ANNOTATION OF THE HUMAN ODONTOBLAST CELL LAYER AND DENTAL PULP PROTEOMES AND N-TERMINOMES

by

Simon Abbey

B.Sc., Queen's University, 1989

D.M.D., The University of British Columbia, 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Craniofacial Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2017

Abstract

The proteome and N-terminome of the human odontoblast cell layer was identified for the first time by shotgun proteomic and terminal amine isotopic labeling of substrates (TAILS) N-terminomic analyses, respectively, and compared with that of human dental pulp stroma from 3rd molar teeth. After reverse-phase liquid chromatography-tandem mass spectrometry, >170,000 spectra from the shotgun and TAILS analyses were matched by four search engines to 4,888 and 12,063 peptides in the odontoblast cell layer and pulp stroma, respectively. Using the Trans-Proteomic Pipeline, I identified 895 and 2,423 unique proteins in these tissues at an FDR of ≤ 1 %. In the odontoblast cell layer proteome I found proteomic evidence for dentin sialophosphoprotein, which is cleaved into dentin phosphoprotein and dentin sialoprotein, proteins that are important in dentin mineralization. Further, 222 proteins of the odontoblast cell layer were not found in the pulp, suggesting many of these proteins are synthesized preferentially by odontoblasts. I also found minor differences in the odontoblast cell layer between the dental pulp proteomes of older and younger donors. The human dental pulp stroma proteome was expanded by 974 new proteins, not previously identified among the 4,123 proteins identified in our previous dental pulp study (Eckhard et al., 2015). Thus, by exploring the proteome of the odontoblast cell layer and expanding the known dental pulp proteome, we found distinct proteome differences when compared with each other and with dentin. The mass spectrometry raw data and metadata have been deposited to ProteomeXchange with the PXD identifier <PXD006557>.

Lay Summary

Dental pulp comprises the central pulpal stroma and the odontoblast periphery. We hypothesized each expresses different proteins. We removed dental pulp tissue from freshly extracted teeth and separated these layers. Protein from each of these samples was broken down into its constituent peptides and these were identified using mass spectrometry. Using software parent proteins were identified from peptides which were unique to individual human proteins. We thus identified 4,888 peptides in the odontoblast sample and 12,063 peptides in the pulp stroma. 222 proteins were found only in the odontoblast sample, suggesting they are synthesized by odontoblasts. We also found differences in the odontoblast cell layer proteomes between older and younger donors. In summary we expanded the human dental pulp proteome by 974 new proteins not previously found in our previous dental pulp study. We also found proteome differences between the odontoblast cell layer and the pulp stroma.

Preface

Dr. Overall and I met in early 2014 to discuss this research project. The Overall lab is involved in the human proteome project and Dr. Overall was interested in taking me on as a graduate student to do proteomics on human dental pulp, adding this tissue to others already being studied in the lab. The terminal amine isotopic labelling of substrates (TAILS) protocol had been developed in the Overall lab previously and we discussed using it on dental pulp tissue (Kleifeld et al., 2011a). Dr. Eckhard had removed one dental pulp from a tooth extracted by Dr. Mathew several months earlier, as a pilot project.

The project began in May 2014. The first phase was collecting the pulp tissue. Dr. Mathew and I coordinated this that summer; I travelled to his office, where he extracted the donor wisdom teeth. I developed the technique for removing the pulp tissue from these teeth and freezing them on dry ice immediately. Together, Dr. Mathew and I collected all pulps studied in this research and in our previous study: he extracted the teeth, I then removed the tooth pulp tissue and froze it in GuHCl buffer. We obtained pulp tissue from a total of 16 donors over the course of the summer.

Dr. U. Eckhard and I worked closely together that summer. We did all experiments together for donors 1-10 of 16 total. Dr. Eckhard worked very closely with me for the first run of the experiment (donors 1 and 2) and for the remainder of the summer I did all the experiments for donors 3-10 with Dr. Eckhard close-by for clarification and consultation.

I was responsible for planning and implementing the histological component of this research, preparing a pulp, sectioning it, and staining it. We had other sections done by the UBC Wax-it service. I did all histology under the guidance of Dr. G. Tharmarajah.

During the summer of 2015 I did initial analysis of LC-MS/MS data for all 16 pulp donors. I prepared spreadsheets of the data for further software analysis by Dr. Eckhard.

