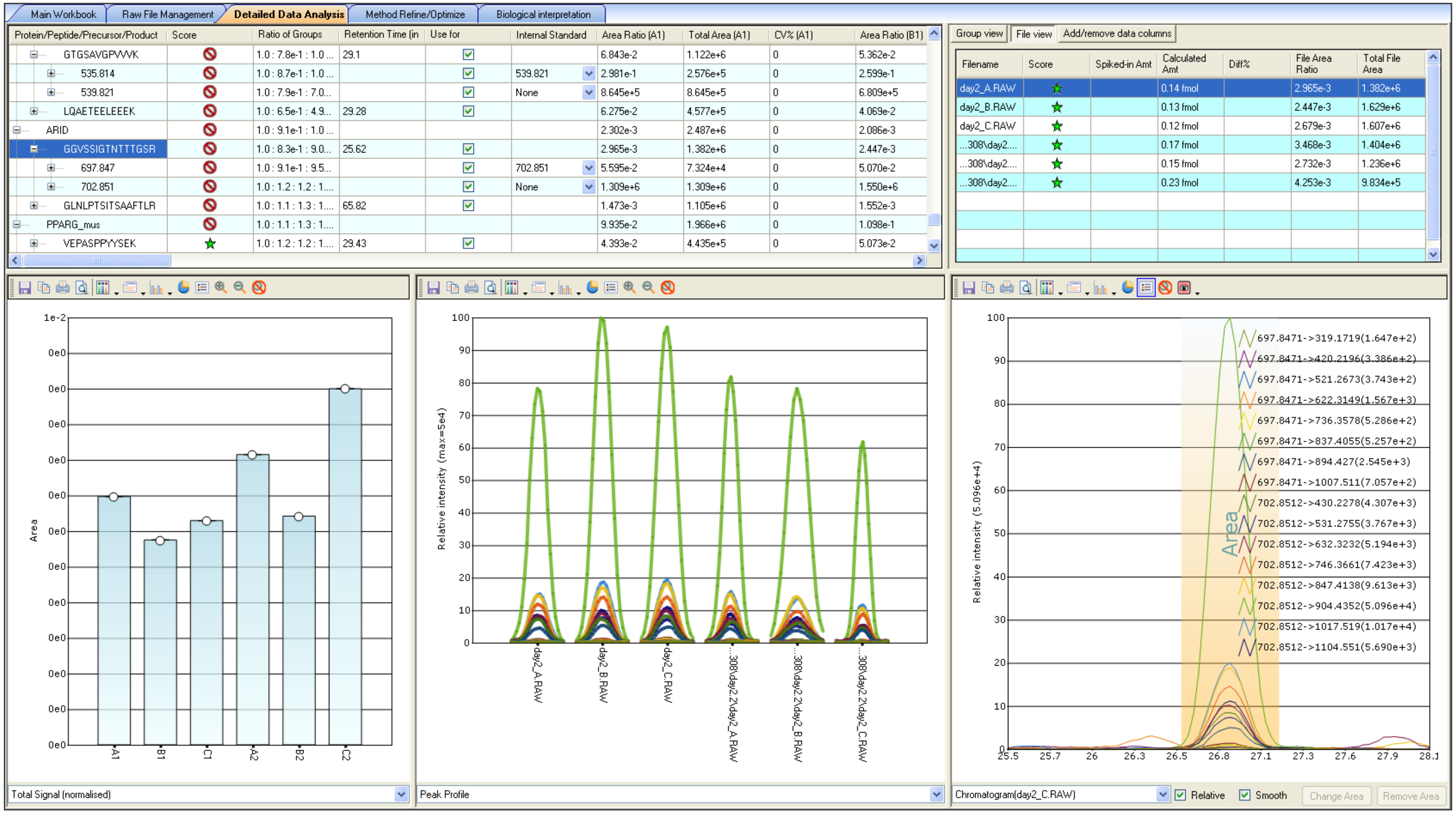
Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
VQLDLGLWDK	★	2.1e-2 : 2.4e-2 : ...	56.93	☑					
593.827	★	1.5e-1 : 1.7e-1 : ...		☑	597.834				
597.834	★	1551000.0 : 175...		☑	None				
FOSL2	🚫	1.0 : 7.4e-1 : 8.1...				6.678e-2	1.580e+6	0	4.915e-2
GTGSAVGPVVVK	🚫	1.0 : 7.8e-1 : 1.0 ...	29.1	☑		6.843e-2	1.122e+6	0	5.362e-2
535.814	🚫	1.0 : 8.7e-1 : 1.0 ...		☑	539.821	2.981e-1	2.576e+5	0	2.599e-1
539.821	🚫	1.0 : 7.9e-1 : 7.0...		☑	None	8.645e+5	8.645e+5	0	6.809e+5
LQAETEELEEEK	🚫	1.0 : 6.5e-1 : 4.9...	29.28	☑		6.275e-2	4.577e+5	0	4.069e-2
ARID	🚫	1.0 : 9.1e-1 : 1.0 ...				2.302e-3	2.487e+6	0	2.086e-3
GGVSSIGTNTTTGSR	🚫	1.0 : 8.3e-1 : 9.0...	25.62	☑		2.965e-3	1.382e+6	0	2.447e-3
GLNLPTSITSAAFTLR	🚫	1.0 : 1.1 : 1.3 : 1....	65.82	☑		1.473e-3	1.105e+6	0	1.552e-3

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	★		3.16 fmol		6.843e-2	1.122e+6
day2_B.RAW	★		2.75 fmol		5.362e-2	8.579e+5
day2_C.RAW	★		3.11 fmol		6.961e-2	7.837e+5
...308\day2....	★		3.82 fmol		7.739e-2	1.084e+6
...308\day2....	★		4.43 fmol		8.109e-2	6.848e+5
...308\day2....	★		3.94 fmol		7.256e-2	5.639e+5

Area
1e-1
9e-2
8e-2
7e-2
6e-2
6e-2
5e-2
4e-2
3e-2
2e-2
2e-2
1e-2
0e0
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
Relative intensity (max=2e4)
100
90
80
70
60
50
40
30
20
10
0
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (1.370e+4)
100
90
80
70
60
50
40
30
20
10
0
28.5
28.8
29.1
29.4
29.6
29.9
30.2
30.4
30.7
31
31.3
Area
535.8138->345.2491(8.187e+2)
535.8138->444.3175(2.959e+2)
535.8138->541.3702(9.867e+2)
535.8138->598.3917(4.495e+3)
535.8138->697.4601(2.119e+3)
535.8138->768.4973(1.176e+3)
535.8138->855.5293(2.441e+2)
535.8138->912.5507(8.539e+2)
539.8209->353.2633(1.624e+3)
539.8209->452.3317(1.129e+3)
539.8209->549.3845(2.411e+3)
539.8209->606.4059(1.370e+4)
539.8209->705.4744(1.030e+4)
539.8209->776.5115(4.161e+3)
539.8209->863.5435(2.851e+3)
539.8209->920.5649(2.864e+3)
Chromatogram(day2_C.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

j

ARID3a



Protein/Peptide/Precursor/Product	Score	Ratio of Groups	Retention Time (in	Use for	Internal Standard	Area Ratio (A1)	Total Area (A1)	CV% (A1)	Area Ratio (B1)
GTGSAVGPVVVK	⊘	1.0 : 7.8e-1 : 1.0 ...	29.1	☑		6.843e-2	1.122e+6	0	5.362e-2
535.814	⊘	1.0 : 8.7e-1 : 1.0 ...		☑	539.821	2.981e-1	2.576e+5	0	2.599e-1
539.821	⊘	1.0 : 7.9e-1 : 7.0...		☑	None	8.645e+5	8.645e+5	0	6.809e+5
LQAETEELEEEK	⊘	1.0 : 6.5e-1 : 4.9...	29.28	☑		6.275e-2	4.577e+5	0	4.069e-2
ARID	⊘	1.0 : 9.1e-1 : 1.0 ...				2.302e-3	2.487e+6	0	2.086e-3
GGVSSIGTNTTTGSR	⊘	1.0 : 8.3e-1 : 9.0...	25.62	☑		2.965e-3	1.382e+6	0	2.447e-3
697.847	⊘	1.0 : 9.1e-1 : 9.5...		☑	702.851	5.595e-2	7.324e+4	0	5.070e-2
702.851	⊘	1.0 : 1.2 : 1.2 : 1....		☑	None	1.309e+6	1.309e+6	0	1.550e+6
GLNLPTSITSAAFTLR	⊘	1.0 : 1.1 : 1.3 : 1....	65.82	☑		1.473e-3	1.105e+6	0	1.552e-3
PPARG_mus	⊘	1.0 : 1.1 : 1.3 : 1....				9.935e-2	1.966e+6	0	1.098e-1
VEPASPPYYSEK	★	1.0 : 1.2 : 1.2 : 1....	29.43	☑		4.393e-2	4.435e+5	0	5.073e-2

Filename	Score	Spiked-in Amt	Calculated Amt	Diff%	File Area Ratio	Total File Area
day2_A.RAW	★		0.14 fmol		2.965e-3	1.382e+6
day2_B.RAW	★		0.13 fmol		2.447e-3	1.629e+6
day2_C.RAW	★		0.12 fmol		2.679e-3	1.607e+6
...308\day2....	★		0.17 fmol		3.468e-3	1.404e+6
...308\day2....	★		0.15 fmol		2.732e-3	1.236e+6
...308\day2....	★		0.23 fmol		4.253e-3	9.834e+5

1e-2
0e0
0e0
0e0
0e0
0e0
0e0
0e0
0e0
0e0
0e0
0e0
0e0
Area
A1
B1
C1
A2
B2
C2
Total Signal (normalised)
100
90
80
70
60
50
40
30
20
10
0
Relative intensity (max=5e4)
day2_A.RAW
day2_B.RAW
day2_C.RAW
...308\day2.2\day2_A.RAW
...308\day2.2\day2_B.RAW
...308\day2.2\day2_C.RAW
Peak Profile
Relative intensity (5.096e+4)
Area
697.8471->319.1719(1.647e+2)
697.8471->420.2196(3.386e+2)
697.8471->521.2673(3.743e+2)
697.8471->622.3149(1.567e+3)
697.8471->736.3578(5.286e+2)
697.8471->837.4055(5.257e+2)
697.8471->894.427(2.545e+3)
697.8471->1007.511(7.057e+2)
702.8512->430.2278(4.307e+3)
702.8512->531.2755(3.767e+3)
702.8512->632.3232(5.194e+3)
702.8512->746.3661(7.423e+3)
702.8512->847.4138(9.613e+3)
702.8512->904.4352(5.096e+4)
702.8512->1017.519(1.017e+4)
702.8512->1104.551(5.690e+3)
25.5
25.7
26
26.3
26.5
26.8
27.1
27.3
27.6
27.9
28.1
Chromatogram(day2_C.RAW)
☑ Relative
☑ Smooth
Change Area
Remove Area

Supplementary Figure 12. (a) – (j) Screen prints of all peptides monitored by SRM for each of the 10 TFs at day two of differentiation.
The Pinpoint window provides a summary of the evaluation of the peptides of interest of the 10 TFs. Subpanels 1 and 2 show the calculated amount of peptide found for each technical replicate. Subpanel 3 shows the total normalized signal of each individual technical replicate. Subpanel 4 shows peak profiles of each technical replicate. Subpanel 5 shows the calculated peak area and chromatogram of peptide transitions. Panel legend and numbering applies to all figures.

Supplementary Methods

3T3-L1 protein extraction

3T3-L1 cells were collected at 6 different differentiation time points (0h, 2h, 24h or Day 1, Day 2, Day 4, and Day 6). At each time-point, petri-plates were rinsed twice with 1X PBS, after which cells were trypsinized, rinsed once with cold 1X PBS, and centrifuged and the pellets were then stored at -80°C. Cells were lysed in cold lysis buffer (10mM HEPES-NaOH at pH 7.9, 1.5mM $MgCl_2$, 10mM KCl, 1mM DTT, 0.5% NP-40, 0.1mM EDTA, 0.1mM EGTA) containing protease inhibitors and phosphates inhibitors (Roche) for 15 min and centrifuged for 10 min at 6,000 rpm (at 4°C) to sediment the nuclei. The pellet was washed twice to remove non-nuclear particles. The cytosolic fraction was collected and stored at -80°C. The isolated nuclei were then washed using protein extraction buffer (20mM HEPES-NaOH at pH 7.9, 25% glycerol, 1.5mM $MgCl_2$, 420mM NaCl, 0.1mM EDTA, 0.1mM EGTA) after adding protease inhibitor and phosphatase inhibitor cocktail tablets (Roche) at 4°C for 30 min and centrifuged for 10 min at 6,000 rpm (at 4°C). Protein concentration was measured for each time point utilizing the Quick Start Bradford Dye Reagent (Bio-Rad). The supernatant containing nuclear proteins was collected and stored at -80 °C.

In vitro protein expression and purification, gel staining, and immunoblotting

*For selection of optimal TF transitions: t*o purify *in vitro*-expressed TFs, proteins were mixed with glutathione sepharose 4B beads (GE Healthcare) and incubated overnight at 4°C on a rotator to enrich for the GST-fusion proteins. The beads were washed 3 times with a saline buffer (50 mM tris-HCl pH 8.0, 150 mM NaCl, 2mM EDTA, 0.1% NP-40, 10% glycerol) and boiled at 95°C for 5 min in protein loading buffer (1-3x depending on expression yield) to release the fusion proteins. The protein-bead mixture was centrifuged at 6’000 rpm for 30 s to sediment the beads. The supernatant fraction was separated by SDS-PAGE in a 10% resolving gel (Tris-Glycine) and stained either with silver nitrate or with Coomassie blue (SimplyBlue Safestain, Invitrogen) for band visualization. Validation of *in vitro*-expressed proteins was performed by Western blot using rabbit anti-GST primary antibodies (Cell signalling) and goat anti-rabbit coupled to HRP secondary antibodies (Santa Cruz) on a nitrocellulose membrane. Validation of PPARγ and RXRα expression was performed again by Western blot using rabbit anti-PPARγ (Santa Cruz, SC-7196) and rabbit anti-RXRα (Santa Cruz, SC-553) primary antibodies, using PARP-1 as a nuclear control (Santa Cruz, SC-1561). Densitometric quantitation analyses were performed using the AlphaDigiDoc 1201 software (Alpha Innotech).

In vitro-expressed TF identification

Samples were prepared for liquid chromatography/tandem mass spectrometry (LC-MS-MS) analysis as indicated in the **Online Methods** section. A mass spectrometer (LTQ Orbitrap XL (Thermo Fisher Scientific) equipped with an ultraperformance LC (UPLC) system (nanoACQUITY; Waters) was used. Peptides were trapped in a custom-made precolumn (Magic C18 AQ stationary phase, 5 µm diameter, 200Å pore, 0.1 × 20 mm, Michrom Bioresource) and separated in a custom-made main column (Magic C18 AQ, 3 µm diameter, 100Å pore, 0.75 × 150 mm), using a run of 53 minutes and a gradient of H2O:ACN:FA 98%:2%:0.1% (solvent A) and ACN:H2O:FA 98%:1.9%:0.1% (solvent B). The gradient of the run was set at a flow rate of 250 nl/min as follows: 100% A for 3 min, 30% B within 36 min, 47% B within 14 min, 90% B within 5 min and held for 5 min and 100% A for 17 min. The MS/MS was operated in an information-dependent mode, in which each full MS analysis was followed by 10 MS/MS acquisitions, during which the most abundant peptides were selected for collision-induced dissociation (CID) to generate tandem mass spectra. The normalized collision energies were set to 35% for CID. Data search was performed using Mascot software (v. 2.3; Matrix Science) and Proteome Discoverer (v.1.3; Thermo Fisher Scientific). The sequences were searched against a concatenated database consisting of a *Triticum aestivum* (wheat) database created from Uniprot database version 12.02 (4,684 sequences), the sequences of the TF constructs expressed complemented with a set of common contaminant proteins sequences. Finally, the results were imported into Scaffold (v. 3.3; Proteome Software) for validation of protein identification, normalization, and comparison of spectral counts. Peptide identifications were accepted if they could be established at a probability of greater than 95% as determined by the PeptideProphet algorithm[1]. Protein identifications were accepted if they were assigned at least two unique validated peptides, and could be established with at least 99% probability as determined by the ProteinProphet algorithm[2].

Proteotypic peptides selection

Peptide proteotypicity: candidate proteotypic peptides were selected using Pinpoint (v.1.1; Thermo Fischer Scientific) and their fragmentation spectra visually checked in Skyline[3] using the spectral library inspection function. The following rules were applied in Pinpoint and

Skyline to the tryptic peptides detected to ensure that only high quality peptides were retained for the analysis:

Peptide uniqueness: As a general rule, the TF-specific peptides selected here were uniquely representative of the TF of interest. In some instances however, empirically observed best-flying candidates were shared among two or more TF isoforms or TF subfamily members and were included in the assay. Discarding such candidates for less-than-optimal ones could compromise detectability. Therefore, the precision of absolute measurements accomplished utilizing such non-unique TF-specific peptides may be affected to a certain extent by TF isoform(s) or/and subfamily counterpart(s). The following non-unique peptides were utilized in the multiplexed assay:

ARID3a: peptide GLNLPTSITSAAFTLR is shared with ARID3b; NFIB: peptide EDFVLTVTGK is shared with NFIA, and GIPLESTDGER is shared with NFIA and NFIC; PIAS3: peptide VSELQVLLGFARG is shared with PIAS2; RXRα: peptide VLTELVSK is shared with RXRβ and ILEAELAVEPK is shared with RXRγ; SMAD2: peptide VETPVLPPVLVPR is shared with SMAD3.

Peptide detectability: only peptides spanning 7-25 amino-acids and falling within a mass range of 700 to 3,000 Da were used. N- or C-terminal peptides which could be degraded or modified in the protein of interest were excluded. The sequences immediately flanking the N- and C-terminal peptides of the target sequence were also excluded from the evaluation (the sequence linking the SH-Quant tag to the N-terminus and the sequence linking the C-terminal peptide and the GST-tag sequence).

Peptide digestability: as the whole methodology is peptide-centric and thus dependent on the completeness and specificity of the digestion step, special care was taken to ensure that only end-product peptides would be selected for the assay. This was especially true for the SH-Quant tag construction which contains a [KK] dibasic sequence at its C-terminal part which could lead to miscleavages. The SH-Quant tag configuration was however retained as systematic attempts to detect miscleavage products originating from the [KK] dibasic sequence using either accurate mass inclusion mass spectrometry or targeted SRM detection were negative (i.e. only the SH-Quant tag end-product was detected). For all the other peptides, only fully tryptic peptides with no miscleavage other than [KP] and [RP] were considered. N- or C- terminal dibasic residues ([KK], [KR], [RK], [RR]) as well as surrounding acidic residues (E, D) in position P’1 and P’2 potentially leading to miscleavages were excluded.

Peptide stability: peptide sequences containing oxidation-sensitive residues, methionine only in this case, were excluded. These peptides which usually ionize well were however kept as secondary choices in case no usable sequences would remain after the selection[4]. Peptides containing potential imide forming residues, [NG], potential deamidating residue, [DP], or N-terminal glutamine (Q) were excluded as well.

Peptide selection: A first manual ranking of the proteotypic peptides was performed using peptide intensities extracted from the MS1 stage with the intention of selecting the best "flyers". This ranking was complemented by a manual evaluation of the tandem MS spectral quality aimed at not excluding low intensity MS1 peaks with intense selective fragment ions (TIC MS2 in **Supplementary Table 2**). Specifically, tandem MS spectra presenting high intensity singly charged y'' ions above the m/z of the parent ion were included in the second screen performed on the triple quadrupole instrument (TSQ Vantage, Thermo Fisher Scientific). In a preliminary SRM run, all selected peptides were targeted. The final best-responding candidates were selected by performing a pilot SRM analysis targeting the heavy labeled versions of the peptides selected in the previous screen spiked in a nuclear extract digest prepared in the same way as for the real experiment. A minimum of two peptides per TF was finally selected for the multiplex assay.