At this point, Dr. Overall decided to divide the project in two. For donors 1, 2 and 11-16 Drs. Eckhard and G. Marino performed all further experiments to process the pulp proteins for LC-MS/MS analysis, did the analysis of these spectra, and prepared the figures for the first paper published from this data (Eckhard et al., 2015). For this first paper, I removed and did initial work on all pulp tissue used, and performed approximately 20 % of the experiments to process this tissue through PreTAILS and TAILS workflows. I also performed initial data analysis using Excel, reviewed the final manuscript, made edits and consulted from an endodontic perspective. I was also responsible for the histology image used in this paper.

Through 2015 and 2016, I performed detailed software analysis on my LC-MS/MS data from donors 3-10, samples for which I had done the experimental work and analysis. In performing the analysis of MS spectra I worked closely with Dr. U. Eckhard and he did the final analysis to prepare our data for publication. I took the 5 day long Trans-Proteomic Pipeline introductory course given by the Seattle Institute of Systems Biology in the fall of 2015. Following this, I understand and can use all the software we employed to analyze our LC-MS/MS data, but my understanding of proteomics is not at the level of Dr. Eckhard and he was critical to final preparation of my data.

In 2016 and 2017 I organized my data for publication, prepared the figures, and wrote the manuscript for a second paper that has been submitted to the Journal of Dental Research. Myself and Dr. Eckard share first author status on this paper.

I presented new research posters, each serving as a landmark for my progress in this project, in every year of my program, January 2015, 2016, and 2017. At this year's research day poster competition I received the M.Sc. student award.

Our research project received ethical approval from the UBC Clinical Research Ethics Board (certificate # H13-03006).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EDTA – Ethyldiamine tetra-acetic acid

FDR – False discovery rate

GuHCl – Guanadinium hydrochloride buffer

HUPO – Human proteome organization

HPP – Human proteome project

LC-MS/MS – Liquid chromatography-tandem mass spectrometry

LPS – Lipopolysacharide, an endotoxin released by Gram negative bacteria

MALDI-TOF-MS – Matrix assisted laser desorption/ionization time-of-flight mass spectrometry

MS – Mass spectrometry

neXtProt - (http://www.nextprot.org)

m/z – Mass/Charge ratio

TAILS – Terminal Amine Isotopic Labelling of Substrates

TPP – Transproteomic pipeline software package for analysis of MS data

UniProt – (http://www.uniprot.org)

Acknowledgements

Many thanks to Drs. Chris Overall and Ulrich Eckhard for their support and assistance in conducting this research project, also to all members of the Overall Lab, Dr Tharmarajah for her help with the histology component of my project, and to the members of my thesis committee. Thanks also to the Canadian Forces for sponsoring me during the initial months of my research, and to the Canadian Academy of Endodontics for their financial support of my project.

Dedication

Dedicated to my children Dylan, Julian, Bronwyn and Elijah

Chapter 1: Introduction

1.1 Basic science background to project

1.1.1 Proteomics

Proteomics is a burgeoning area of scientific research aimed at identifying the individual proteins found in tissues. It is the large-scale study of proteins. To some degree it is an evolution from genomics, the deciphering and characterization of the human and other genomes. But while genomics examines the fixed blueprint of chromosomal DNA, proteomics attempts to understand and quantitate active processes of gene regulation and expression in both health and disease (Anderson and Anderson, 1998). The goal of the International Human Proteome Project (HPP) is mapping the entire human proteome to improve our understanding of human cellular biology and facilitate development of diagnostic, prognostic, therapeutic, and preventive medical applications. Approximately 20,300 genes are identified in the human genome which code for proteins (Paik et al., 2012). The goal of the HPP is to identify at least one protein from each of these genes in at least one human tissue.

1.1.2 Transcriptomics

Proteins are difficult molecules to characterize because of the ubiquity of protein modifications, both at the post-transcriptional RNA level and post-translational protein level. These are independent processes. Transcriptomics is the branch of life science between genomics and proteomics: the study of nucleotide chains of mRNA isolated from tissue. The genome codes for a set of basic protein building blocks, but this is only the starting point and biological processes modify these to create the almost infinite functionality required for life. Transcriptomics reveals changes that occur between mRNA transcription and protein translation. It is estimated that post-transcriptional RNA modifications are responsible for as many as a 5-fold increase in

proteins as the number of genes alone (potentially coding for as many as 100,000 proteins (Sharon et al., 2013).