Evaluation of heavy TF isotope incorporation and miscleavage during SRM

All *in vitro*-expressed heavy TFs were assessed for light isotope (K & R) misincorporation prior to SRM quantification. The amounts of isotope misincorporation (usually ≤ 5%) were corrected in the final calculations of absolute endogenous concentrations. In addition, all SH-Quant tags were monitored for trypsin miscleavage during a dedicated LC-MS run targeting the SH-Quant tag and its potential miscleavage products to assess the extent of digestion completeness.

Supplementary references

49. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical chemistry* **74**, 5383-5392 (2002).

50. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical chemistry* **75**, 4646-4658 (2003).

51. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010).

52. Huillet, C. et al. Accurate quantification of cardiovascular biomarkers in serum using Protein Standard Absolute Quantification (PSAQ) and selected reaction monitoring. *Molecular & cellular proteomics : MCP* **11**, M111 008235 (2012).

II.III DECODING THE ADIPOGENIC GENE REGULATORY NETWORK USING QUANTITATIVE TARGETED PROTEOMICS

Future work:

Decoding the adipogenic gene regulatory network using quantitative targeted proteomics

INTRODUCTION

Adipogenesis is probably one of the best characterized cell differentiation processes in biology as it has been subject to great attention primarily because of the vital role that the adipose tissue has in energy homeostasis and because excessive fat storage leading to obesity can be directly linked to a plethora of diseases (E.Calle and R.Kaaks, Nature Reviews Cancer, 2004). This extensive knowledge is in large part due to the availability of immortalized cell lines that faithfully mimic what happens *in vivo* (Rosen *et al.*, Genes dev., 2000). As a result, key players involved in the adipogenic transcriptional network have been identified (Figure 1) although this has so far only yielded a rather qualitative picture of the mechanisms at play.

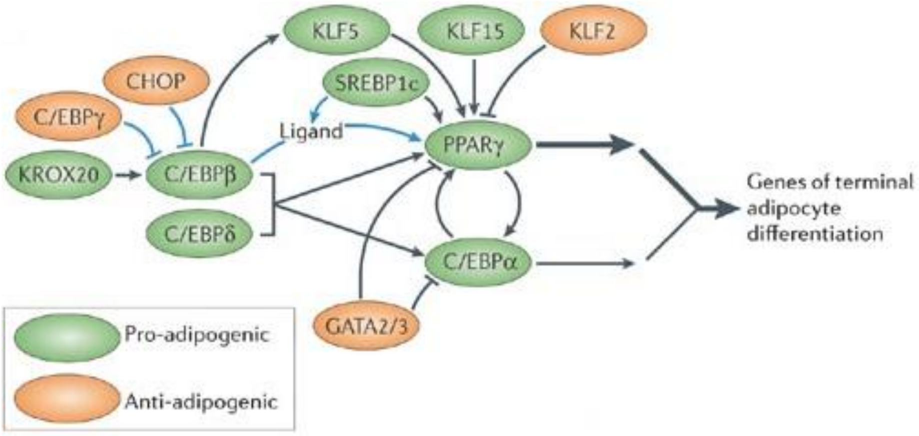


Figure 1: Transcriptional cascade regulating adipogenesis. (Rosen and MacDougald, Nature Reviews Molecular Cell Biology, 2006)

AIMS OF THE PROJECT: QUANTIFICATION OF ADIPOGENIC KEY PLAYERS VIA TARGETED PROTEOMICS

Building upon the development of our targeted quantitative approach aimed at quantifying lowly abundant TFs, we decided to expand the analytical capabilities of our methodology by integrating into the workflow a series of consecutive SRM runs (each able of measuring the abundance of 10 TFs at the time maximum), with the aim of increasing the number of proteins that can be scrutinized in one study. As often biological processes are rather complex, involving the interplay of many proteins (e.g.

signaling pathways), it would be extremely useful to devise a quantitative approach that can be adapted to cover a larger number of participant (a prerequisite in the majority of systems biology investigations). In our case, since we are interested in the study of adipogenesis, it would be of utmost importance to be able to measure fluctuations in the abundance of the many TFs involved in this process of cell differentiation. To this end, we aim to quantify in absolute amounts the TFs that lie at the core of the adipogenic gene regulatory network (12 proteins, therefore 2 SRM assays) (the complete list is presented in Table 1) and possibly increase the number of selected proteins in a second phase. Since the regulatory capabilities of TFs are largely dictated by their cellular abundance, the quantitative data obtained will be of utmost importance to understand the gene regulatory mechanisms at play. In order to better understand the fine-tuning mechanisms that take place in the nucleus upon changes in the external or internal environment, comparative quantitative studies can be envisaged. In particular, we aim at perturbing the regulatory network by knocking down an important adipogenic co-repressor such as SMRT (Raghav *et al.*, Molecular Cell, 2012) and explore the concurrent changes in adipogenic TF copy number. This will allow us to study the dynamic characteristics of the underlying regulatory machinery during terminal differentiation of 3T3-L1 cells by comparing knock-down versus wild-type. The copy per cell counts obtained will be used to develop a computational model that will elucidate or predict the dynamic properties of gene regulatory mechanisms underlying adipogenesis. Ultimately, the same methodology could be utilized to investigate in a comprehensive manner gene regulatory networks involved in other cellular processes of choice, including those involving a large number of players.

TF	**SwisssProt ID**	**TF**	**Swissprot ID**
Cebpα	P53566	Klf5	Q9Z0Z7
Cebpβ	P28033	Klf15	Q9EPW2
Cebpδ	Q00322	Krox20	P08152
Cebpγ	P53568	Pparγ	P37238
Gata3	P23772	Srebp1	Q9WTN3
Klf2	Q60843	Chop	P35639

Table 1: List of transcription factors involved in the core adipogenic gene regulatory network.

III. CONCLUSIONS

Regulation of gene expression, which is orchestrated in cells by gene regulatory networks, is the result of a complex interplay between genes and their DNA-binding partners, transcription factors. Due to their vital role in most cellular processes, understanding the mechanisms by which transcription factors regulate gene expression is of crucial importance in biology. Thanks to their particularly high sensitivity and specificity, MS-based proteomics is quickly becoming the *gold standard* in the characterization and quantification of DNA binding TFs.

Originally, we aimed to develop a gene-centered proteomic platform to capture TFs and co-TFs in the form of functional complexes in a systematic manner utilizing a novel technique called PICh (Proteomics of Isolated Chromatin segments), targeting β-globin-like gene regulatory elements in an immortalized human myelogenous leukaemia cell line. Although extensive work has been accomplished in characterizing the key players (in terms of TFs) involved in many biological processes, including the one controlling β-globin gene switching (depicting a rather comprehensive qualitative picture of the major GRNs), almost no quantitative information was available in the literature. Thus, while our efforts were initially calibrated towards dissecting TF-complex formation to elucidate novel mechanisms of gene regulation, we became soon aware of the importance of accurately quantifying DNA-binding proteins. Moreover, deriving TF copy numbers has been of longstanding interest in regulatory biology since the DNA binding ability and thus the regulatory capabilities of these proteins strongly depends on their cellular concentration[5-7]. To date, only a handful of studies have so far provided estimates on the absolute *in vivo* abundance of animal TFs. These measurements were achieved with indirect, immuno-based methods though (which thanks to signal amplification are still considered to be the gold standard in the quantification of low-abundant proteins), whose additional drawbacks include the dependence on antibodies which are available for only a limited number of TFs. Our intent was to propose a viable option to such alternative methodologies, since they ultimately prove to be too time-consuming and impractical at large scale.

We focused on Selected Reaction Monitoring (SRM), which appeared to be particularly well suited for targeting low abundant proteins such as TFs, which are difficult to identify and considerably more difficult to quantify with standard shotgun LC-MS-based approaches. Together with our difficulties to implement the PICh approach, which seemed to be shared and acknowledged by many colleagues (thus raising doubts about the reproducible nature of the PICh method), we therefore decided to put the original project on hold and focus instead fully on the problem of TF quantification. Specifically, we aimed to establish a novel analytical platform that focuses on the use of spectral libraries of TF peptide libraries, based on *in vitro* protein expression proteomic parameter selection, for the high-throughput development of SRM assays in an adipogenic model. In particular, we intended to exploit the TF proteotypic peptide library that we had already started populating with the previous project as to obtain the parameters needed for the implementation of such assays (e.g. elution time and charge state of the peptides). With parameter selection being a crucial aspect, a variety of predictive tools had been appositely developed. Unfortunately, predictions tend to fail to fully capture the peculiar behavior of peptides in a mass spectrometer, therefore have to be utilized with care. We found out that by utilizing experimental data we obtain better peptide candidates, improving thereby the assay as a whole (choosing inadequate peptides may prevent quantification). In this regard, starting from our vast murine TF ORF collection, we were able to create a peptide Atlas that contains more than 1000 TF peptide spectra, representing some 800 unique peptides belonging to approximately 100 TFs. This vast effort was the first of its kind; we aimed at creating a comprehensive TF-specific spectral repository to make available to the public. TF-specific data can be utilized as a guide for future SRM assays or to improve protein identification. Although a significant effort has been devoted since the beginning of our project to address TF-specific peptide information paucity, reliable TF peptide data is still lacking in the majority of databases.

Hence, capitalizing on the TF-specific data acquired, we developed a state-of-the-art, targeted SRM-based assay, which combines high sensitivity and technological innovation to enable the monitoring of absolute copy number changes of TFs of interest during specific biological processes using *in vitro*-expressed, isotopically-labelled protein standards. Our pipeline significantly extends recent efforts to specifically tailor MS methodologies to quantify low-abundant proteins such as TFs, by offering several advantages. First, the use of full-length protein standards allow us to bypass the use of expensive, accurately quantified synthetic peptide standards. Second, current absolute protein quantification methodologies relying on full-length protein internal standards currently require that the latter are quantified in a separate experiment, thus limiting the measurements to one standard at a time. We significantly simplify the internal protein standard quantification step, which, in our assay, is performed *in situ* along with the quantification of its endogenous counterpart (two quantification steps are collapsed into one unique assay). This in turn circumvents the need for intermediate protein storage as well as separate quantification procedures, effectively eliminating two significant sources of measurement variation. Third, the use of a proteotypic-peptide tag previously employed for quantitative interaction studies, the SH-Quant tag, allowed us to design and validate 9 additional quantotypic peptide-tag variants, simply by changing at the most by one or two amino acids within the original tag sequence at selected positions. We used the tag variants to develop a multiplex SRM-based assay that allows the *in situ* quantification of up to 10 proteins in one single assay. Importantly, we were able to demonstrate that the upscale features the same overall accuracy and sensitivity compared to the single protein analysis. In conclusion, the multiplexing of the assay provides a sensitive, cost-effective, and rapid strategy to simultaneously quantify small sets of proteins in one unique integrated assay. The multiplexing capability constitutes an important step towards addressing the need for assay upscaling, which remains one of the main limitations in current SRM-based assays. Moreover, our quantification methodology therefore constitutes a powerful and even cost-effective alternative to quantitative immunoassays to determine fluctuations in protein abundance by presenting the additional advantage of being amicable to multiplexing. The expanded analytical capabilities offered by the multiplexed workflow should be of great value in the modeling of absolute, dynamic changes of entire pathways as well as in the field of biomarker discovery.

Once the workflow had been established and extensively tested, we applied it to uniquely assess the *in vivo* expression changes of the adipogenic master regulators PPARγ and RXRα as well as eight other TFs during the terminal phase of adipogenesis. We found that the absolute abundance of TFs differs quite substantially, but that their dynamic range during fat cell differentiation is limited, varying at most five-fold. Finally, we illustrated the importance of deriving this type of data in building a quantitative model of genome-wide TF DNA binding by focusing on the adipogenic master regulator PPARγ. Specifically, we sought to mechanistically explain the paradoxical finding of a significantly increased number of PPARγ binding sites during the final stage of differentiation despite a concurrent saturation in the number of PPARγ molecules. The ensuing model provided unique, quantitative insights into TF DNA binding. Whereas it had previously been suggested that other factors may contribute to binding site selectivity, we revealed that the DNA binding profile of PPARγ can be faithfully modeled by considering its own copy number, simple thermodynamic principles, and chromatin accessibility, thus emphasizing the importance of both protein copy number and chromatin remodeling in dictating TF DNA binding behavior during adipogenesis.

To conclude, the work accomplished during my PhD thesis, the development of the SRM-based methodology for the quantification of low-abundant protein species in particular, represents an important effort to expand the analytical and practical capabilities of targeted proteomics approaches while preserving measurement accuracy. Possible improvements would include the further expanding the number of reference-peptide variants to increase the assay’s upscale capacity. The resulting increase in throughput capabilities will allow the elucidation and dynamic modeling of entire biological processes or networks under different experimental conditions.

Curriculum Vitae

Address :	Rue de la Blancherie 17, 1022 Chavannes-près-Renens, Switzerland
Phone:	+41/78/616-20-24, jovan.simicevic@epfl.ch, jovansimicevic@yahoo.com
Nationality:	Swiss
Date of birth:	20.09.1976

EDUCATION

Ph.D. IN BIOENGINEERING (05/2008-08/2012)

Thesis: Dissecting gene regulatory networks with targeted quantitative proteomics
Swiss Institute of Technology at Lausanne (EPFL) – Laboratory of Systems Biology and Genetics (Prof. Deplancke), Lausanne, Switzerland

MASTER'S DEGREE IN PROTEOMICS AND BIOINFORMATICS (07/2007-03/2008)

Thesis: High Throughput Proteomics Imaging using Fluorescent Peptidic Markers
University of Geneva – Department of Structural Biology and Bioinformatics, Geneva, Switzerland

CERTIFICATE IN BIOINFORMATICS – BIOLOGY TRACK (09/2005-06/2006)

UCSD, University of California San Diego, San Diego, CA , U.S.A.

MASTER OF ARTS IN APPLIED MICROECONOMICS (01/2004-01/2005)

Thesis: A Random Walk model for the Biotech Industry (Time Series analysis)
San Diego State University, San Diego, CA, U.S.A. (G.P.A. 3.83 out of 4.00)

BACHELOR OF ARTS IN ECONOMICS (10/1996-03/2001)

University of Lausanne (HEC), Lausanne, Switzerland

WORK EXPERIENCE

INTERNSHIP IN MEDICAL BIOCHEMISTRY (08/2007-02/2008)
University of Geneva – Department of Structural Biology and Bioinformatics
Laboratory of Professor D. Hochstrasser – Clinical Proteomics Group
Project: High Throughput Proteomics Imaging using Fluorescent Peptidic Markers

STATISTICIAN/ANALYST (06/2005-11/2005)
J.H.Cohn, San Diego, CA, U.S.A.
Perform multiple tests, database research, statistical analysis, forecasting.

INTERNSHIP IN DATA ANALYSIS (10/2002-12/2002)
Merrill Lynch, San Diego, CA, U.S.A
Perform database research, and account analysis, forecasting.

AUDIT ANALYST (10/2001-07/2002)
Deloitte & Touche, Lugano, Switzerland
Perform multiple tests, database research, and data analysis.

SKILLS

Microsoft Office , Windows, Mac OSX, UNIX/LINUX
Perl, R, E-views, JMP, SAS, SQL

Bioinformatics (including MS tools such as Mascot, Pinpoint, Skyline)

DNA and protein gel electrophoresis, Immunoassays

Mass Spectrometry-based techniques (including SRM/MRM)

Genetic analysis tools (Cloning, mutagenesis)

Cell culture (including murine 3T3-L1 and human HEK293, K562 cell lines)

LANGUAGES

Fluent: English, French, Italian, Serbian. Good knowledge: German, Spanish.

PERSONAL INTERESTS

Surfing, Waterpolo, Swimming, Golf, History, Anthropology, Economics, Political Science.

Imaging mass spectrometry using peptide isoelectric focusing

Ali R. Vaezzadeh, Jovan Simicevic, Alexis Chauvet, Patrice François, Catherine G. Zimmermann-Ivol, Pierre Lescuyer, Jacques P. M. Deshusses, Denis F. Hochstrasser, Rapid Communications in Mass Spectrometry, Volume 22 Issue 17 (2008), Pages 2667 – 2676

DNA-centered approaches to characterize regulatory protein-DNA interaction complexes Molecular Biosystems, 2010

J. Simicevic, B. Deplancke, Molecular Biosystems, 2010, 6, 462-468.

Decoding Absolute copy number analysis of transcription factors during cellular differentiation using a multiplex, targeted proteomics approach (submitted)

Jovan Simicevic, Adrian W. Schmid, Benjamin Zoller, Sunil K. Raghav ,Irina Krier, Carine Gubelmann, Frédérique Lisacek, Felix Naef, Marc Moniatte, Bart Deplancke, submitted

- “Swiss Proteomics Society 2007 scientific meeting”, Lausanne (Switzerland):

- Pushing the limits of the Shotgun IPG-IEF workflow

- “3rd All-SystemsX.ch-Day 2009”, Bern (Switzerland)

- Development of a gene-centered proteomic platform for the systematic identification of DNA binding proteins and complexes

- “4th EUPA Meeting 2010”, Estoril (Portugal) + oral presentation:

- Novel proteomic approach for the absolute quantification of transcription factors controlling adipogenesis

- “4th All-SystemsX.ch-Day 2010”, Geneva (Switzerland) + oral presentation:

- Novel proteomic approach for the absolute quantification of transcription factors controlling adipogenesis

- “SPS PhD Student Symposium 2010”, Basel (Switzerland) + oral presentation:

- Novel proteomic approach for the absolute quantification of transcription factors controlling adipogenesis

- “HUPO 10th World Congress 2011”, Geneva (Switzerland):

- Absolute quantification of a large set of transcription factors involved in the adipogenic gene regulatory network via a mass spectrometry-based approach

ANNEX

Supplementary Table 1: list of the 10 selected TFs

Gene symbol	Name	DNA-binding domain	IPI
RXRa	Retinoic acid receptor RXR-alpha	zf-C4	P28700
Nfib	Nuclear factor 1 B-type	MH1	P97863
Pias3	E3 SUMO-protein ligase PIAS3	SAP	O54714
Pias4	E3 SUMO-protein ligase PIAS4	SAP	Q9JM05
Fosl2	Fos-related antigen 2	bZIP_1	P47930
Rarg	Retinoic acid receptor gamma	zf-C4	P18911
PPARg	Peroxisome proliferator-activated receptor gamma	zf-C4	P37238
Arid3a	AT-rich interactive domain-containing protein 3A	ARID	Q62431
Nr2c1	Nuclear receptor subfamily 2 group C member 1	zf-C4	Q0VGP8
Smad2	Mothers against decapentaplegic homolog 2	MH1	Q62432

Supplementary Table 2: proteotypic peptide selection for the 10 TFs (multiplex)

PeptideAtlas query done on 14Apr2012

Biosequence Name	Protein Name/Accession.
Peptide Accession	Peptide Atlas accession number, beginning with PAp followed by 9 digits.
Pre AA	Preceding (towards the N terminus) amino acid
Sequence	Amino acid sequence of detected pepide, including any mass modifications.
Fol AA	Following (towards the C terminus) amino acid
Peptide Length	Length of peptide
Combined Predictor Score	Score genereated based on STEPP, PSieve, ESPP, APEX and DPred scores
PSieve	Predicted peptide score calculated by Peptide Sieve algorithm
ESPP	Predicted peptide score calculated by ESPP algorithm
DPred	Predicted peptide score calculated by Detectability Predictor algorithm
APEX	Predicted peptide score calculated by APEX algorithm
STEPP	Predicted peptide score calculated by STEPP algorithm
N SP Mapping	Number of SwissProt primary protein mapping
N SP-varsplic Mapping	Number of SwissProt primary and alternatively-spliced protein mapping
N SP-nsSNP Mapping	Number of SwissProt primary and alternatively-spliced protein mapping, plus nsSNP mapping,wherein all Swiss-Prot-annotated nsSNPs have been expanded out to sequence with context so that any nsSNP-containing peptides are properly mapped.
N ENSP Mapping	Number of Ensembl protein mapping
N ENSG Mapping	Number of Ensembl gene mapping
N IPI Mapping	Number of IPI protein mapping
N Human Mapping	Number of Human protein mapping, including SwissProt, IPI and Ensembl Proteins
N Mouse Mapping	Number of Mouse protein mapping, including SwissProt, IPI and Ensembl Proteins
N Yeast Mapping	Number of Yeast protein mapping, including SwissProt, SGD and Ensembl Proteins
Intensities	MS1 intensities as determined using MaxQuant on OrbitrapXL data. Bar length proportional to intensity. The longer the higher.
MQ.rank.MS1	MS1 intensity ranking. Bar length inversly proportional to ranking. The shorter the better.
TIC MS2 2+	Ranking of Total Ion Current in tandemMS spectrum as extracted from Scaffold for doubly charged ions. Bar length inversly proportional to ranking. The shorter the better.
TIC MS2 3+	Ranking of Total Ion Current in tandemMS spectrum as extracted from Scaffold for triply charged ions. Bar length inversly proportional to ranking. The shorter the better.
Highlights PeptideAtlas Predictions:	Number of protein mapped by peptide sequence between 1 and 3 were highlighted in green (Peptide sequence specificity)
Highlights PeptideAtlas Predictions:	Number of protein mapped by peptide sequence above 3 were not highlighted (Peptide sequence specificity)
Highlight Sequence	Proteotypic peptides selected after filtering