1.1.3 Post-translational protein modifications

Once proteins are translated, post-translational modifications always occur. This may be as simple as cleavage of the N-terminal methionine residue, may involve other internal cleavages, or consist of different modifications such as phosphorylation, acetylation, or the removal of signal peptides. There are over 300 known post-translational modifications which are estimated to produce in the order of 10 times more final protein products than the approximately 100,000 (from transcriptomics) translated peptide sequences, creating in the order of 1,000,000 different proteins (Beck-Sickinger and Mörl, 2006; Zhao and Jensen, 2009). These modified proteins may be isoforms of a protein family with similar but related functions, or a translated polypeptide may be cleaved into several proteins with independent functions. For example, following translation dentin sialophosphoprotein is cleaved and folded into dentin sialoprotein, dentin phosphoprotein, and dentin glycoprotein, each with independent functions within the dentin matrix (Hargreaves et al., 2012; Jágr et al., 2012). Such a discrete cleavage into three proteins with different but related roles is probably uncommon, but evidence shows that a single translated protein can be modified in many different ways so that the different final protein products may appear in different tissues and have different functions.

1.1.4 Protein variability with time

A final layer to the complexity of proteomics is its mutability with time. Although epigenetics has demonstrated that changes in chromosomes can affect gene expression the genome is fixed and does not change throughout the life of an organism. Genetics is the basic blueprint from which the infinite adaptability and complexity of proteomics is created downstream through

transcription, translation and post-translational modifications. Protein expression changes constantly depending on the needs of a cell and biologic demands placed on tissues. Some proteins are common, for example type I collagen is frequently a component of extracellular matrix. As such Type I collagen is a relatively stable element within these tissues. However, other proteins may be more reactive, may be needed only in small quantities or at rare intervals, for example due to a stressful event or lack of a nutrient. These latter proteins will be more difficult to identify.

1.1.5 The Human Proteome Project

The basic strategy of the human proteome project is to do proteomic studies on as many tissues as possible in order to identify as many proteins as possible. The dental pulp is an accessible tissue that could contribute significantly to this work. Some work has already been done on human pulp proteomics, which I will review. I will also discuss dentin proteomics because the odontoblast layer of dental pulp generates dentin. Thus, pulp and dentin proteomics are closely related and it is beneficial to be familiar with both in the study of dental pulp.

1.2 Dental pulp

1.2.1 Tooth structure

The dental pulp is the soft tissue core of a tooth. Developmentally, the human tooth is created from an infolding of dental epithelium and ectomesenchymal tissue, where the epithelial layer produces ameloblasts, which form the enamel layer of a developing tooth. These cells subsequently undergo programmed cell death when enamel formation is complete. At the same time, mesenchymal tissue, interior to the forming enamel, develops into odontoblasts, which begin to produce dentin; unlike ameloblasts, odontoblasts remain vital and persist following

tooth eruption. Enamel is the hardest tissue in the body, while dentin is the second. This is a function of their mineral content: enamel is 96 % inorganic, 4 % water and organic matrix, while dentin is 70 % inorganic, 20 % organic matrix, 10 % water (Hargreaves et al., 2012). The line between dentin and the soft tissue pulp is demarcated by the odontoblast layer, which forms the outermost layer of the pulp abutting the predentin. Predentin is the initial stage of new dentin formation, consisting of mineralized dentin organic matrix newly formed by odontoblasts. Odontoblasts form a continuous unicellular layer of columnar cells at the pulp-dentin border, and these cells extend processes into the dentinal tubules.

1.2.2 Odontoblast function within the pulp organ

Odontoblasts perform many roles in the dental pulp organ. Odontoblasts probably communicate with pulp neurons since the odontoblast process penetrate into the dentinal tubules and may be responsible for detecting stimuli that cause tooth sensation. However, histology studies have shown that axons penetrate the odontoblast layer and also enter dentinal tubules, although probably not as deeply as the odontoblast processes. These nerve fibers could also be responsible for dental sensation. Dental nerve fibers also play a defensive role in initiating pulpal inflammatory responses following stimulation by bacterial toxins, *e.g.* Lipopolysacharide (LPS), through the release of neuropeptides (Henry and Hargreaves, 2007). Odontoblasts could be responsible, or involved in the initiation of this neuropeptide release. Odontoblasts are also known to be the primary cell type responsible for dentin formation, maintenance and repair. Odontoblasts can be renewed and regenerated throughout life, as new dental pulp stem cells are stimulated to differentiate to replace odontoblasts that have been destroyed by advancing carious lesions, trauma, etc. (Hargreaves et al., 2012).

1.3 Literature review of related studies

Prior to our group's work there are only two published studies that directly examine human dental pulp proteomics *in vivo*; several other studies examined pulp tissue culture proteomics *in vitro*, and several more looked at dentin proteomics.