PeptideAtlas Legend

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities	MQ.MS1.Rank.1	TIC MS2 Rank 2+	TIC MS2 Rank 3+
4	Q62431	PAp00739	K	GLNLPTSITSAAFTLR	T	16	0.92	0.97	0.59	0.77	0.07	0.38	2	2	2	6	2	3	12	11			1	
8	Q62431	PAp00388	K	LPVTPLGLAASTNGSSITPAPK	I	22	0.9	0.92	0.4	0.57	0.06	0.71	1	1	1	3	1	2	0	6				1
18	Q62431	PAp00598	K	GGLVEVINK	K	9	0.3	0.31	0.73	0.43	0.05	0.65	1	1	1	3	1	2	3	6	21190000	7	2	
26	Q62431		K	YLYPYECER	R	9	0.05	0.02	0.19	0.38	0.08	-0.48	1	1	1	3	1	2	0	6	10529000	9	3	
2	Q62431		R	EGTLSSPALHGSVLEGAGHAEGDR	H	24	0.93	0.98	0.58	0.76	0.09	0.62	1	1	1	3	1	2	0	6	42674000	1		
3	Q62431	PAp00741	K	EEDSAIPITVPGR	L	13	0.93	0.98	0.92	0.58	0.06	1.02	1	1	1	2	1	1	3	4	28904000	3		
1	Q62431		R	GLSSPNELQAAIDSNR	R	16	0.93	0.94	0.9	0.77	0.08	0.71	1	1	1	3	1	2	0	6	19637000	8		
5	Q62431		R	AAAAGLGHPSSPGGSEDGPPISGDEDTAR	E	29	0.92	0.98	0.61	0.68	0.05	0.74	1	1	1	4	1	3	0	8	22595000	5		
6	Q62431	PAp00480	R	QSFGGSLFAYSPSGAHSMLPSPK	L	23	0.91	0.99	0.27	0.66	0.09	0.26	1	1	1	3	1	2	0	6				
7	Q62431		R	TGASGSTVSGGQVGLPGVSTPTMSSTSNNSLP	-	32	0.91	0.95	0.49	0.83	0.01	0.53	1	1	1	3	1	2	0	6				
9	Q62431		R	QAPPPPPPEPTGVR	A	14	0.88	0.8	0.34	0.79	0.06	0.71	1	1	1	4	1	3	0	8	21925000	6		
10	Q62431		K	GGVSSIGTNTTTGSR	T	15	0.84	0.7	0.38	0.88	0.06	0.34	1	1	1	3	1	2	0	6				
11	Q62431		K	LQAVMETLIQR	Q	11	0.82	0.77	0.42	0.83	0.02	0.22	1	1	1	4	1	3	0	8	760390	10		
12	Q62431	PAp00741	K	MALVADEQQR	L	10	0.81	0.62	0.74	0.66	0.11	0.28	1	1	1	3	1	2	3	6				
13	Q62431		R	AVQQSFLAMTAQLPMNIR	I	18	0.74	0.67	0.13	0.68	0.07	0.18	1	1	1	3	1	2	0	6				
14	Q62431	PAp00481	R	LPVSLAGHPVVAAQAAAVQAAAAQAAVAAQAAALEQLR	E	38	0.63	0.67	0.4	0.64	0.01	0.04	1	1	1	3	1	2	3	6	29589000	2		
15	Q62431		R	TQMAALAAMR	A	10	0.61	0.53	0.56	0.62	0.07	0.13	1	1	1	4	1	3	0	8				
16	Q62431		K	QLYELDADPK	R	10	0.56	0.54	0.54	0.37	0.05	0.68	2	2	2	6	2	3	3	11	24772000	4		
17	Q62431		R	HLMDVGSDDDDTK	S	13	0.38	0.42	0.3	0.6	0.01	0.64	1	1	1	3	1	2	0	6				
19	Q62431	PAp00060	K	EFLDDLFSFMQK	R	12	0.26	0.28	0.18	0.59	0.06	0.51	2	2	2	6	2	3	6	11				
20	Q62431	PAp00750	K	QVLDLFMLYVLVTEK	G	15	0.14	0.2	0.12	0.67	0.01	0.34	1	1	1	3	1	2	3	6				
21	Q62431		R	LGPGPAHPSHMASQMPPPDHGDWTFEEQFK	Q	30	0.13	0.18	0.4	0.57	0.01	0.38	1	1	1	3	1	2	0	6				
22	Q62431		K	LESTEPPEK	K	9	0.06	0.01	0.23	0.26	0.06	0.54	1	1	1	3	1	2	0	6				
23	Q62431	PAp00596	R	INSQASESR	Q	9	0.06	0.01	0.23	0.36	0.05	0.11	1	1	1	3	1	2	3	6				
24	Q62431		R	QDSAVSLTSANGSNSISMSVEMNGIVYTGVLFAQPPPPTAPSAPGK	G	46	0.05	0.11	-1	0.53	0	0.39	1	1	1	3	1	2	0	6				
25	Q62431		R	TTMTDEDR	E	8	0.05	0	0.19	0.49	0.01	0.13	1	1	1	4	1	3	0	8				
27	Q62431		K	APPAQAFR	G	8	0.05	0.04	0.46	0.23	0	0.33	1	1	1	3	1	2	0	6				
28	Q62431		R	IPIMAK	Q	6	0.04	0.01	0.4	0.13	0.02	0.34	2	2	2	6	2	3	12	11				
29	Q62431		R	GTPVNR	I	6	0.04	0.01	0.09	0.27	0.02	0.12	2	2	2	6	2	3	6	11				
30	Q62431		R	GDGGPR	M	6	0.04	0	0.18	0.18	0.03	0.14	1	1	1	3	1	2	0	6				
31	Q62431		R	MLSGPER	L	7	0.04	0.04	0.18	0.13	0.02	0.05	1	1	1	3	1	2	0	6				
32	Q62431		R	QELEAR	Q	6	0.04	0.01	0.1	0.13	0.02	-0.01	2	3	3	7	2	5	13	15				
33	Q62431		R	EPENAR	M	6	0.03	0.01	0.04	0.16	0	0.05	1	1	1	4	1	3	0	8				
34	Q62431		K	WEEQELEELGEEEEEEEEEDDFEEEEEEEEGLGPPESASLGTAGLFTR	K	48	0.02	0.22	-1	-1	0	0.69	1	1	1	3	1	2	0	6				

Proteotypic Peptides

PeptideAtlas Predictions

Arid3a_Q62431

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities.y	MQ.MS1.Rank.2	TIC MS2 Rank 2+	TIC MS2 Rank 3+
1	P47930		R	GTGSAVGPVVVK	Q	12	0.92	0.83	0.66	0.78	0.1	0.61	1	1	1	1	1	2	0	4	3430300	1	1	
5	P47930	PAp01458	K	LQAETEELEEEK	S	12	0.78	0.75	0.37	0.6	0.02	0.57	1	1	1	1	1	2	6	4	2874200	2	2	
11	P47930		R	DEQLSPEEEEK	R	11	0.06	0.07	0.21	0.42	0	0.47	1	1	1	1	1	2	9	4	1007800	3		
9	P47930		R	SPPTSGLQSLR	G	11	0.18	0.18	0.63	0.69	0.02	0.33	1	1	1	1	1	2	0	4	918780	4		
13	P47930		K	EIAELQK	E	7	0.04	0.02	0.27	0.09	0	0.39	3	4	4	8	3	6	30	18	711070	5		
2	P47930	PAp01750	R	SSSSGDQSSDSLNSPTLLAL	-	20	0.92	0.91	0.48	0.82	0.04	0.81	1	1	1	1	1	2	7	4				
3	P47930	PAp00154	-	MYQDYPGNFDTSSR	G	14	0.85	0.93	0.33	0.61	0.01	0.33	1	1	1	1	1	2	3	4				
4	P47930	PAp00068	R	GSSGSPAHAESYSSGGGGQQK	F	21	0.85	0.85	0.36	0.78	0.01	0.34	1	1	1	1	1	2	3	4				
6	P47930		R	SHPYSPLPGLASVPGHMALPRPGVIK	T	26	0.73	0.8	0.22	0.68	0	-0.05	1	1	1	1	1	2	9	4				
7	P47930		K	QEPPEEDSPSSSAGMDK	T	17	0.72	0.67	0.25	0.62	0.01	0.76	1	1	1	1	1	2	0	4				
8	P47930	PAp00583	K	LEFMLVAHGPVCK	I	13	0.41	0.53	0.19	0.5	0.06	-0.19	1	1	1	1	1	1	9	3				
10	P47930		K	TIGTTVGR	R	8	0.07	0.02	0.38	0.67	0	0.28	1	1	1	1	1	2	9	4				
12	P47930		R	VDMPGSGSAFIPTINAITTSQDLQWMVQPTVITSMSNPYPR	S	41	0.04	0	0.1	0.44	0	-0.03	1	1	1	1	1	2	6	4				
14	P47930		K	ISPEER	R	6	0.03	0.01	0.11	0.08	0.01	0.05	2	4	4	8	3	6	9	18				

Proteotypic peptides

PeptideAtlas Predictions

Fosl2_P47930

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities	MQ.MS1.Rank.1	TIC MS2 Rank 2+	TIC MS2 Rank 3+
10	P97863	PAp01440;	K	SGVFNVSELVR	V	11	0.82	0.77	0.66	0.7	0.02	0.38	1	3	3	9	1	6	14	18	38160000	1	3	
8	P97863	PAp00512(	K	NPPGYLEDSFVK	S	12	0.86	0.77	0.79	0.67	0.01	0.95	1	3	3	9	1	6	14	18	31199000	2	2	
16	P97863	PAp00518;	R	TPPPPSPLPFPTQAILPPAPSSYFSHPTIR	Y	30	0.16	0.23	0.33	0.66	0	0.37	1	3	3	8	1	5	17	16	19483000	4		1
3	P97863	PAp00779;	K	GIPLESTDGER	L	11	0.9	0.89	0.84	0.7	0.04	0.59	4	20	20	32	4	30	70	82	19467000	5	1	
11	P97863	PAp00411(	R	YPPHLNPQDTLK	N	12	0.8	0.78	0.42	0.6	0	0.73	1	3	3	8	1	5	17	16	7370200	7	4	2
12	P97863	PAp00482(	K	KPEKPLFSSTSPQDSSPR	L	18	0.8	0.9	0.32	0.54	0.04	-0.32	1	3	3	9	1	6	0	18	8142100	6		
15	P97863	PAp00482;	R	TPITQGTGVNFPIGEIPSQPYYHDMNSGVNLQR	S	33	0.52	0.58	0.19	0.63	0.01	0.26	1	3	3	9	1	5	14	17	2442100	9		
24	P97863	PAp00483;	K	HPCCVLSNPDQK	G	12	0.04	0	0.28	0.44	0.05	-0.68	1	3	3	10	1	7	17	20	1492900	10		
17	P97863	PAp00762;	R	LDLVMVILFK	G	10	0.11	0.11	0.26	0.68	0.02	0.27	4	20	20	32	4	30	70	82	1484200	11		
5	P97863		K	TISIDENMEPSPTGDFYPSPNSPAAGSR	T	28	0.89	0.96	0.51	0.66	0	0.66	1	3	3	9	1	6	0	18	841470	13		
1	P97863		R	TPILPANVQNYGLNIIGEPFLQAETSN	-	27	0.91	0.96	0.05	0.73	0.02	1.01	1	2	2	5	1	3	7	10				
2	P97863		R	EDFVLTVTGK	K	10	0.9	0.86	0.72	0.71	0.05	0.76	2	10	10	20	2	17	34	47				
4	P97863		K	NYVPSYDPSSPQTSQPNSSGQVVGK	V	25	0.89	0.98	0.45	0.65	0.01	0.51	1	1	1	4	1	3	0	8				
6	P97863		R	DPSFLHQQQLR	I	11	0.89	0.86	0.39	0.54	0.12	0.23	1	1	1	3	1	2	5	6				
7	P97863	PAp01152;	R	LSTFPQHHHPGIPGVAHSVISTR	T	23	0.88	0.97	0.23	0.77	0	0.2	1	3	3	9	1	6	17	18				
9	P97863	PAp00526;	K	ELDLFLAYYVQEQDSGQSGSPSHSDPAK	N	28	0.85	0.9	0.4	0.56	0.01	0.6	1	3	3	9	1	7	1	19				
13	P97863	PAp00778;	R	AIAYTWFNLQAR	K	12	0.62	0.68	0.22	0.67	0	0.07	1	3	3	10	1	7	16	20				
14	P97863	PAp00513;	K	DELLSEKPEIK	Q	11	0.61	0.68	0.51	0.31	0.05	0.21	1	3	3	10	1	7	17	20				
18	P97863		K	SPHCTNPALCVQPHHITVSVK	E	21	0.09	0.12	0.31	0.66	0.06	-0.56	1	3	3	9	1	7	17	19				
19	P97863		R	HLYPSTSEDTLGITWQSPGTWASLVPFQVSNR	T	32	0.07	0.1	0.09	0.61	0	-0.01	1	2	2	5	1	3	7	10				
20	P97863		R	DQDMSSPTTMK	K	11	0.06	0.01	0.25	0.56	0.01	0.38	1	3	3	9	1	6	17	18				
21	P97863		R	YVGLSPR	D	7	0.06	0.05	0.31	0.2	0.08	0.16	1	1	1	4	1	3	7	8				
22	P97863		R	ICDWTMNQNGR	H	11	0.06	0.02	0.24	0.55	0.06	-0.59	1	2	2	5	1	3	7	10				
23	P97863		R	SLSSPPSSK	R	9	0.05	0	0.21	0.3	0.01	0.34	1	3	3	9	1	6	17	18				
25	P97863		-	MMYSPICLTQDEFHPFIEALLPHVR	A	25	0.04	0.02	0.05	0.42	0	-0.41	1	3	3	5	1	5	15	13				

Proteotypic peptides

PeptideAtlas Predictions

Nfib_P97863

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities.y	MQ.MS1.Rank.2	TIC MS2 Rank 2+	TIC MS2 Rank 3+
4	Q505F1		R	SPLAATPTFVTDSETAR	S	17	0.93	0.86	0.71	0.69	0.14	0.81	1	2	2	3	1	3	0	8	30690000	1	1	
6	Q505F1		K	AYVEFQDYITR	T	11	0.92	0.9	0.24	0.69	0.14	0.3	1	1	1	3	1	2	0	6	17958000	2	2	
26	Q505F1		R	AFDTLAK	A	7	0.06	0.02	0.29	0.36	0.04	0.47	2	2	2	5	2	4	20	11	8729600	3	3	
21	Q505F1		K	GLIGNVR	I	7	0.09	0.1	0.46	0.46	0.03	0.28	1	1	1	3	1	2	0	6	4016500	4	4	
3	Q505F1		K	EGPLLSESHVAFR	L	13	0.94	0.98	0.8	0.72	0.14	0.68	1	1	1	3	1	2	0	6	3445800	5		1
22	Q505F1		K	VFDLCVVCGDK	A	11	0.08	0	0.33	0.79	0.08	-0.59	1	2	2	3	1	3	9	8	2228400	6	5	
32	Q505F1		K	NLVYSCR	G	7	0.04	0.01	0.24	0.41	0.02	-0.64	1	2	2	3	1	3	11	8	1975700	7		
1	Q505F1		K	QFILANHEGSTPGK	V	14	0.95	0.98	0.78	0.69	0.17	0.95	1	2	2	3	1	3	0	8	1013700	8		
19	Q505F1		R	IDSVIPHILK	M	10	0.44	0.3	0.43	0.45	0.14	0.56	1	1	1	3	1	2	5	6	486940	9		
23	Q505F1		R	TYPDDTYR	L	8	0.07	0.03	0.25	0.59	0.03	0.15	1	1	1	3	1	2	5	6	323440	10		
24	Q505F1		K	LQEFCNSMVK	L	10	0.07	0.05	0.25	0.46	0.09	-0.46	1	1	1	3	1	2	4	6	111760	11		
2	Q505F1		K	IQIVTALDHSTQGK	Q	14	0.94	0.98	0.67	0.78	0.13	0.82	1	2	2	3	1	3	0	8				
5	Q505F1		R	SAGLLDSGMFVNIHPSGIK	T	19	0.92	0.97	0.23	0.74	0.08	0.61	1	1	1	3	1	2	0	6				
7	Q505F1		K	AIVLFSPDHPGLENMELIER	F	20	0.9	0.92	0.05	0.67	0.04	0.99	1	1	1	3	1	2	0	6				
8	Q505F1		K	DLSHCGGDMPVVQSLR	N	16	0.9	0.93	0.6	0.7	0.1	-0.32	1	1	1	3	1	2	0	6				
9	Q505F1		R	NGDTSFGAFHHDIQTNGDVSR	A	21	0.89	0.79	0.32	0.72	0.11	0.5	1	1	1	3	1	2	0	6				
10	Q505F1		R	LTMPSPMPEYLNVHYIGESASR	L	22	0.88	0.95	0.11	0.65	0.03	0.42	1	1	1	3	1	2	11	6				
11	Q505F1		K	VFLTTPDAAGVNQLFFTSPDLSAPHLQLLTEK	S	32	0.85	0.91	0.04	0.75	0	0.42	1	2	2	3	1	3	0	8				
12	Q505F1	PAp00773	R	LMNATITEELFFK	G	13	0.84	0.8	0.13	0.63	0.06	0.53	1	1	1	3	1	2	5	6				
13	Q505F1		K	ALTPGESSACQSPEEGMEGSPHLIAGEPSFVEK	E	33	0.72	0.76	0.62	0.57	0	0.11	1	1	1	3	1	1	0	5				
14	Q505F1		K	LCIDGHEYAYLK	A	12	0.64	0.55	0.16	0.6	0.14	-0.28	1	1	1	3	1	2	0	6				
15	Q505F1		K	TEPAMLMAPDK	A	11	0.63	0.6	0.65	0.46	0.02	0.69	1	1	1	3	1	2	0	6				
16	Q505F1		K	AESCQGDLSTLASVVTSLANLGK	A	23	0.6	0.67	0.2	0.71	0	-0.09	1	1	1	3	1	2	0	6				
17	Q505F1		K	MEPADYNSQIIGHSL	-	15	0.57	0.96	0.61	-1	0.05	1.64	1	1	1	3	1	2	0	6				
18	Q505F1		-	MATIEEIAHQIIDQQMGEIVTEQQTGQK	I	28	0.55	0.59	0.26	0.56	0.02	0.35	1	2	2	3	1	3	0	8				
20	Q505F1		K	SLMEHIFK	L	8	0.13	0.3	0.13	0.28	0.03	0.29	1	1	1	3	1	2	0	6				
25	Q505F1		R	HYGAITCEGCK	G	11	0.06	0	0.19	0.67	0.09	-0.77	1	2	2	3	1	3	0	8				
27	Q505F1		K	SSNCAASTEK	I	10	0.06	0.03	0.18	0.63	0.02	-0.24	1	2	2	3	1	3	9	8				
28	Q505F1		R	LLFLSMHWALSIPSFQALGQENSISLVK	A	28	0.05	0.01	0.02	0.65	0	-0.1	1	1	1	3	1	2	10	6				
29	Q505F1		K	SPDQGPNK	V	8	0.05	0	0.13	0.37	0.01	0.35	1	2	2	3	1	3	0	8				
30	Q505F1		K	QDSVQCER	K	8	0.05	0	0.21	0.64	0	-0.46	1	2	2	3	1	3	9	8				
31	Q505F1		R	CIAFGMK	Q	7	0.04	0.01	0.38	0.33	0.01	-0.23	1	2	2	3	1	3	9	8				
33	Q505F1		K	DCVINK	H	6	0.04	0	0.2	0.26	0.02	-0.37	1	2	2	3	1	3	0	8				
34	Q505F1		K	AYWNELFTLGLAQCWQVMNVATILATFVNCLHSSLQQDK	M	39	0.03	0	0.01	0.53	0	-0.98	1	1	1	3	1	2	0	6				
35	Q505F1		R	KPIEVSR	E	7	0.03	0.01	0.17	0.21	0.01	-0.46	1	2	2	3	1	3	9	8				

Proteotypic peptides

PeptideAtlas Predictions

Nr2c1_Q505F1

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities	MQ.MS1.Rank.1	TIC MS2 Rank 2+	TIC MS2 Rank 3+
9	O54714	PAp004425	R	VSSIVAPGSSLR	E	12	0.88	0.87	0.83	0.77	0.01	0.34	1	3	3	4	1	4	0	11	3787000	1	2	
1	O54714	PAp015268	K	LTADPDSEVATTSLR	V	15	0.94	0.98	0.83	0.78	0.08	0.93	1	3	3	5	1	4	5	12	3032300	2	4	
7	O54714		R	FEEAHFTFALTPQQLQQILTSR	E	22	0.9	0.99	0.1	0.74	0.04	0.26	1	3	3	5	1	5	0	13	2822600	4		1
2	O54714		K	GALTSGHQPSSVLR	S	14	0.93	0.97	0.7	0.85	0.06	0.43	1	3	3	4	1	4	0	11	2537800	5	8	2
11	O54714	PAp015286	R	LSATVPNTIVVNWSSEFGR	N	19	0.85	0.83	0.08	0.77	0.05	0.33	1	3	3	5	1	5	9	13	2456100	6	3	4
6	O54714		R	NYSLSVYLVR	Q	10	0.9	0.8	0.37	0.67	0.17	0.09	1	3	3	5	1	5	5	13	1369700	7	5	
4	O54714	PAp015342	R	EGHGGPLPSGPSLTGCR	S	17	0.91	0.85	0.5	0.78	0.14	-0.04	1	3	3	4	1	4	5	11	805050	8		3
17	O54714		K	CDYTIQVQLR	F	10	0.14	0.06	0.36	0.65	0.15	-0.57	1	3	3	5	1	5	9	13	573440	9	7	
12	O54714		K	TLGPSDLSLLSLPPGTSPVGSPGPLAPIPPTLLTPGTLLGPK	R	42	0.84	0.92	0.35	0.58	0	0.37	1	2	2	3	1	3	0	8				5
3	O54714		R	VSELQVLLGFAGR	N	13	0.92	0.91	0.43	0.72	0.13	0.36	2	8	8	10	2	10	22	28			1	
24	O54714		K	ALHLLK	S	6	0.05	0.02	0.22	0.08	0.06	0.29	3	9	9	13	3	11	33	33			6	
30	O54714		R	LTVPCR	A	6	0.03	0	0.29	0.18	0.03	-0.48	2	4	4	6	2	5	5	15	2851300	3		
5	O54714		R	GTPSHFLGPLAPTLGSSHR	S	19	0.9	0.99	0.37	0.72	0.02	0.45	1	3	3	4	1	4	0	11				
8	O54714		K	HCPVTSAAIPALPGSK	G	16	0.88	0.82	0.78	0.61	0.11	0.03	1	3	3	4	1	4	0	11				
10	O54714		R	VEVIDLTIESSSDEEDLPPTK	K	21	0.88	0.95	0.24	0.49	0	1.19	1	3	3	4	1	4	5	11				
13	O54714		K	EASEVCPPPGYGLDGLQYSAVQEGIQPESK	K	30	0.82	0.91	0.45	0.53	0.01	0.18	1	3	3	4	1	4	0	11				
14	O54714		R	QLTAGTLLQK	L	10	0.78	0.6	0.68	0.66	0.09	0.42	1	3	3	5	1	5	5	13				
15	O54714	PAp015349	K	RPSRPINITPLAR	L	13	0.74	0.94	0.31	0.63	0	-0.97	1	3	3	5	1	5	9	13				
16	O54714		K	LCPLPGYLPPTK	N	12	0.22	0.23	0.45	0.39	0.13	0.01	1	3	3	5	1	5	9	13				
18	O54714		R	ALTCAHLQSFDAALYLQMNEK	K	21	0.08	0.24	0.08	0.54	0	-0.39	1	3	3	5	1	4	5	12				
19	O54714		R	SSTPAPPPGR	V	10	0.06	0.08	0.28	0.37	0	0.45	1	3	3	4	1	4	0	11				
20	O54714	PAp008570	-	MAELGELK	H	8	0.06	0.04	0.55	0.36	0	0.45	1	2	2	2	1	3	6	7				
21	O54714	PAp001381	K	KPTWTCPVCDK	K	11	0.06	0	0.29	0.66	0.09	-1.12	1	3	3	5	1	4	5	12				
22	O54714		R	EVLPGAK	C	7	0.05	0.01	0.21	0.27	0.03	0.57	2	5	5	6	2	7	13	18				
23	O54714		K	SSCAPSVQMK	I	10	0.05	0.01	0.33	0.39	0.04	-0.36	1	3	3	6	1	5	10	14				
25	O54714		R	VSLMCPLGK	M	9	0.05	0.01	0.49	0.24	0.04	-0.22	2	8	8	9	2	9	18	26				
26	O54714		R	FCLCETSCPQEDYFPPNLFVK	V	21	0.04	0	0.13	0.69	0.01	-0.82	1	3	3	5	1	5	9	13				
27	O54714		K	HELLAK	A	6	0.04	0.02	0.16	0.15	0.01	0.32	1	3	3	6	1	6	10	15				
28	O54714		K	HMVMSFR	V	7	0.04	0.01	0.19	0.17	0.04	-0.08	1	2	2	2	1	3	6	7				
29	O54714		K	NGAEPK	R	6	0.04	0	0.03	0.06	0.03	0.3	1	3	3	5	1	5	9	13				
31	O54714		R	NPDHSR	A	6	0.03	0	0.05	0.09	0.01	-0.02	3	9	9	12	3	11	23	32				
32	O54714		K	APYESLIIDGLFMEILNSCSDCDEIQFMEDGSWCPMKPK	K	39	0.03	0	0.05	0.37	0	-0.99	1	3	3	5	1	4	0	12				
33	O54714		R	SDVISLD	-	7	0.03	0.04	0.34	-1	0.06	1.11	1	3	3	4	1	4	0	11				
34	O54714		R	EVDMHPPLPQPVHPDVTMKPLPFYEVYGELIRPTTLASTSSQR	F	43	0.02	0.01	0.18	-1	0	-0.35	1	2	2	3	1	3						

Proteotypic peptides

PeptideAtlas Predictions

Pias3_O54714

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities	MQ.MS1.Rank.1	TIC MS2 Rank 2+	TIC MS2 Rank 3+
20	Q9JM05	PAp020391	R	VSLICPLVK	M	9	0.06	0.05	0.37	0.31	0.07	-0.12	1	1	1	1	1	2	3	4	11113000	1	4	
11	Q9JM05	PAp020391	K	LQESPCIFALTPR	Q	13	0.5	0.49	0.65	0.52	0.09	-0.3	1	1	1	2	1	2	3	5	8903900	3	2	
10	Q9JM05	PAp020391	K	SYSVALYLVR	Q	10	0.52	0.36	0.38	0.68	0.13	0.09	1	1	1	3	1	3	3	7	8281900	4	1	
1	Q9JM05		R	TPLSGPTVDYPVLYGK	Y	16	0.94	0.97	0.58	0.8	0.12	0.75	1	1	1	2	1	2	0	5	7449100	5	6	
17	Q9JM05		K	AVQVVLR	I	7	0.07	0.03	0.34	0.47	0.04	0.18	1	1	1	2	1	2	3	5	7087600	7	5	
3	Q9JM05	PAp020391	R	ITVTWGNYGK	S	10	0.88	0.78	0.42	0.68	0.11	0.33	1	1	1	3	1	3	3	7	4887800	10	7	
27	Q9JM05		R	LSVPCR	A	6	0.04	0.01	0.24	0.13	0.06	-0.54	1	1	1	2	2	3	3	6	4345200	11	3	
16	Q9JM05		K	ILSECEGADEIEFLAEGSWRPIR	A	23	0.07	0.19	0.09	0.52	0.03	-0.68	1	1	1	1	1	1	0	3	3452300	13		1
15	Q9JM05	PAp01098!	R	ICYSDTSCPQEDQYPPNIAVK	V	21	0.07	0.09	0.24	0.57	0.05	-0.45	1	1	1	2	1	2	3	5	607960	18		2
6	Q9JM05		R	QLTSSDLLQR	L	10	0.82	0.6	0.6	0.74	0.13	0.17	1	1	1	3	1	3	0	7	9916700	2		
7	Q9JM05	PAp00533	R	ALQLVQFDCSPELFK	K	15	0.78	0.66	0.2	0.54	0.16	-0.16	1	1	1	2	1	2	3	5	3462400	12		
19	Q9JM05		K	YLNGLGR	L	7	0.06	0.05	0.36	0.29	0.06	0.14	1	1	1	2	1	2	3	5	2070400	15		
5	Q9JM05	PAp00747	R	VSDLQMLLGFVGR	S	13	0.86	0.83	0.22	0.68	0.08	0.27	1	1	1	2	1	2	3	5	838390	16		
2	Q9JM05	PAp01542(	R	LDPDSEIATTGVR	V	13	0.94	0.95	0.94	0.75	0.07	0.98	1	1	1	1	1	2	3	4				
4	Q9JM05		K	SAEPGPQAPRPLDPLALHSMPR	T	22	0.88	0.93	0.48	0.64	0.04	0.12	1	1	1	2	1	2	0	5				
8	Q9JM05		K	LPFFNMLDELLKPTELVPQSAEK	L	23	0.76	0.76	0.14	0.72	0.01	0.32	1	1	1	2	1	2	0	5				
9	Q9JM05	PAp00501!	K	VNHSYCSVPGYYPSNKPGVEPK	R	22	0.66	0.7	0.34	0.46	0.09	-0.45	1	1	1	2	1	2	3	5				
12	Q9JM05		-	MAAELVEAK	N	9	0.19	0.21	0.58	0.4	0.06	0.56	1	1	1	2	1	2	3	5				
13	Q9JM05		K	TGADVVDLTLDSSSSSEDEDEDEDDDEDEDEGPRPK	R	36	0.13	0.24	0.34	0.43	0	0.62	1	1	1	1	1	1	0	3				
14	Q9JM05		K	GLVPAC	-	6	0.1	0.01	0.45	0.7	0.04	0.36	1	1	1	1	1	1	3	3				
18	Q9JM05		R	QVEMIR	N	6	0.06	0.02	0.29	0.68	0.01	-0.07	2	2	2	3	2	3	10	8				
21	Q9JM05		K	ELYETR	Y	6	0.06	0.01	0.09	0.7	0	-0.04	1	1	1	2	1	2	3	5				
22	Q9JM05		K	HELVTR	A	6	0.04	0.03	0.11	0.16	0.05	0.03	1	1	1	2	1	2	3	5				
23	Q9JM05		K	RPCRPINLTHLMYLSSATNR	I	20	0.04	0.08	0.18	0.73	0.01	-1.66	1	1	1	2	1	2	3	5				
24	Q9JM05		R	ELQPGVK	A	7	0.04	0.04	0.15	0.15	0	0.45	1	1	1	2	1	2	3	5				
25	Q9JM05	PAp020391	R	AETCAHLQCFDAVFYLQMNEK	K	21	0.04	0.03	0.05	0.47	0.04	-0.85	1	1	1	1	1	1	3	3				
26	Q9JM05		K	NMVMSFR	V	7	0.04	0.01	0.25	0.23	0.01	-0.05	1	1	1	2	1	2	3	5				
28	Q9JM05		K	KPTWMCPVCDKPAAYDQLIIDGLLSK	I	26	0.03	0	0.11	0.62	0.01	-1.38	1	1	1	1	1	1	0	3				
29	Q9JM05		K	TLKPEVR	L	7	0.03	0.02	0.22	0.2	0	-0.44	1	1	1	2	1	2	3	5				
30	Q9JM05		K	HPELCK	A	6	0.03	0.01	0.08	0.07	0.02	-0.4	1	1	1	3	1	3	3	7				

🟨 Proteotypic peptides

🟩 PeptideAtlas Predictions

Pias4_Q9JM05

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities.y	MQ.MS1.Rank.2	TIC MS2 Rank 2+	TIC MS2 Rank 3+
1	P37238	PAp007322	R	SVEAVQEITEYAK	N	13	0.95	0.97	0.72	0.81	0.13	0.91	1	2	2	2	1	3	11	7	24270000	1	1	
4	P37238	PAp00892	K	LLAEISSDIDQLNPESADLR	A	20	0.9	0.99	0.1	0.62	0	1.03	1	2	2	2	1	4	15	8	3060000	3	2	1
21	P37238		R	IFQGCQFR	S	8	0.05	0.02	0.26	0.48	0.03	-0.52	1	2	2	2	1	3	11	7	2658200	4	4	
7	P37238		K	VEPASPPYYSEK	T	12	0.84	0.67	0.51	0.61	0.1	0.71	1	2	2	2	1	4	20	8	2360500	5	3	
17	P37238		K	HLYDSYIK	S	8	0.08	0.1	0.13	0.26	0.09	0.26	1	2	2	2	1	4	16	8	2360400	6	5	
5	P37238	PAp01530(	K	LNHPESSQLFAK	V	12	0.88	0.92	0.58	0.64	0.01	0.68	1	2	2	2	1	3	11	7	1463900	8	6	
9	P37238		K	NIPGFINLDLNDQVTLLK	Y	18	0.65	0.56	0.06	0.65	0.04	0.87	1	2	2	2	1	3	0	7	0	11	7	
15	P37238		K	HITPLQEQSK	E	10	0.15	0.11	0.41	0.56	0.07	0.49	1	2	2	2	1	3	11	7	3179100	2		
6	P37238		R	QIVTEHVQLLHVIK	K	14	0.85	0.71	0.27	0.63	0.13	0.43	1	2	2	2	1	3	0	7	2042400	7		
2	P37238	PAp01540	K	DGVLISEGQGFMTR	E	14	0.94	0.98	0.8	0.79	0.14	0.54	1	2	2	2	1	3	11	7	846270	9		
20	P37238		K	FEFAVK	F	6	0.05	0.04	0.35	0.3	0	0.37	2	3	3	3	2	4	24	10	829700	10		
3	P37238	PAp01537	K	TETDMSLHPLLQEIYK	D	16	0.91	0.91	0.25	0.68	0.07	0.95	1	2	2	2	1	3	11	7				
8	P37238		R	KPFGDFMEPK	F	10	0.68	0.62	0.46	0.56	0.08	-0.02	1	2	2	2	1	3	11	7				
10	P37238		R	ADPMVADYK	Y	9	0.46	0.41	0.59	0.65	0.03	0.59	1	2	2	2	1	4	0	8				
11	P37238		K	SPFVIYDMNSLMMGEDK	I	17	0.32	0.32	0.07	0.67	0.05	0.56	1	2	2	2	1	3	11	7				
12	P37238		K	TQLYNRPHEEPSNSLMAIECR	V	21	0.22	0.42	0.26	0.66	0.04	-0.92	1	2	2	2	1	4	0	8				
13	P37238		K	LQEYQSAIK	V	9	0.17	0.25	0.51	0.35	0.06	0.34	1	2	2	2	1	4	20	8				
14	P37238		K	ASGFHYGVHACEGCK	G	15	0.16	0.07	0.23	0.73	0.17	-0.74	4	5	5	7	4	7	38	19				
16	P37238		K	CLAVGMSHNAIR	F	12	0.09	0.14	0.42	0.69	0.02	-0.47	1	2	2	2	1	4	14	8				
18	P37238		K	YGVHEIIYTMLASLMNK	D	17	0.08	0.08	0.04	0.69	0	0.23	1	2	2	2	1	3	11	7				
19	P37238		R	AILTGK	T	6	0.05	0.02	0.33	0.26	0.02	0.42	3	4	4	4	3	9	25	17				
22	P37238		K	SFPLTK	A	6	0.04	0.01	0.18	0.2	0.03	0.31	1	2	2	2	1	5	17	9				
23	P37238		R	MPQAEK	E	6	0.03	0.01	0.07	0.06	0.01	0.16	3	6	6	5	3	9	15	20				
24	P37238		R	CDLNCR	I	6	0.03	0	0.12	0.16	0.03	-1.04	1	2	2	2	1	4	17	8				
25	P37238		K	FNALELDDSDLAIFIAVIILSGDRPGLLNVKPIEDIQDNLLQALELQLK	L	49	0.02	0	-1	-1	0	-0.02	1	2	2	2	1	3	11	7				

Proteotypic peptides

PeptideAtlas Predictions

Pparg_P37238

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities	MQ.MS1.Rank.1	TIC MS2 Rank 2+	TIC MS2 Rank 3+
16	P18911	PAp00409	R	GLGQPDLPK	E	9	0.08	0.04	0.39	0.33	0.05	0.66	1	1	1	2	1	2	5	5	4498200	3	1	
9	P18911	PAp01528	K	LQEPLLEALR	L	10	0.81	0.72	0.6	0.51	0.08	0.42	1	3	3	2	1	3	10	8	7300400	1	2	
18	P18911		K	FSELATK	C	7	0.05	0.07	0.21	0.25	0.02	0.36	2	7	7	5	2	7	25	19	3495900	5	3	
12	P18911		R	VQLDLGLWDK	F	10	0.44	0.44	0.44	0.55	0.04	0.48	1	3	3	2	1	3	12	8	1161900	7	4	
1	P18911	PAp01540	R	LPGFTGLSIADQITLLK	A	17	0.93	0.98	0.18	0.76	0.1	0.6	1	3	3	2	1	3	13	8	378120	11	5	
15	P18911	PAp00164	K	SSGYHYGVSSCEGCK	G	15	0.13	0.05	0.24	0.71	0.16	-0.73	1	3	3	3	1	4	9	10	1233300	6	6	1
23	P18911		R	VYKPCFVCNDK	S	11	0.04	0	0.29	0.39	0.08	-1.18	1	3	3	3	1	4	10	10	3934600	4	7	2
6	P18911		R	LFAPGALGPGSGYPGAGFPFAFPGALR	G	27	0.86	0.86	0.27	0.87	0.02	0.19	1	1	1	2	1	2	0	5				3
19	P18911		K	IVEFAK	R	6	0.05	0.02	0.38	0.21	0	0.41	2	7	7	5	2	7	24	19	5391100	2		
17	P18911		K	AACLDILMLR	I	10	0.06	0	0.33	0.5	0.07	-0.35	1	3	3	2	1	3	13	8	452460	10		
2	P18911		R	YTPEQDTMTFSDGLTLNR	T	18	0.91	0.96	0.21	0.72	0.06	0.66	3	9	9	11	3	9	41	29				
3	P18911	PAp01540	K	EEGSPDSYELSPQLEELITK	V	20	0.9	0.96	0.37	0.63	0	1.01	1	3	3	3	1	4	12	10				
4	P18911	PAp00895	R	GSPPFEMLSPSFR	G	13	0.89	0.96	0.65	0.49	0.05	0.41	1	1	1	2	1	2	5	5				
5	P18911		K	AHQETFPSLCQLGK	Y	14	0.88	0.88	0.61	0.65	0.11	-0.26	2	7	7	6	2	8	25	21				
7	P18911	PAp00444	K	MEIPGPMPPLIR	E	12	0.84	0.74	0.49	0.54	0.1	0.44	1	3	3	2	1	3	10	8				
8	P18911		K	EMASLSVETQSTSSEEMVPSSPSPPPPPR	V	29	0.81	0.87	0.54	0.4	0.02	0.5	1	1	1	2	1	2	5	5				
10	P18911		R	EMLENPEMFEDDSSKPGPHPK	A	21	0.77	0.83	0.51	0.34	0.03	0.43	1	3	3	2	1	3	0	8				
11	P18911		K	ASSEDEAPGGQGK	R	13	0.54	0.52	0.16	0.58	0.04	0.61	1	3	3	2	1	3	0	8				
13	P18911		K	YTTNSSADHR	V	10	0.3	0.26	0.12	0.63	0.11	0.15	2	7	7	6	2	8	24	21				
14	P18911		R	RPSQPYMFPR	M	10	0.19	0.46	0.34	0.45	0.01	-0.58	1	3	3	2	1	3	10	8				
20	P18911		K	NMVYTCHR	D	8	0.04	0	0.15	0.55	0.04	-0.71	2	5	5	8	2	6	25	19				
21	P18911		K	CFEVGMSK	E	8	0.04	0.01	0.25	0.41	0.03	-0.41	2	7	7	6	2	8	38	21				
22	P18911		R	MDLEEPEK	V	8	0.04	0.02	0.34	0.13	0	0.37	1	3	3	2	1	3	10	8				
24	P18911		K	NCIINK	V	6	0.03	0	0.26	0.13	0.01	-0.33	2	5	5	9	2	6	25	20				
25	P18911		R	GQSPQPDQGP	-	10	0.02	0.01	0.28	-1	0.03	0.92	1	3	3	2	1	3	0	8				
26	P18911		R	TQMHNAGFGPLTDLVFAFAGQLLPLEMDDTETGLLSAICLICGDR	M	45	0.02	0	0.01	-1	0	-0.69	1	3	3	2	1	3	10	8				

Proteotypic peptides

PeptideAtlas Predictions

Rarg_P18911

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities	MQ.MS1.Rank.3	TIC MS2 Rank 2+	TIC MS2 Rank 3+
7	P28700		K	HFLPLDFSTQVNSSSLNSPTGR	G	22	0.89	0.98	0.18	0.66	0.04	0.38	1	1	1	2	1	1	0	4	26781000	2	10	1
6	P28700	PAp00747855	R	AGWNELLIASFSHR	S	14	0.9	0.94	0.26	0.52	0.11	0.2	3	4	4	12	3	8	39	24	20517000	3	4	4
2	P28700	PAp00746849	K	ILEAELAVEPK	T	11	0.94	0.84	0.66	0.58	0.23	0.9	2	2	2	8	2	5	7	15	20343000	4	3	
3	P28700	PAp01449315	K	GLSNPAEVEALR	E	12	0.94	0.96	0.93	0.74	0.08	0.75	1	1	1	4	1	3	4	8	18740000	5	2	
10	P28700	PAp01926880	R	AIVLFNPDSK	G	10	0.79	0.69	0.85	0.34	0.07	0.7	1	1	1	4	1	3	4	8	16150000	7	1	
1	P28700	PAp00752323	R	NSAHSAGVGAIFDR	V	14	0.95	0.85	0.69	0.85	0.24	0.52	2	3	3	8	2	6	42	17	8635300	8	6	2
18	P28700		R	VLTELVSK	M	8	0.08	0.04	0.41	0.48	0.03	0.42	3	4	4	12	3	8	45	24	8107900	10	5	
15	P28700	PAp01926875	K	VYASLEAYCK	H	10	0.08	0.05	0.38	0.52	0.09	-0.33	1	1	1	4	1	3	4	8	4544800	14	7	
16	P28700	PAp00029245	K	CLEHLFFFK	L	9	0.08	0	0.14	0.34	0.16	-0.26	3	4	4	12	3	8	39	24	3409500	16	9	5
14	P28700	PAp00010394	K	HYGVYSCEGCK	G	11	0.08	0	0.17	0.66	0.14	-0.81	3	4	4	14	3	9	47	27	1128800	21	11	
8	P28700	PAp00606610	R	IPHFSELPLDDQVILLR	A	17	0.89	0.91	0.13	0.63	0.05	0.75	1	1	1	4	1	3	4	8	42668000	1	8	3
13	P28700	PAp00745638	K	QLFTLVEWAK	R	10	0.36	0.36	0.18	0.71	0.05	0.28	3	4	4	12	3	8	46	24	17025000	6		
5	P28700	PAp01791977	K	VPAHPSGNMASFTK	H	14	0.92	0.97	0.6	0.69	0.04	0.69	1	1	1	5	1	4	4	10	8514500	9		
11	P28700		R	NENEVESTSSANEDMPVEK	I	19	0.7	0.7	0.37	0.47	0.01	0.78	1	1	1	4	1	3	0	8	6568800	11		
9	P28700	PAp01926882	K	DGILLATGLHVHR	N	13	0.87	0.89	0.43	0.58	0.05	0.45	3	4	4	12	3	8	39	24	5716300	13		
12	P28700	PAp01926887	K	TETYVEANMGLNPSSPNDPVTNICQAADK	Q	29	0.52	0.55	0.41	0.67	0.01	0.21	1	1	1	4	1	3	4	8	3176600	17		
19	P28700	PAp01926868	K	TELGCLR	A	7	0.04	0.02	0.3	0.5	0.01	-0.62	2	3	3	8	2	6	42	17	2345200	18		
23	P28700	PAp01638909	K	DLTYTCR	D	7	0.04	0.01	0.2	0.35	0.02	-0.67	1	1	1	5	1	4	4	10	1264100	20		
4	P28700		K	LIGDTPIDTFLMEMLEAPHQAT	-	22	0.93	0.99	0.05	0.75	0.04	1.43	1	1	1	4	1	3	0	8				
17	P28700	PAp01456379	K	HICAICGDR	S	9	0.08	0	0.26	0.57	0.15	-0.88	2	2	2	9	2	6	7	17				
20	P28700		R	EAVQEER	Q	7	0.04	0.01	0.18	0.29	0	0.11	3	4	4	12	3	8	47	24				
21	P28700	PAp01926863	K	YPEQPGR	F	7	0.04	0.02	0.23	0.26	0	0.1	2	2	2	8	2	5	7	15				
22	P28700		K	DCLIDK	R	6	0.04	0.01	0.24	0.16	0.04	-0.31	2	2	2	9	2	6	7	17				
24	P28700		R	DMQMDK	T	6	0.03	0	0.09	0.09	0.02	0.04	2	2	2	8	2	5	7	15				
25	P28700		K	CLAMGMK	R	7	0.03	0	0.36	0.13	0.01	-0.32	1	1	1	4	1	3	4	8				

Proteotypic peptides

PeptideAtlas Predictions

Rxra_P28700

list	Biosequence Name	Peptide Accession	Pre AA	Sequence	Fol AA	Peptide Length	Combined Predictor Score	PSieve	ESPP	DPred	APEX	STEPP	N SP Mapping	N SP-varsplic Mapping	N SP-nsSNP Mapping	N ENSP Mapping	N ENSG Mapping	N IPI Mapping	N Human Mapping	N Mouse Mapping	Intensities.y	MQ.MS1.Rank.2	TIC MS2 Rank 2+	TIC MS2 Rank 3+
1	Q62432	PAp004979	R	VGETFHASQPSLTVDGFTDPSNSER	F	25	0.94	0.94	0.51	0.7	0.19	0.7	1	2	2	4	1	2	8	8	490670	2		1
12	Q62432	PAp005012	R	YGWHPATVCK	I	10	0.08	0.11	0.21	0.6	0.05	-0.34	2	3	3	5	2	4	18	12	199240	4	3	
3	Q62432	PAp003856	R	VETPVLPPVLVPR	H	13	0.85	0.79	0.52	0.58	0.06	0.56	3	4	4	9	3	8	24	21	8815200	1	1	
8	Q62432	PAp000023	R	FCLGLLSNVNR	N	11	0.14	0.03	0.28	0.77	0.14	-0.39	5	6	6	13	5	13	37	32			2	
16	Q62432	PAp005281	K	*GWGAEYR*	R	7	0.05	0.05	0.21	0.36	0.04	0.03	2	3	3	5	2	4	18	12	367170	3		
2	Q62432	PAp004950	K	SAGGSGGAGGGEQNGQEEK	W	19	0.89	0.91	0.34	0.79	0.03	0.49	1	2	2	6	1	2	8	10				
4	Q62432		-	MSSILPFTPPVVK	R	13	0.58	0.62	0.55	0.43	0.02	0.47	1	2	2	6	1	2	8	10				
5	Q62432		K	AIENCEYAFNLK	K	12	0.57	0.61	0.41	0.51	0.07	-0.21	1	2	2	6	1	2	9	10				
6	Q62432	PAp015255	K	DEVCVNPYHYQR	V	12	0.36	0.41	0.19	0.46	0.13	-0.38	2	3	3	9	2	7	17	19				
7	Q62432	PAp004727	R	WPDLHSHHELK	A	11	0.22	0.37	0.19	0.31	0.06	0.3	1	2	2	6	1	2	9	10				
9	Q62432	PAp004321	K	GLPHVIYCR	L	9	0.1	0.07	0.32	0.65	0.08	-0.39	5	6	6	17	5	16	39	39				
10	Q62432	PAp001493	K	VLTQMGSPSVR	C	11	0.09	0.03	0.92	0.55	0.03	0.27	1	2	2	3	1	2	7	7				
11	Q62432	PAp003955	K	IFNNQEFAALLAQSVNQGFEAVYQLTR	M	27	0.09	0.08	0.04	0.75	0.01	0.12	2	3	3	5	2	4	18	12				
13	Q62432		R	NATVEMTR	R	8	0.07	0.06	0.53	0.6	0	0.13	1	2	2	4	1	2	8	8				
14	Q62432		R	QTVTSTPCWIELHLNGPLQWLDK	V	23	0.07	0.11	0.06	0.65	0.02	-0.29	2	3	3	4	2	4	17	11				
15	Q62432	PAp010515	K	AITTQNCNTK	C	10	0.06	0	0.28	0.6	0.04	-0.23	1	2	2	6	1	2	8	10				
17	Q62432		K	IPPGCNLK	I	8	0.05	0.01	0.5	0.32	0.04	-0.12	2	3	3	5	2	4	18	12				
18	Q62432		R	LDELEK	A	6	0.04	0.01	0.22	0.1	0.04	0.26	5	6	6	12	5	8	57	26				
19	Q62432		K	CVTIPSTCSEIWGLSTANTVDQWDTTGLYSFSEQTR	S	36	0.04	0	0.07	0.63	0	-0.66	1	1	1	3	1	1	0	5				
20	Q62432		R	LYYIGGEVFAECLSDSAIFVQSPNCNQR	Y	28	0.04	0	0.02	0.59	0	-0.86	2	3	3	5	2	4	18	12				
21	Q62432		R	LQVSHR	K	6	0.03	0.01	0.07	0.12	0.01	-0.11	5	6	6	17	5	16	42	39				

Proteotypic peptides

PeptideAtlas Predictions

Smad2_Q62432

Supplementary Table 3: calculations of PPARγ and RXRα copy-number per cell (1 TF per assay)

		MS intensity	# cells	ug	ugNE/cell	copies/cell*
		1st bio.repl.				
PPARγ	Day 0	0.039	21825000	2025	9.27835E-05	2179
	2h	0.037	26475000	1876	7.08593E-05	1579
	Day1	0.028	36150000	2048	5.66528E-05	955
	Day2	0.069	39375000	2104	5.34349E-05	2220
	Day4	0.153	61125000	4227	6.91534E-05	6372
	Day6	0.072	56486250	3809	6.74323E-05	2924
RXRα	Day 0	0.33	21825000	2025	9.27835E-05	18438
	2h	0.252	26475000	1876	7.08593E-05	10753
	Day1	0.518	36150000	2048	5.66528E-05	17672
	Day2	1.399	39375000	2104	5.34349E-05	45018
	Day4	0.883	61125000	4227	6.91534E-05	36772
	Day6	0.357	56486250	3809	6.74323E-05	14497
		2nd bio.repl.				
PPARγ	Day 0	0.049	24975000	1380	5.52553E-05	1630
	2h	0.04	24450000	1420	5.80777E-05	1399
	Day1	0.092	29287500	1532	5.2309E-05	2898
	Day2	0.281	42000000	2036	4.84762E-05	8203
	Day4	0.303	48000000	2819	5.87292E-05	10716
	Day6	0.299	43950000	3174	7.22184E-05	13003
RXRα	Day 0	0.245	24975000	1380	5.52553E-05	8152
	2h	0.314	24450000	1420	5.80777E-05	10982
	Day1	0.627	29287500	1532	5.2309E-05	19751
	Day2	0.945	42000000	2036	4.84762E-05	27587
	Day4	0.468	48000000	2819	5.87292E-05	16552
	Day6	0.281	43950000	3174	7.22184E-05	12221
		3rd bio.repl.				
PPARγ	Day 0	0.108	24900750	1270	5.10025E-05	3317
	2h	0.074	29025000	1260	4.34109E-05	1935
	Day1	0.182	45234000	1500	3.31609E-05	3634
	Day2	0.342	53700000	2050	3.8175E-05	7862
	Day4	0.308	56212500	2500	4.44741E-05	8249
	Day6	0.235	57412500	2620	4.56347E-05	6458
RXRα	Day 0	0.191	24900750	1270	5.10025E-05	5866
	2h	0.321	29025000	1260	4.34109E-05	8392

	Day1	0.459	45234000	1500	3.31609E-05	9166
	Day2	1.099	53700000	2050	3.8175E-05	25265
	Day4	0.646	56212500	2500	4.44741E-05	17301
	Day6	0.445	57412500	2620	4.56347E-05	12229

calculations of 10 TFs copy-number per cell (multiplex)

		4th bio.repl.				
RXRα	Day 0	0.174	22650000	1264	5.58057E-05	5847
	Day 2	0.551	20175000	1466	7.26642E-05	24111
	Day4	0.307	37275000	2249	6.03353E-05	11155
PPARγ	Day 0	0.07	22650000	1264	5.58057E-05	2352
	Day 2	0.194	20175000	1466	7.26642E-05	8489
	Day4	0.176	37275000	2249	6.03353E-05	6395
Pias3	Day 0	0.023	22650000	1264	5.58057E-05	773
	Day 2	0.011	20175000	1466	7.26642E-05	481
	Day4	0.007	37275000	2249	6.03353E-05	254
Pias4	Day 0	0.172	22650000	1264	5.58057E-05	5780
	Day 2	0.174	20175000	1466	7.26642E-05	7614
	Day4	0.136	37275000	2249	6.03353E-05	4941
NFIB	Day 0	5.659	22650000	1264	5.58057E-05	190178
	Day 2	6.984	20175000	1466	7.26642E-05	305608
	Day4	4.703	37275000	2249	6.03353E-05	170879
SMAD2	Day 0	0.431	22650000	1264	5.58057E-05	14484
	Day 2	0.341	20175000	1466	7.26642E-05	14922
	Day4	0.32	37275000	2249	6.03353E-05	11627
RARg	Day 0	0.129	22650000	1264	5.58057E-05	4335
	Day 2	0.106	20175000	1466	7.26642E-05	4638
	Day4	0.104	37275000	2249	6.03353E-05	3779
NR2C1	Day 0	0.043	22650000	1264	5.58057E-05	1445
	Day 2	0.041	20175000	1466	7.26642E-05	1794
	Day4	0.037	37275000	2249	6.03353E-05	1344
FOSL2	Day 0	0.077	22650000	1264	5.58057E-05	2588
	Day 2	0.093	20175000	1466	7.26642E-05	4070
	Day4	0.073	37275000	2249	6.03353E-05	2652
ARID3a	Day 0	0.015	22650000	1264	5.58057E-05	504
	Day 2	0.013	20175000	1466	7.26642E-05	569
	Day4	0.013	37275000	2249	6.03353E-05	472

Supplementary Table 4: primer design for the reference peptide-tag variants for (multiplexing)

Tag name (amino acid substitution)		nucleotide sequence (5’ to 3’)
SH24-pF3A-GST (I4G)	FW	atcgatggaaaaagcggccgatggcacaagtttgtacaaaaaagc
	RV	gcttttttgtacaaacttgtgccatcggccgctttttccatcgat
SH25-pF3A-GST (I4A)	FW	atcgatggaaaaagcggccgatgccacaagtttgtacaaaaaagc
	RV	gcttttttgtacaaacttgtggcatcggccgctttttccatcgatt
SH26-pF3A-GST (I4V)	FW	cgatggaaaaagcggccgatgtcacaagtttgtacaaaa
	RV	ttttgtacaaacttgtgacatcggccgctttttccatcg
SH27-pF3A-GST (I4F)	FW	cgatggaaaaagcggccgatttcacaagtttgtacaaaa
	RV	ttttgtacaaacttgtgaaatcggccgctttttccatcg
SH28-pF3A-GST (D3E)	FW	cgatggaaaaagcggccgagatcacaagtttgtacaaaa
	RV	ttttgtacaaacttgtgatctcggccgctttttccatcg
SH29-pF3A-GST (D3E- I4G)	FW	gatcgatggaaaaagcggccgagggcacaagtttgtacaaaaaagctg
	RV	cagcttttttgtacaaacttgtgccctcggccgctttttccatcgatc
SH30-pF3A-GST (D3E-I4A)	FW	gatcgatggaaaaagcggccgaggccacaagtttgtacaaaaaagctg
	RV	cagcttttttgtacaaacttgtggcctcggccgctttttccatcgatc
SH31-pF3A-GST (D3E-I4V)	FW	cgatcgatggaaaaagcggccgaggtcacaagtttgtacaaaaaag
	RV	cttttttgtacaaacttgtgacctcggccgctttttccatcgatcg
SH32-pF3A-GST (D3E-I4F)	FW	cgatcgatggaaaaagcggccgagttcacaagtttgtacaaaaaag
	RV	cttttttgtacaaacttgtgaactcggccgctttttccatcgatcg

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
RXR	VLTELVSK	444.7736	333.2127	y3
RXR	VLTELVSK	444.7736	446.2968	y4
RXR	VLTELVSK	444.7736	575.3394	y5
RXR	VLTELVSK	444.7736	676.387	y6
RXR	VLTELVSK	444.7736	789.4711	y7
RXR	VLTELVSK[HeavyK]	448.7807	341.2269	y3
RXR	VLTELVSK[HeavyK]	448.7807	454.311	y4
RXR	VLTELVSK[HeavyK]	448.7807	583.3536	y5
RXR	VLTELVSK[HeavyK]	448.7807	684.4012	y6
RXR	VLTELVSK[HeavyK]	448.7807	797.4853	y7
RXR	AIVLFNPDSK	552.3084	349.1712	y3
RXR	AIVLFNPDSK	552.3084	446.224	y4
RXR	AIVLFNPDSK	552.3084	560.2669	y5
RXR	AIVLFNPDSK	552.3084	707.3353	y6
RXR	AIVLFNPDSK	552.3084	820.4194	y7
RXR	AIVLFNPDSK	552.3084	919.4878	y8
RXR	AIVLFNPDSK	552.3084	1032.572	y9
RXR	AIVLFNPDSK[HeavyK]	556.3155	357.1854	y3
RXR	AIVLFNPDSK[HeavyK]	556.3155	454.2382	y4
RXR	AIVLFNPDSK[HeavyK]	556.3155	568.2811	y5
RXR	AIVLFNPDSK[HeavyK]	556.3155	715.3495	y6
RXR	AIVLFNPDSK[HeavyK]	556.3155	828.4336	y7
RXR	AIVLFNPDSK[HeavyK]	556.3155	927.502	y8
RXR	AIVLFNPDSK[HeavyK]	556.3155	1040.586	y9
RXR	ILEAELAVEPK	606.3477	373.2076	y3
RXR	ILEAELAVEPK	606.3477	472.276	y4
RXR	ILEAELAVEPK	606.3477	543.3131	y5
RXR	ILEAELAVEPK	606.3477	656.3972	y6
RXR	ILEAELAVEPK	606.3477	785.4398	y7
RXR	ILEAELAVEPK	606.3477	856.4769	y8
RXR	ILEAELAVEPK	606.3477	985.5195	y9
RXR	ILEAELAVEPK	606.3477	1098.604	y10
RXR	ILEAELAVEPK[HeavyK]	610.3548	381.2218	y3
RXR	ILEAELAVEPK[HeavyK]	610.3548	480.2902	y4
RXR	ILEAELAVEPK[HeavyK]	610.3548	551.3273	y5
RXR	ILEAELAVEPK[HeavyK]	610.3548	664.4114	y6
RXR	ILEAELAVEPK[HeavyK]	610.3548	793.454	y7
RXR	ILEAELAVEPK[HeavyK]	610.3548	864.4911	y8
RXR	ILEAELAVEPK[HeavyK]	610.3548	993.5337	y9
RXR	ILEAELAVEPK[HeavyK]	610.3548	1106.618	y10
RXR	GLSNPAEVEALR	628.3356	359.2396	y3
RXR	GLSNPAEVEALR	628.3356	488.2822	y4
RXR	GLSNPAEVEALR	628.3356	587.3506	y5
RXR	GLSNPAEVEALR	628.3356	716.3932	y6
RXR	GLSNPAEVEALR	628.3356	787.4303	y7
RXR	GLSNPAEVEALR	628.3356	884.483	y8
RXR	GLSNPAEVEALR	628.3356	998.526	y9
RXR	GLSNPAEVEALR	628.3356	1085.558	y10
RXR	GLSNPAEVEALR	628.3356	1198.642	y11
RXR	GLSNPAEVEALR[HeavyR]	633.3398	369.2479	y3
RXR	GLSNPAEVEALR[HeavyR]	633.3398	498.2904	y4
RXR	GLSNPAEVEALR[HeavyR]	633.3398	597.3589	y5
RXR	GLSNPAEVEALR[HeavyR]	633.3398	726.4014	y6
RXR	GLSNPAEVEALR[HeavyR]	633.3398	797.4385	y7
RXR	GLSNPAEVEALR[HeavyR]	633.3398	894.4913	y8
RXR	GLSNPAEVEALR[HeavyR]	633.3398	1008.534	y9
RXR	GLSNPAEVEALR[HeavyR]	633.3398	1095.566	y10
RXR	GLSNPAEVEALR[HeavyR]	633.3398	1208.65	y11

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
Pparg	HLYDSYIK	519.7663	423.2596	y3
Pparg	HLYDSYIK	519.7663	510.2917	y4
Pparg	HLYDSYIK	519.7663	625.3186	y5
Pparg	HLYDSYIK	519.7663	788.382	y6
Pparg	HLYDSYIK[HeavyK]	523.7734	431.2738	y3
Pparg	HLYDSYIK[HeavyK]	523.7734	518.3058	y4
Pparg	HLYDSYIK[HeavyK]	523.7734	633.3328	y5
Pparg	HLYDSYIK[HeavyK]	523.7734	796.3961	y6
Pparg	VEPASPPYYSEK	683.8298	363.1869	y3
Pparg	VEPASPPYYSEK	683.8298	526.2502	y4
Pparg	VEPASPPYYSEK	683.8298	689.3135	y5
Pparg	VEPASPPYYSEK	683.8298	786.3663	y6
Pparg	VEPASPPYYSEK	683.8298	883.4191	y7
Pparg	VEPASPPYYSEK	683.8298	970.4511	y8
Pparg	VEPASPPYYSEK	683.8298	1041.488	y9
Pparg	VEPASPPYYSEK	683.8298	1138.541	y10
Pparg	LNHPESSQLFAK	685.8567	365.2178	y3
Pparg	LNHPESSQLFAK	685.8567	478.3018	y4
Pparg	LNHPESSQLFAK	685.8567	606.3604	y5
Pparg	LNHPESSQLFAK	685.8567	693.3925	y6
Pparg	LNHPESSQLFAK	685.8567	780.4245	y7
Pparg	LNHPESSQLFAK	685.8567	909.4671	y8
Pparg	LNHPESSQLFAK	685.8567	1006.52	y9
Pparg	LNHPESSQLFAK	685.8567	1143.579	y10
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	371.2011	y3
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	534.2644	y4
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	697.3278	y5
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	794.3805	y6
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	891.4333	y7
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	978.4653	y8
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	1049.502	y9
Pparg	VEPASPPYYSEK[HeavyK]	687.8369	1146.555	y10
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	373.232	y3
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	486.316	y4
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	614.3746	y5
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	701.4067	y6
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	788.4387	y7
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	917.4813	y8
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	1014.534	y9
Pparg	LNHPESSQLFAK[HeavyK]	689.8638	1151.593	y10
Pparg	SVEAVQEITEYAK	733.8722	381.2127	y3
Pparg	SVEAVQEITEYAK	733.8722	510.2553	y4
Pparg	SVEAVQEITEYAK	733.8722	611.303	y5
Pparg	SVEAVQEITEYAK	733.8722	724.387	y6
Pparg	SVEAVQEITEYAK	733.8722	853.4296	y7
Pparg	SVEAVQEITEYAK	733.8722	981.4882	y8
Pparg	SVEAVQEITEYAK	733.8722	1080.557	y9
Pparg	SVEAVQEITEYAK	733.8722	1151.594	y10
Pparg	SVEAVQEITEYAK	733.8722	1280.636	y11
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	389.2269	y3
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	518.2695	y4
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	619.3171	y5
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	732.4012	y6
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	861.4438	y7
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	989.5024	y8
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	1088.571	y9
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	1159.608	y10
Pparg	SVEAVQEITEYAK[HeavyK]	737.8793	1288.651	y11

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
PIAS3	VSSIVAPGSSLR	586.8353	375.2345	y3
PIAS3	VSSIVAPGSSLR	586.8353	462.2665	y4
PIAS3	VSSIVAPGSSLR	586.8353	519.288	y5
PIAS3	VSSIVAPGSSLR	586.8353	616.3408	y6
PIAS3	VSSIVAPGSSLR	586.8353	687.3779	y7
PIAS3	VSSIVAPGSSLR	586.8353	786.4463	y8
PIAS3	VSSIVAPGSSLR	586.8353	899.5303	y9
PIAS3	VSSIVAPGSSLR	586.8353	986.5624	y10
PIAS3	VSSIVAPGSSLR	586.8353	1073.594	y11
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	385.2428	y3
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	472.2748	y4
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	529.2963	y5
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	626.349	y6
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	697.3861	y7
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	796.4545	y8
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	909.5386	y9
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	996.5706	y10
PIAS3	VSSIVAPGSSLR[HeavyR]	591.8394	1083.603	y11
PIAS3	VSELQVLLGFAGR	694.8984	303.177	y3
PIAS3	VSELQVLLGFAGR	694.8984	450.2454	y4
PIAS3	VSELQVLLGFAGR	694.8984	507.2668	y5
PIAS3	VSELQVLLGFAGR	694.8984	620.351	y6
PIAS3	VSELQVLLGFAGR	694.8984	733.435	y7
PIAS3	VSELQVLLGFAGR	694.8984	832.5034	y8
PIAS3	VSELQVLLGFAGR	694.8984	960.562	y9
PIAS3	VSELQVLLGFAGR	694.8984	1073.646	y10
PIAS3	VSELQVLLGFAGR	694.8984	1202.689	y11
PIAS3	VSELQVLLGFAGR	694.8984	1289.721	y12
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	313.1852	y3
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	460.2537	y4
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	517.2751	y5
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	630.3592	y6
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	743.4432	y7
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	842.5117	y8
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	970.5703	y9
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	1083.654	y10
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	1212.697	y11
PIAS3	VSELQVLLGFAGR[HeavyR]	699.9025	1299.729	y12

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012

Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
PIAS4	SYSVALYLVR	585.8295	387.2709	y3
PIAS4	SYSVALYLVR	585.8295	550.3342	y4
PIAS4	SYSVALYLVR	585.8295	663.4183	y5
PIAS4	SYSVALYLVR	585.8295	734.4554	y6
PIAS4	SYSVALYLVR	585.8295	833.5238	y7
PIAS4	SYSVALYLVR	585.8295	920.5558	y8
PIAS4	SYSVALYLVR	585.8295	1083.619	y9
PIAS4	SYSVALYLVR[HeavyR]	590.8336	397.2791	y3
PIAS4	SYSVALYLVR[HeavyR]	590.8336	560.3425	y4
PIAS4	SYSVALYLVR[HeavyR]	590.8336	673.4265	y5
PIAS4	SYSVALYLVR[HeavyR]	590.8336	744.4637	y6
PIAS4	SYSVALYLVR[HeavyR]	590.8336	843.532	y7
PIAS4	SYSVALYLVR[HeavyR]	590.8336	930.5641	y8
PIAS4	SYSVALYLVR[HeavyR]	590.8336	1093.627	y9
PIAS4	LDPDSEIATTGVR	687.3489	331.2083	y3
PIAS4	LDPDSEIATTGVR	687.3489	432.256	y4
PIAS4	LDPDSEIATTGVR	687.3489	533.3036	y5
PIAS4	LDPDSEIATTGVR	687.3489	604.3408	y6
PIAS4	LDPDSEIATTGVR	687.3489	717.4248	y7
PIAS4	LDPDSEIATTGVR	687.3489	846.4674	y8
PIAS4	LDPDSEIATTGVR	687.3489	933.4995	y9
PIAS4	LDPDSEIATTGVR	687.3489	1048.526	y10
PIAS4	LDPDSEIATTGVR	687.3489	1145.579	y11
PIAS4	LDPDSEIATTGVR	687.3489	1260.606	y12
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	341.2166	y3
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	442.2642	y4
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	543.3119	y5
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	614.349	y6
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	727.4331	y7
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	856.4757	y8
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	943.5077	y9
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	1058.535	y10
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	1155.587	y11
PIAS4	LDPDSEIATTGVR[HeavyR]	692.3531	1270.614	y12

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012
Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
RARG	GLGQPDLPK	462.761	357.2491	y3
RARG	GLGQPDLPK	462.761	472.276	y4
RARG	GLGQPDLPK	462.761	569.3288	y5
RARG	GLGQPDLPK	462.761	697.3874	y6
RARG	GLGQPDLPK	462.761	754.4088	y7
RARG	GLGQPDLPK	462.761	867.4929	y8
RARG	GLGQPDLPK[HeavyK]	466.7681	365.2633	y3
RARG	GLGQPDLPK[HeavyK]	466.7681	480.2902	y4
RARG	GLGQPDLPK[HeavyK]	466.7681	577.343	y5
RARG	GLGQPDLPK[HeavyK]	466.7681	705.4016	y6
RARG	GLGQPDLPK[HeavyK]	466.7681	762.423	y7
RARG	GLGQPDLPK[HeavyK]	466.7681	875.5071	y8
RARG	LQEPLLEALR	591.348	359.2396	y3
RARG	LQEPLLEALR	591.348	488.2822	y4
RARG	LQEPLLEALR	591.348	601.3662	y5
RARG	LQEPLLEALR	591.348	714.4503	y6
RARG	LQEPLLEALR	591.348	811.5031	y7
RARG	LQEPLLEALR	591.348	940.5457	y8
RARG	LQEPLLEALR	591.348	1068.604	y9
RARG	VQLDLGLWDK	593.8269	448.2185	y3
RARG	VQLDLGLWDK	593.8269	561.3026	y4
RARG	VQLDLGLWDK	593.8269	618.324	y5
RARG	VQLDLGLWDK	593.8269	731.4081	y6
RARG	VQLDLGLWDK	593.8269	846.4351	y7
RARG	VQLDLGLWDK	593.8269	959.5191	y8
RARG	VQLDLGLWDK	593.8269	1087.578	y9
RARG	LQEPLLEALR[HeavyR]	596.3521	369.2479	y3
RARG	LQEPLLEALR[HeavyR]	596.3521	498.2904	y4
RARG	LQEPLLEALR[HeavyR]	596.3521	611.3745	y5
RARG	LQEPLLEALR[HeavyR]	596.3521	724.4586	y6
RARG	LQEPLLEALR[HeavyR]	596.3521	821.5114	y7
RARG	LQEPLLEALR[HeavyR]	596.3521	950.554	y8
RARG	LQEPLLEALR[HeavyR]	596.3521	1078.613	y9
RARG	VQLDLGLWDK[HeavyK]	597.834	456.2327	y3
RARG	VQLDLGLWDK[HeavyK]	597.834	569.3168	y4
RARG	VQLDLGLWDK[HeavyK]	597.834	626.3383	y5
RARG	VQLDLGLWDK[HeavyK]	597.834	739.4223	y6
RARG	VQLDLGLWDK[HeavyK]	597.834	854.4492	y7
RARG	VQLDLGLWDK[HeavyK]	597.834	967.5333	y8
RARG	VQLDLGLWDK[HeavyK]	597.834	1095.592	y9

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012
Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass m/z [M+2H]2+	Product ion mass m/z [M+H]+	Transition y-ion
FOSL2	GTGSAVGPVVVK	535.8138	345.2491	y3
FOSL2	GTGSAVGPVVVK	535.8138	444.3175	y4
FOSL2	GTGSAVGPVVVK	535.8138	541.3702	y5
FOSL2	GTGSAVGPVVVK	535.8138	598.3917	y6
FOSL2	GTGSAVGPVVVK	535.8138	697.4601	y7
FOSL2	GTGSAVGPVVVK	535.8138	768.4973	y8
FOSL2	GTGSAVGPVVVK	535.8138	855.5293	y9
FOSL2	GTGSAVGPVVVK	535.8138	912.5507	y10
FOSL2	GTGSAVGPVVVK	535.8138	1013.598	y11
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	353.2633	y3
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	452.3317	y4
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	549.3845	y5
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	606.4059	y6
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	705.4744	y7
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	776.5115	y8
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	863.5435	y9
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	920.5649	y10
FOSL2	GTGSAVGPVVVK[HeavyK]	539.8209	1021.613	y11
FOSL2	LQAETEELEEEK	724.3435	405.1974	y3
FOSL2	LQAETEELEEEK	724.3435	534.2401	y4
FOSL2	LQAETEELEEEK	724.3435	647.3241	y5
FOSL2	LQAETEELEEEK	724.3435	776.3667	y6
FOSL2	LQAETEELEEEK	724.3435	905.4093	y7
FOSL2	LQAETEELEEEK	724.3435	1006.457	y8
FOSL2	LQAETEELEEEK	724.3435	1135.5	y9
FOSL2	LQAETEELEEEK	724.3435	1206.537	y10
FOSL2	LQAETEELEEEK	724.3435	1334.595	y11
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	413.2116	y3
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	542.2542	y4
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	655.3383	y5
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	784.3809	y6
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	913.4235	y7
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	1014.471	y8
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	1143.514	y9
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	1214.551	y10
FOSL2	LQAETEELEEEK[HeavyK]	728.3506	1342.609	y11

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012
Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
Smad2	GWGAEYR	419.6957	338.1817	y2
Smad2	GWGAEYR	419.6957	467.2243	y3
Smad2	GWGAEYR	419.6957	538.2614	y4
Smad2	GWGAEYR	419.6957	595.2829	y5
Smad2	GWGAEYR	419.6957	781.3622	y6
Smad2	GWGAEYR[HeavyR]	424.6998	348.19	y2
Smad2	GWGAEYR[HeavyR]	424.6998	477.2326	y3
Smad2	GWGAEYR[HeavyR]	424.6998	548.2697	y4
Smad2	GWGAEYR[HeavyR]	424.6998	605.2912	y5
Smad2	GWGAEYR[HeavyR]	424.6998	791.3705	y6
Smad2	VETPVLPPVLVPR	708.4346	371.2396	y3
Smad2	VETPVLPPVLVPR	708.4346	583.392	y5
Smad2	VETPVLPPVLVPR	708.4346	680.4448	y6
Smad2	VETPVLPPVLVPR	708.4346	777.4976	y7
Smad2	VETPVLPPVLVPR	708.4346	890.5817	y8
Smad2	VETPVLPPVLVPR	708.4346	989.6501	y9
Smad2	VETPVLPPVLVPR	708.4346	1086.703	y10
Smad2	VETPVLPPVLVPR	708.4346	1187.75	y11
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	381.2479	y3
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	494.3319	y4
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	593.4003	y5
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	690.4531	y6
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	787.5059	y7
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	900.5899	y8
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	999.6583	y9
Smad2	VETPVLPPVLVPR[HeavyR]	713.4388	1096.711	y10

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012
Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
Nr2C1	AYVEFQDYITR	702.8433	389.2502	y3
Nr2C1	AYVEFQDYITR	702.8433	552.3135	y4
Nr2C1	AYVEFQDYITR	702.8433	667.3404	y5
Nr2C1	AYVEFQDYITR	702.8433	795.399	y6
Nr2C1	AYVEFQDYITR	702.8433	942.4674	y7
Nr2C1	AYVEFQDYITR	702.8433	1071.51	y8
Nr2C1	AYVEFQDYITR	702.8433	1170.578	y9
Nr2C1	AYVEFQDYITR	702.8433	1333.642	y10
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	399.2584	y3
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	562.3217	y4
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	677.3487	y5
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	805.4073	y6
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	952.4757	y7
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	1081.518	y8
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	1180.587	y9
Nr2C1	AYVEFQDYITR[HeavyR]	707.8474	1343.65	y10
Nr2C1	SPLAATPTFVTDSETAR	882.4441	347.2032	y3
Nr2C1	SPLAATPTFVTDSETAR	882.4441	476.2458	y4
Nr2C1	SPLAATPTFVTDSETAR	882.4441	563.2778	y5
Nr2C1	SPLAATPTFVTDSETAR	882.4441	678.3047	y6
Nr2C1	SPLAATPTFVTDSETAR	882.4441	779.3524	y7
Nr2C1	SPLAATPTFVTDSETAR	882.4441	878.4208	y8
Nr2C1	SPLAATPTFVTDSETAR	882.4441	1223.59	y11
Nr2C1	SPLAATPTFVTDSETAR	882.4441	1395.675	y13
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	573.2861	y5
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	789.3607	y7
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	888.4291	y8
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	1136.545	y10
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	1233.598	y11
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	1334.646	y12
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	1405.683	y13
Nr2C1	SPLAATPTFVTDSETAR[HeavyR]	887.4482	1476.72	y14

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012
Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass m/z [M+2H]2+	Product ion mass m/z [M+H]+	Transition y-ion
Nfib	EDFVLTVTGK	554.7978	305.1814	y3
Nfib	EDFVLTVTGK	554.7978	404.2498	y4
Nfib	EDFVLTVTGK	554.7978	505.2975	y5
Nfib	EDFVLTVTGK	554.7978	618.3815	y6
Nfib	EDFVLTVTGK	554.7978	717.45	y7
Nfib	EDFVLTVTGK	554.7978	864.5184	y8
Nfib	EDFVLTVTGK	554.7978	979.5453	y9
Nfib	EDFVLTVTGK[HeavyK]	558.8049	313.1956	y3
Nfib	EDFVLTVTGK[HeavyK]	558.8049	412.264	y4
Nfib	EDFVLTVTGK[HeavyK]	558.8049	513.3117	y5
Nfib	EDFVLTVTGK[HeavyK]	558.8049	626.3958	y6
Nfib	EDFVLTVTGK[HeavyK]	558.8049	725.4642	y7
Nfib	EDFVLTVTGK[HeavyK]	558.8049	872.5326	y8
Nfib	EDFVLTVTGK[HeavyK]	558.8049	987.5595	y9
Nfib	GIPLESTDGER	587.2909	304.161	y2
Nfib	GIPLESTDGER	587.2909	361.1825	y3
Nfib	GIPLESTDGER	587.2909	476.2094	y4
Nfib	GIPLESTDGER	587.2909	577.2571	y5
Nfib	GIPLESTDGER	587.2909	664.2891	y6
Nfib	GIPLESTDGER	587.2909	793.3317	y7
Nfib	GIPLESTDGER	587.2909	906.4158	y8
Nfib	GIPLESTDGER	587.2909	1003.469	y9
Nfib	GIPLESTDGER[HeavyR]	592.295	314.1693	y2
Nfib	GIPLESTDGER[HeavyR]	592.295	371.1907	y3
Nfib	GIPLESTDGER[HeavyR]	592.295	486.2177	y4
Nfib	GIPLESTDGER[HeavyR]	592.295	587.2654	y5
Nfib	GIPLESTDGER[HeavyR]	592.295	674.2974	y6
Nfib	GIPLESTDGER[HeavyR]	592.295	803.34	y7
Nfib	GIPLESTDGER[HeavyR]	592.295	916.424	y8
Nfib	GIPLESTDGER[HeavyR]	592.295	1013.477	y9

Decoding adipogenic gene regulatory mechanisms using targeted quantitative proteomics. Simicevic et al. 2012
Supplementary Table 5 - 10TF transitions monitored

Transcription Factor	Sequence	Parent ion mass	Product ion mass	Transition
		m/z [M+2H]2+	m/z [M+H]+	y-ion
ARID	GGVSSIGTNTTTGSR	697.8471	319.1719	y3
ARID	GGVSSIGTNTTTGSR	697.8471	420.2196	y4
ARID	GGVSSIGTNTTTGSR	697.8471	521.2673	y5
ARID	GGVSSIGTNTTTGSR	697.8471	622.3149	y6
ARID	GGVSSIGTNTTTGSR	697.8471	736.3578	y7
ARID	GGVSSIGTNTTTGSR	697.8471	837.4055	y8
ARID	GGVSSIGTNTTTGSR	697.8471	894.427	y9
ARID	GGVSSIGTNTTTGSR	697.8471	1007.511	y10
ARID	GGVSSIGTNTTTGSR	697.8471	1094.543	y11
ARID	GGVSSIGTNTTTGSR	697.8471	1181.575	y12
ARID	GGVSSIGTNTTTGSR	697.8471	1280.644	y13
ARID	GGVSSIGTNTTTGSR	697.8471	1337.665	y14
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	329.1802	y3
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	430.2278	y4
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	531.2755	y5
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	632.3232	y6
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	746.3661	y7
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	847.4138	y8
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	904.4352	y9
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	1017.519	y10
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	1104.551	y11
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	1191.583	y12
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	1290.652	y13
ARID	GGVSSIGTNTTTGSR[HeavyR]	702.8512	1347.673	y14
ARID	GLNLPTSITSAAFTLR	831.4646	389.2502	y3
ARID	GLNLPTSITSAAFTLR	831.4646	536.3185	y4
ARID	GLNLPTSITSAAFTLR	831.4646	607.3557	y5
ARID	GLNLPTSITSAAFTLR	831.4646	678.3928	y6
ARID	GLNLPTSITSAAFTLR	831.4646	765.4248	y7
ARID	GLNLPTSITSAAFTLR	831.4646	866.4725	y8
ARID	GLNLPTSITSAAFTLR	831.4646	979.5566	y9
ARID	GLNLPTSITSAAFTLR	831.4646	1066.589	y10
ARID	GLNLPTSITSAAFTLR	831.4646	1167.636	y11
ARID	GLNLPTSITSAAFTLR	831.4646	1264.689	y12
ARID	GLNLPTSITSAAFTLR	831.4646	1377.773	y13
ARID	GLNLPTSITSAAFTLR	831.4646	1491.816	y14
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	399.2584	y3
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	546.3268	y4
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	617.364	y5
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	688.4011	y6
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	775.4331	y7
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	876.4808	y8
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	989.5648	y9
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	1076.597	y10
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	1177.645	y11
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	1274.697	y12
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	1387.781	y13
ARID	GLNLPTSITSAAFTLR[HeavyR]	836.4688	1501.824	y14

Supplementary Note
Absolute copy number analysis of transcription factors during cellular differentiation using a multiplex, targeted proteomics approach

1 Statistical Modeling of protein binding

1.1 Statistical Mechanics Approach

The first question we want to address is how many protein molecules on average are bound to specific sites in a given cell. The approach used to compute the occupancy of the different specific sites is based on statistical mechanics. The main idea consists in enumerating all the different configurations by which one can place N proteins on the accessible genome weighted by the Boltzmann factor $\exp(-\beta E)$, which determines the likelihood of a configuration depending on its energy E. Typically, this is achieved by means of the partition function which describes the statistical properties of the system at thermodynamical equilibrium.

For the sake of simplicity, we made the following assumptions regarding our system: 1) we considered only one species of protein, 2) all the proteins are assumed to be on the DNA [1, 2] (either at specific sites or non-specific sites), 3) we do not consider hindrance between possibly (but very unlikely) overlapping sites, 4) we model k categories of specific sites of different affinity and all of these sites are assumed to be stronger than non-specific sites. Consequently, the system can be parametrized in the following way. N is the total number of proteins in the nucleus, M the size of the accessible genome, m_i the number of specific sites of category i, n_i the number of proteins bound to sites of category i and the energy E_i associated with each category of sites i. Therefore, the partition function Z of the system can be written as:

$$Z(N) = \sum_{n_1,n_2,\ldots,n_k=0}^{N} \binom{M-m}{N-n} \exp\left(-\beta(N-n)E_0\right) \left\{ \prod_{i=1}^{k} \binom{m_i}{n_i} \exp\left(-\beta n_i E_i\right) \right\}$$

with $m = \sum_{i=1}^{k} m_i$ the total number of specific sites, $n = \sum_{i=1}^{k} n_i$ the number of proteins bound to specific sites and E_0 the energy of an non-specific site. We can compute the average number of proteins $\bar{n}_i$ which are bound to specific sites of category i, with the derivative of the log of the partition function with respect to the energy E_i:

$$\bar{n}_i = -\frac{1}{\beta}\frac{\partial}{\partial E_i}\log Z(N)$$

It follows that the mean occupancy p_i of a site i is given by:

$$p_i = \frac{\bar{n}_i}{m_i}$$

In the following, it is convenient to rewrite the partition function in term of the affinities $X_i = \exp\left(-\beta(E_i - E_0)\right)$ of the different sites where the reference energy is chosen as the energy of the non-specific sites E_0. Consequently, the affinity of a non-specific site is $X_0 = 1$ and the partition function is now given by:

$$Z(N) = \sum_{n_1,n_2,...,n_k=0}^{N} \binom{M-m}{N-n}\left\{\prod_{i=1}^{k}\binom{m_i}{n_i}X_i^{n_i}\right\} \tag{1}$$

It is possible to compute an approximate expression for the average number of occupied specific sites $\bar{n}_i$ and their occupancy p_i in two different regimes: either the number of proteins is in excess compared to number of specific sites, or the number of specific sites is much larger than the number of proteins.

1.2 Regime 1: Excess of Proteins over Specific Sites

In this first regime, we assume that the number of proteins N is much larger than the total number of specific sites m but that N is still small compared to the size of the accessible genome M, namely:

$$M \gg N \gg m = \sum_{i=1}^{k} m_i \quad \text{with} \quad m_i \geq n_i \quad \forall i \in \{1, 2, ..., k\}$$

The average number of occupied sites and the mean occupancy of a site can be derived rigorously from the partition function by approximating the binomial coefficients. Here, we present a more intuitive approach leading to the same results. By setting $m_i = 1$ and $m_j = 0\,\forall j \neq i$ in the partition function (1), we can compute the probability $P(b|N, m = 1)$ that a specific site i is bound given that we have only one specific site in the genome and N proteins. We find that:

$$P(b|N, m = 1) = \frac{\frac{N}{M}X_i}{\frac{M-N}{M} + \frac{N}{M}X_i} \simeq \frac{\frac{N}{M}X_i}{1 + \frac{N}{M}X_i}$$

which is the famous Hill function where the occupation of the site depends on the concentration of proteins N/M and the affinity of the site X_i. Since there are many proteins compared to the number of specific sites, each site can be seen as independent from others, namely we neglect the depletion of the protein pool due to binding at specific sites. Therefore, the average number of specific sites which are occupied $\bar{n}_i$ is given by $P(b|N, m = 1)$ multiplied by the number of sites m_i:

$$\bar{n}_i = m_i\frac{\frac{N}{M}X_i}{1 + \frac{N}{M}X_i}$$

The mean occupancy $p_i = \bar{n}_i/m_i$ of a site of category i is then identical to $P(b|N, m = 1)$:

$$p_i = \frac{\frac{N}{M}X_i}{1 + \frac{N}{M}X_i}$$

1.3 Regime 2: Excess of Specific Sites over Proteins

In the second regime, we assume that the number of specific sites m_i and the number of non-specific sites $m_0 = M - m$ are much larger than the number of proteins N. This can be stated as:

$$M = \sum_{i=0}^{k} m_i \quad \text{with} \quad m_i \gg N \geq n_i \quad \forall i \in \{0, 1, ..., k\}$$

From the partition function (1), one can compute the probability $P(b|N = 1, \{m_i\})$ that one protein is bound to a specific site given the set of specific sites $\{m_i\}$ and that the total number of proteins is one, this probability is given by:

$$P(b|N = 1, \{m_i\}) = \frac{\frac{m_i}{M}X_i}{\frac{m_0}{M} + \sum_{i=1}^{k} \frac{m_i}{M}X_i}$$

The probability that one protein is bound to a specific site now depends on the concentrations of sites m_i/M and the affinities X_i of the different sites. Since the number of sites is much larger than the number of proteins, each protein can be considered as being independent from each other. The average number of proteins $\bar{n}_i$ which are bound to specific site i is then given by $P(b|N = 1, \{m_i\})$ times the number of proteins N:

$$\bar{n}_i = N\frac{\frac{m_i}{M}X_i}{\frac{m_0}{M} + \sum_{i=1}^{k} \frac{m_i}{M}X_i}$$

Thus, the expression for the mean occupancy $p_i = \bar{n}_i/m_i$ of a site of category i is similar that in the first regime except that the denominator now depends on the concentrations of sites m_i/M:

$$p_i = \frac{\frac{N}{M}X_i}{\frac{m_0}{M} + \sum_{i=1}^{k} \frac{m_i}{M}X_i}$$

In the above expression, since $X_0 = 1$, the denominator can be expressed in term of the average affinity of the sites $\bar{X}$ which depends on the distribution of affinities in the accessible genome $P(X_i) = m_i/M$, therefore we obtain:

$$p_i = \frac{\frac{N}{M}X_i}{\sum_{i=0}^{k} \frac{m_i}{M}X_i} = \frac{N}{M\bar{X}}X_i$$

1.4 Typical Parameters

The size of the accessible mice genome is typically less than 10% of the full genome, consequently $M \sim 10^8$ bp. Regarding the number of proteins, the measured number for $\mathrm{PPAR}\gamma$ and $\mathrm{RXR}\alpha$ were approximately in the range of $N \sim 10^3$ proteins. The number of binding sites which were detected is approximately $m \sim 10^3$ sites [3], it is important to realize that those represent the stronger sites, weaker sites are most likely not detected. Finally, the dissociation constant K_d is typically in the order of $\mu\mathrm{M}$ for a non-specific site and nM for a strong site, consequently we can assume that a strong site has an affinity roughly 10^3 times larger than a non-specific one [2].

Given those numbers, the first regime (Section 1.2) where the proteins are assumed to be in excess is clearly not adequate, since the number of detected binding sites is in the range of the number of proteins 10^3. Despite the fact that in our case the strongest sites might be fewer than the number of proteins N, most of the sites will actually be in large excess compared to N, therefore the second regime (Section 1.3) is appropriate to describe the occupancies of the different sites in our ChIP-seq experiments. In fact, it also predicts very well the behavior of the strongest sites (cf. Supplementary Fig. 15), because the strongest sites are not saturated due to the competition with the very large number of less favorable sites. Consequently, we expect a linear relationship between the occupancy and the affinity of the sites.

$$p_i = \frac{N}{M\bar{X}} X_i \tag{2}$$

If the non-specific sites dominate, namely the affinities X_i decrease fast enough compared to the increase in number of sites m_i, then the mean affinity of the sites will be close to one $\bar{X} \simeq 1$. This might not be true in practice, we will see later how to estimate the coefficient $\bar{X}$ (Section 2.4).

2 Link between ChIP-seq signal and affinity

2.1 Simple ChIP-seq Model

In a ChIP-seq experiment, the signal, i.e. the average number of fragments S_i for any sites of type i, should reflect the mean occupancy of the sites p_i. Of course, the occupancy is not the only contribution to the signal, the efficiency and specificity of the antibody as well as the number of cells play an important role, but those effects can be implicitly included in a simple model:

$$S_i = Ap_i + B$$

where as a first approximation the signal S_i is proportional to the occupancy p_i plus some background B [4]. Since in our case the occupancy goes as $p_i \simeq \frac{N}{M\bar{X}} X_i$, we expect a correlation between the affinity X_i and the mean signal S_i:

$$S_i = \frac{AN}{M\bar{X}} X_i + B \tag{3}$$

2.2 Link between Energy and PWM Log-likelihood

In practice, one does not necessarily know the binding energy of the sites of interest, but if we know the position weight matrix (PWM), it is then possible to make the link between each sequence and its binding energy [4, 5]. In the PWM framework the binding energy of the site $\varepsilon = -\beta E$ is assumed to be the sum of independent contributions of each base $\varepsilon_k(b)$. If the binding sites are characterized by an average energy $\bar{\varepsilon}$, one can express the likelihood $q(s)$ that a randomly chosen binding sites of length L will have sequence s:

$$q(s) = \prod_{k=1}^{L} \frac{\exp\left(\lambda \varepsilon_k(s_k)\right)}{\sum_b \exp\left(\lambda \varepsilon_k(b)\right)} = \prod_{k=1}^{L} f_k(s_k)$$

where the selection parameter λ ensures that $\bar{\varepsilon} = \sum_s \varepsilon(s) q(s)$ and $f_k(b)$ is the frequency of observation of base b at position k. Therefore, the log-likelihood $z(s) = \log\left(q(s)\right)$ is related to energy by the factor λ plus a constant:

$$z = \log\left(q\right) = \lambda \varepsilon + \text{cst}$$

We can express the energy of the sites with respect to the consensus sequence which is assumed to be the best sequence with the largest log-likelihood $z_{max} = \log\left(q_{max}\right)$:

$$-\beta(E_i - E_c) = \frac{1}{\lambda}(z_i - z_{max})$$

where E_c is the energy of the consensus. Finally the affinities X_i in term of the log-likelihood are given by:

$$X_i = X_{max} \exp\left(\frac{1}{\lambda}(z_i - z_{max})\right) \tag{4}$$

where $X_{max} = \exp\left(-\beta(E_c - E_0)\right)$ is the affinity of the consensus sequence. In the following we will assume that $X_{max} = 10^3$ which is the typical magnitude for the strongest sites compared to non-specific ones [2].

2.3 Estimating the Parameter λ

Using our simple model for the ChIP-seq signal, we can now relate the log-likelihood score z_i of the different sites with the signal. Indeed, from equation (3) and (4), the signal is given by:

$$S_i = A' \exp\left(\frac{1}{\lambda} z_i\right) + B \tag{5}$$

where A' is a new proportionality constant. Therefore, it is possible to estimate the λ parameter of the $\mathrm{PPAR}\gamma$ motif from the ChIP-seq signal. This enables us to make the link between the energy scale and the log-likelihood score.

We focused on $\mathrm{PPAR}\gamma$ at day 0, day 2 and day 6, we used the motif from the JASPAR CORE database [6] and we screened the accessible genome with FIMO [7] from the MEME suite for all the sequences which had a p-value $< 10^{-3}$. Due to limited sequencing depth of available DHS data, we used the H3K27ac regions from Tarjei et al. [8] as a proxy for the accessible genome. We verified that these regions cover most of the reported $\mathrm{PPAR}\gamma$ sites. The size of these regions were similar between the three time points, $M_0 = 5.97 \cdot 10^7$ bp, $M_2 = 6.99 \cdot 10^7$ bp and $M_6 = 6.76 \cdot 10^7$ bp. For each of those three time points, we estimated the parameter λ using the mean ChIP-seq signal for different category of sites, namely we separated the sequences in 6 bins of log-likelihood covering the range obtained with FIMO (cf. Supplementary Fig. 16). We estimated the background level for the three time points by taking a window of 200 bp shifted by 1000 bp from each peak, we obtained $B_0 = 1.273$, $B_2 = 1.146$ and $B_6 = 2.004$. Since there is not much ChIP signal at day 0 (very few $\mathrm{PPAR}\gamma$ proteins before induction), we estimated the λ parameter as the average of day 2 and 6, we obtained $\lambda \simeq 3.36$.

2.4 Estimating $\bar{X}$

In the second regime, where the specific sites are in excess compared to proteins, we showed in Section 1.3 that the occupancy can be expressed as follows:

$$p_i = \frac{\frac{N}{M} X_i}{\sum_{i=0}^{k} \frac{m_i}{M} X_i} = \frac{N}{M\bar{X}} X_i$$

where $\bar{X}$ is actually the average affinity of the sites $\bar{X} = \sum_{i=0}^{k} X_i P(X_i)$. We estimated $\bar{X}$ using the theoretical distribution of sites with respect to the log-likelihood $\rho(z)$ given by the PWM of $\mathrm{PPAR}\gamma$:

$$\bar{X} = \int_{z_{min}}^{z_{max}} X(z)\rho(z)\mathrm{d}z \tag{6}$$

where $X(z)$ is given by equation (4) above the non-specific threshold $z^\star = z_{max} - \lambda \log(X_{max})$ which corresponds to the limit where non-specific binding starts to dominate. Therefore, $X(z)$ can be expressed as:

$$X(z) = \begin{cases} X_{max} \exp\left(\frac{1}{\lambda}(z - z_{max})\right) & z > z^\star \\ X_0 = 1 & z \leq z^\star \end{cases} \tag{7}$$

We approximated the probability density distribution $\rho(z)$ as a normal distribution $\mathcal{N}(z; \bar{z}, \sigma_z)$. We sampled the distribution from the PWM in order to estimate the mean $\bar{z} = -32.69$ and the standard deviation $\sigma_z = 5.33$. Performing the integral (6), we obtained $\bar{X} \simeq 1.93$. A correction to the normal distribution based on the saddle-point approximation [9] did not change much the result ($\bar{X} \simeq 1.97$).

3 Predicting the number of binding sites

3.1 Occupancy Threshold

In order to predict the number of binding sites that should be detected in our ChIP-seq data, we assume that there is a certain occupancy threshold p_t above which we will start detecting the sites. We can now express this threshold in term of the log-likelihood. Indeed, from equation (2) and (4) the occupancy can be expressed as :

$$p_t = \frac{N X_{max}}{M\bar{X}} \exp\left(\frac{1}{\lambda}(z_t - z_{max})\right) \tag{8}$$

Inverting this relation gives the log-likelihood threshold z_t:

$$z_t = \lambda \log \left(\frac{M \bar{X} p_t}{N X_{max}} \right) + z_{max} \tag{9}$$

Given this threshold z_t and the density of sites with respect to the log-likelihood, we will be able to predict the number of binding sites $N_{sites}(z_t)$, by counting how many sites in the accessible genome (section 2.3) have a log-likelihood larger than z_t.

3.2 Number of Binding Sites

We used the same approach than the previous section 2.3 to screen the accessible genome in order to model the cumulative number of sites $N_{sites}(z)$ with respect to the log-likelihood z. In order to obtain a smooth representation, it is convenient to parametrize the tail of $N_{sites}(z)$ as a power-law:

$$N_{sites}(z) = C(z_0 - z)^k \tag{10}$$

where k is the power law exponent, C and z_0 are constants. We estimated the parameters C, k and z_0 for $\mathrm{PPAR}\gamma$ at day 0, day 2 and day 6 using non-linear least squares, this representation gives excellent results (cf. Supplementary Fig. 17). Knowing the log-likelihood threshold z_t above which we should start to detect sites, we can predict the number of sites by evaluating (10) in z_t:

$$N_{sites}(z_t) = C(z_0 - z_t)^k \quad \text{with} \quad z_t = \lambda \log \left(\frac{M \bar{X} p_t}{N X_{max}} \right) + z_{max}$$

We determined the log-likelihood threshold z_t for the three different time points using equation (9) and the measured proteins copy number $N_0 = 2376, N_2 = 6095, N_6 = 7462$ which were average over the three biological replicates. The occupancy threshold p_t is a free parameter which was assumed to be the same between each time points, it was chosen so that the squared error between the detected number of sites and the predicted one is minimized, we obtained $p_t \simeq 1.35\%$ (cf. Supplementary Fig. 18) which is comparable to % input values in a good ChIP experiment [10]. We remind the reader that % input would correspond to occupancy in the case of an ideal ChIP (100% efficient antibody).

3.3 Independent Validation

We tested our model with a different set of ChIP-seq data provided by Siersbaek et al. [11] although they only performed ChIP-seq on $\mathrm{PPAR}\gamma$ at day 2 and 6 after induction. Following a similar approach than in section 2.3 and using the same DNA accessibility marks, we fitted the selection parameter and obtained $\lambda \simeq 2.69$. Using the measured copy number as in section 3.2 and assuming no detected $\mathrm{PPAR}\gamma$ binding site at day 0, we then minimized the squared error between the detected and predicted number of binding sites for the three time points, and we obtained an occupancy threshold of $p_t \simeq 0.98\%$ (cf. Supplementary Fig. 18) which is in the same range as the value we obtained in the previous section (section 3.2). Despite the new selection parameter, the prediction for both data set are in good agreement. If we average both the selection parameter and the number of detected binding sites from [3] and [11], the model is then in excellent agreement with the average detected binding sites (cf. Supplementary Fig. 18).

3.4 Discussion

Our model makes a few assumptions that merit discussions. First, since we did not know the affinity of the best sites *in vivo*, we assumed that the strongest sites, namely sites corresponding to the consensus sequence, were 1000 fold stronger than non-specific sites ($X_{max} = 10^3$), which is realistic [2]. In addition, we computed how the predicted occupancies p_i depend on this number trough the ratio $X_i/\bar{X}$ (cf. Supplementary Fig. 19). We observe two regimes: one when $X_{max} > 10^3$ where $X_i/\bar{X}$ becomes insensitive to X_{max}. Secondly when $X_{max} < 10^3$ (which is unlikely biologically), the predicted occupancies would decrease, but this could be compensated by lowering the occupancy threshold p_t. Thus, our conclusions on the behavior of the number of sites vs proteins number do not strongly depend on the presumed value of X_{max}.

As a second assumption, we did not model explicitly the dimer between $\mathrm{PPAR}\gamma$ and $\mathrm{RXR}\alpha$, which is justified since $\mathrm{PPAR}\gamma$ is in limiting amounts. Moreover, we neglected the hindrance due to other protein species which could distort the predicted occupancies. Indeed, many proteins are interacting with DNA and thereby obstruct the accessible sites [12]. In particular, this might reduce the size of the genomic regions that are effectively accessible for $\mathrm{PPAR}\gamma$ and consequently the occupancy of the sites might be higher. Nevertheless, the values derived for the occupancy (2) will not be strongly affected since the product $M\bar{X}$ would only be mildly changed if we assume that additional proteins would essentially deplete non-specific or weak $\mathrm{PPAR}\gamma$ binding sites. Indeed, if we assume that one third of the transcription factors of the mouse (~ 850 TFs) are expressed at an average level of 10^4 copies [13], we would obtain $N_{TF} \simeq 8.5 \cdot 10^6$ molecules in the nucleus which is still small compared to the measured accessible genome size $M \simeq 6.8 \cdot 10^7$.

References

[1] Y Kao-Huang, A Revzin, A P Butler, P O'Conner, D W Noble, and P H von Hippel. Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 74(10):4228–4232, October 1977.

[2] Mark D Biggin. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Developmental Cell*, 21(4):611–626, October 2011.

[3] R Nielsen, T A Pedersen, D Hagenbeek, P Moulos, R Siersbaek, E Megens, S Denissov, M Borgesen, K J Francoijs, S Mandrup, and H G Stunnenberg. Genome-wide profiling of PPARγ:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes & Development*, 22(21):2953–2967, November 2008.

[4] B C Foat, A V Morozov, and H J Bussemaker. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics*, 22(14):e141–e149, July 2006.

[5] O G Berg and P H von Hippel. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *Journal of molecular biology*, 193(4):723–750, February 1987.

[6] J C Bryne, E Valen, M H E Tang, T Marstrand, O Winther, I da Piedade, A Krogh, B Lenhard, and A Sandelin. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Research*, 36(Database):D102–D106, December 2007.

[7] C E Grant, T L Bailey, and W S Noble. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018, March 2011.

[8] Tarjei S Mikkelsen, Zhao Xu, Xiaolan Zhang, Li Wang, Jeffrey M Gimble, Eric S Lander, and Evan D Rosen. Comparative Epigenomic Analysis of Murine and Human Adipogenesis. *Cell*, 143(1):156–169, October 2010.

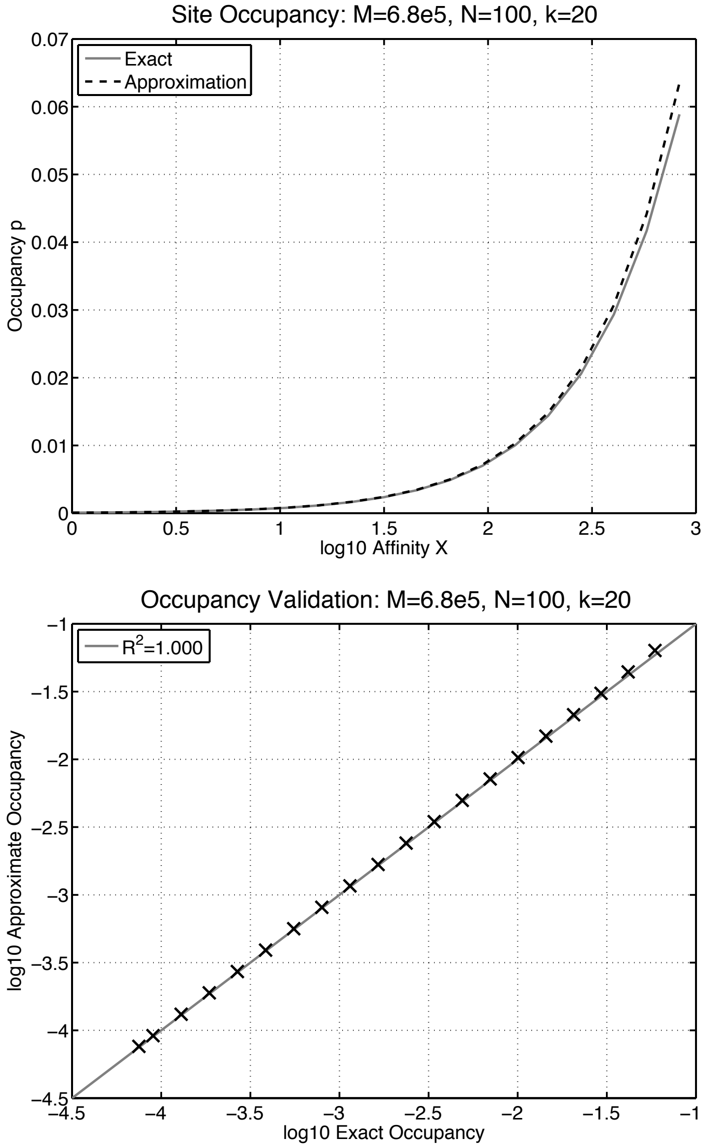
[9] M Djordjevic. A Biophysical Approach to Transcription Factor Binding Site Discovery. *Genome Research*, 13(11):2381–2390, November 2003.

[10] Guillaume Rey, François Cesbron, Jacques Rougemont, Hans Reinke, Michael Brunner, and Felix Naef. Genome-Wide and Phase-Specific DNA-Binding Rhythms of BMAL1 Control Circadian Output Functions in Mouse Liver. *PLoS Biology*, 9(2):e1000595, February 2011.

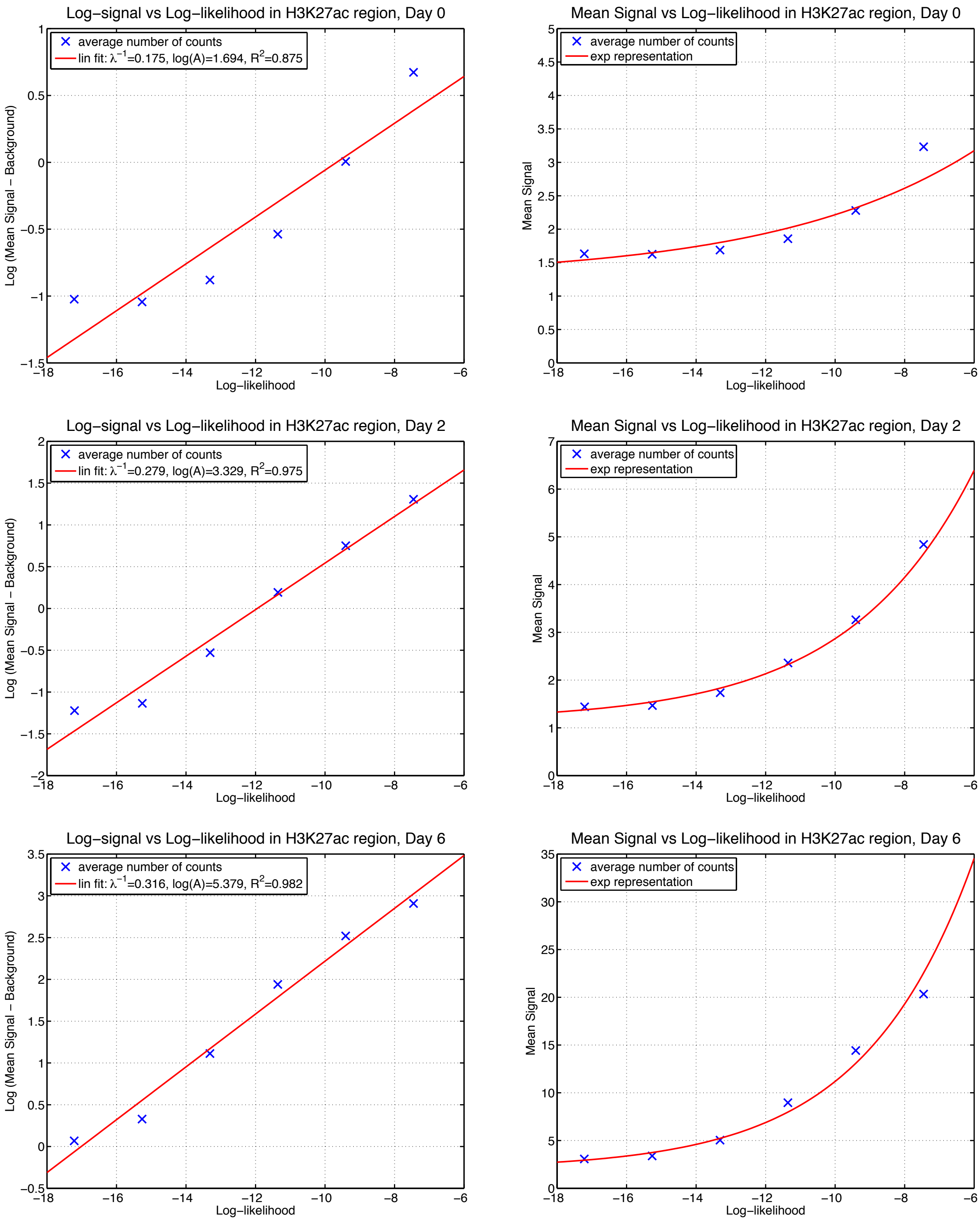
[11] Rasmus Siersbaek, Ronni Nielsen, Sam John, Myong-Hee Sung, Songjoon Baek, Anne Loft, Gordon L Hager, and Susanne Mandrup. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *The EMBO Journal*, 30(8):1459–1472, April 2011.

[12] Marco J Morelli, Rosalind J Allen, and Pieter Rein ten Wolde. Effects of Macromolecular Crowding on Genetic Networks. *Biophysj*, 101(12):2882–2891, December 2011.

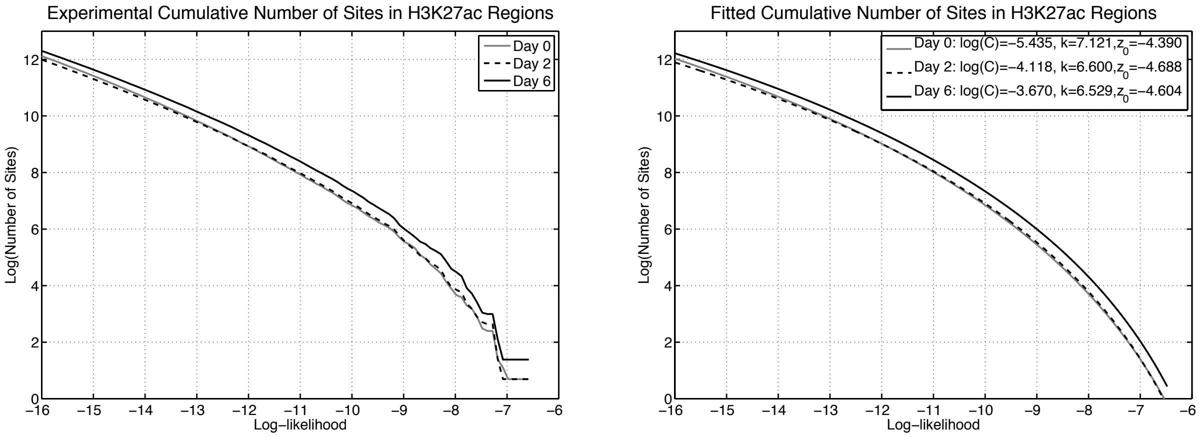
[13] S K Kummerfeld. DBD: a transcription factor prediction database. *Nucleic Acids Research*, 34(90001):D74–D81, January 2006.



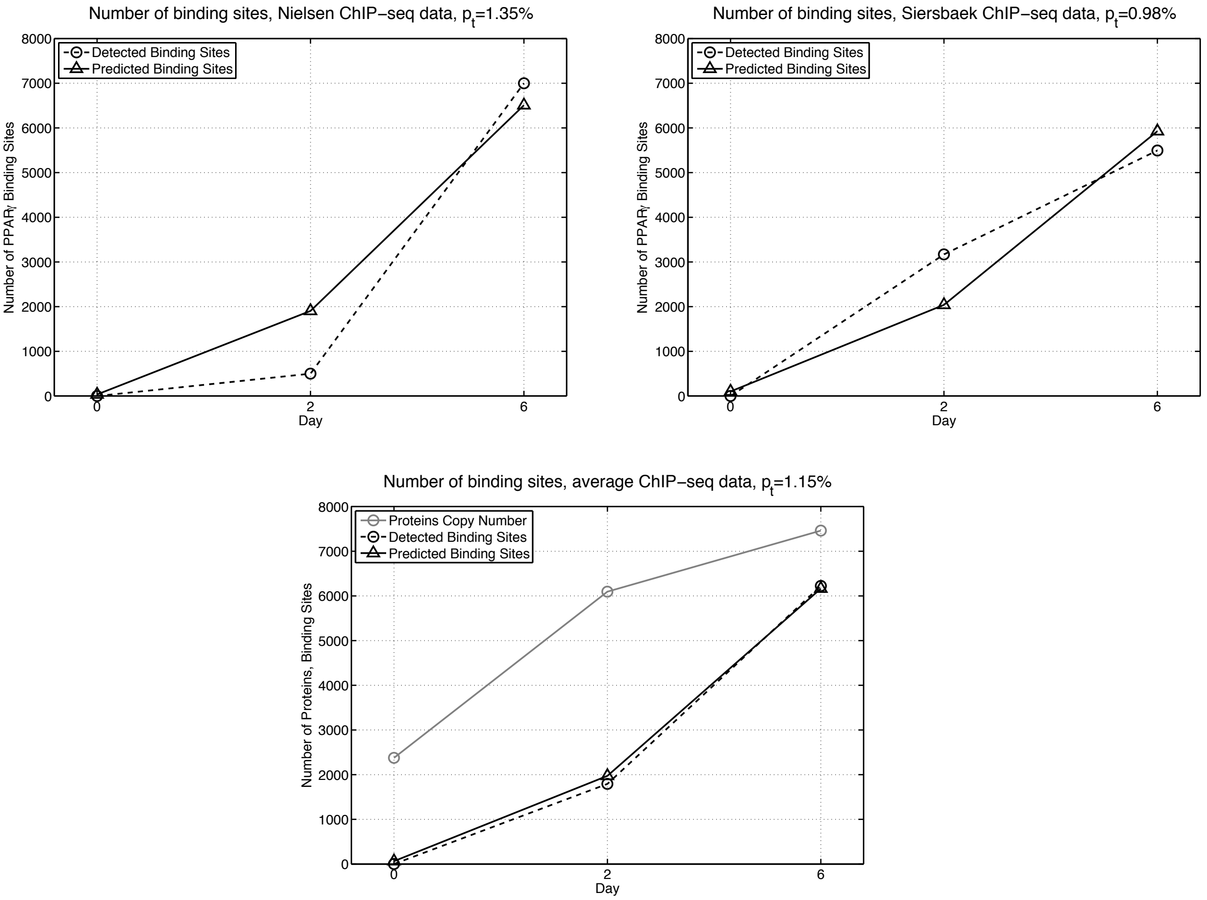
Supplementary Figure 15. Top: site occupancy p with respect to affinity X, computed with the approximate expression (2) and exactly from the partition function (1), for a reduced system with comparable concentration of proteins than $\mathrm{PPAR}\gamma$ at day 6 (cf. Section 2.3). The theoretical distribution of sites with respect to the affinity given by the PWM of $\mathrm{PPAR}\gamma$ was discretized in $k = 20$ categories, covering the range of affinities $[X_0 = 1, X_{max} = 1000]$. Bottom: comparison between the approximate and exact occupancy for the same system. Even if the stronger sites are in low amount compared to the number of proteins ($m_{19} = 1$, $m_{18} = 2$, ..., $m_{14} = 80 < N = 100$), the approximation (2) still predicts very well their occupancy.



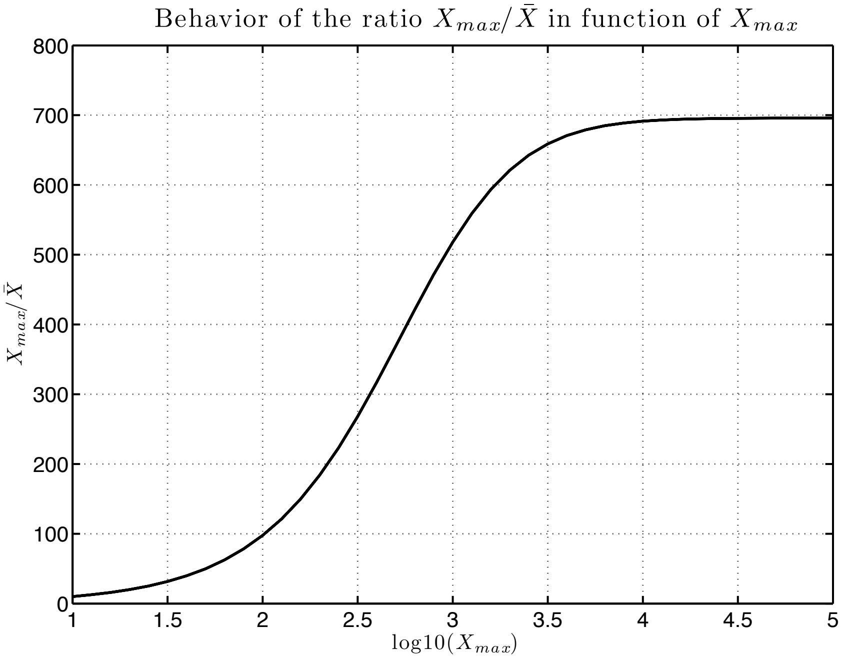
Supplementary Figure 16. Left column: fitting of the selection parameter λ from the mean ChIP-seq signal at day 0, day 2 and day 6 in log-space. Right column: representation of expression (5) in real space, using the estimated parameters on the left.



Supplementary Figure 17. Left: cumulative number of sites found in H3K27ac regions with FIMO [7] at day 0, day 2 and day 6. Right: cumulative number of sites approximated as a power law, this representation gives excellent results $R^2 > 0.99$.



Supplementary Figure 18. Left: predicted number of $\mathrm{PPAR}\gamma$ binding sites based on Nielsen et al. ChIP-seq data [3]. Right: predicted number of $\mathrm{PPAR}\gamma$ binding sites based on Siersbaek et al. ChIP-seq data. [11] Bottom: predicted number of $\mathrm{PPAR}\gamma$ binding sites compared to average number of detected sites in both [3] and [11].



Supplementary Figure 19. Effect of X_{max} on the ratio $X_{max}/\bar{X}$ which appears in the occupancy expression (8). If $X_{max} > 10^3$, the ratio becomes insensitive to X_{max}. On the other hand, if $X_{max} < 10^3$, the occupancy will decrease.