1.3.1 Pääkkönen et al. 2005

Pääkkönen et al. 2005 analyzed pulp tissue prepared in a similar manner to me from extracted wisdom teeth, but examined both the mRNA gene transcripts, and tissue sample protein extract (Pääkkönen et al., 2005). Also of note Pääkkönen's group divided pulp tissue samples into two groups: 1) those from healthy wisdom teeth and 2) those from carious wisdom teeth. Pääkkönen's study was commendable in that it attempted to find proteomic differences between healthy and diseased tissues so that proteins could be identified which are expressed under different conditions. By characterizing both the transcriptome, using cDNA microarrays, and the proteome, using 2D gel electrophoresis and MS analysis, the authors hypothesized that differences would be found between healthy pulps and carious pulps. This makes intuitive sense since caries bacteria stimulate pulp tissue to form tertiary dentin and initiate activation of both the innate and adaptive immune systems.

Some differences were found in transcription, but the heterogeneity of samples meant these differences may well have appeared stochastically. In terms of mRNA, they found evidence that 396 genes were expressed in at least one sample of the 1081 cDNAs screened. This was assessed by fluorescence level for a given cDNA, which was quantified to indicate if the gene was transcribed or not, and if this signal was 5x this base level for expression it was classified as 'highly expressed.' But for calculating differences between carious and heathy tooth structure a difference of 1.4x was considered to represent a 'true difference.' These

factors are somewhat arbitrary, but the data produced shows that the same proteins may be active in both samples but more are transcribed in either healthy or carious pulp depending on the gene. This shows that biologic function is not only influenced by the presence or absence of proteins but by their relative numbers and proportions. The proteomics analyses found evidence of 96 proteins. In terms of our results, these numbers are very small. Individual cells probably require as many as 10,000-30,000 proteins at a given point in time to carry out their biological functions (Alberts et al., 2002). Thus, by identifying only 96 proteins, approximately 1 % or less of the total proteome, they had a low chance of finding differences caused by pulpal response to caries. Of these 96 proteins, only 12 were translated from the expressed cDNAs and none from the highly expressed cDNAs that had been identified in the transcriptomics portion of the study. Before we can confidently move into an area where we can identify differences in proteomics between diseased and healthy tissues, we should ideally establish a consistent baseline, *e.g.* multiple studies from multiple centers find the same or a similar proteome for healthy dental pulp, and this should be a large proteome in the order of 10,000 plus proteins so we can be confident most of the proteins found in the dental tissue are represented. Also, proteomics is even now not quantitative and differences in the number of peptides found may reflect relative differences in peptide stability under experimental conditions as well as the number of parent proteins expressed in the tissue sample.

1.3.1.1 Two dimensional (2D) Gel electrophoresis

Pääkkönen's study used 2D gel electrophoresis to separate proteins based on their mass and isoelectric points. Briefly proteins are solubilized and electrophoresed based on isoelectric point; these gel samples are then loaded onto SDS-PAGE gels where they are subjected to another electric field and move through the gel at a rate inversely proportional to their mass. There was no difference found between the protein spots on the healthy pulp gel compared with the

carious pulp gel, so expression of these 96 proteins was approximately the same in both samples. The authors suggested this could be because the carious pulp samples they selected were from teeth that had only small or moderate sized carious lesions, meaning inflammation would be limited and localized to a small area of the pulp leaving the majority of pulp tissue in a healthy condition. The final phase of 2D gel proteomics is to cut each protein dot from the gel, digest the protein into peptide fragments, and put these through MS to produce spectra which can identify the parent protein. This technique is effective at identifying proteins since the mass, isoelectric point, and MS peptide identification, can all be used to confirm the protein is correctly identified. However, it is laborious, and proteins must be present in sufficient quantities to appear as a stained spot on the gel to be identified. Typically studies of this nature identify hundreds of proteins, where complete tissue proteomes are probably in excess of 10,000 proteins.

1.3.2 Wei et al. Cell culture study

Wei et al. took a very different approach in studying pulp proteomics (Wei et al., 2008). The authors took a tissue culture of dental pulp stem cells, and induced these with β-glycerophosphate, which stimulates dental pulp stem cells to differentiate into odontoblasts. This is an important step in the formation of reparative dentin when new odontoblasts are required to replace cells destroyed by rapidly advancing caries. Induced cell culture protein extract was compared to non-induced control culture and protein spots were compared on 2D gels using immunofluorescence to identify differences in expression. Spots that were over- or under-expressed in the induced cell culture were excised from the gels, dissolved, put through trypsin digest and MALDI-TOF-MS analysis. 23 proteins were identified to be up or down regulated in cultured dental pulp stem cells induced to develop into odontoblasts. The authors grouped these as cytoskeletal proteins, nuclear proteins, calcium binding proteins, proteins

involved in matrix synthesis, metabolic enzymes, cell signaling or proteins with unknown functions. This is a commendable study, showing how different environmental conditions change protein expression. However, this is a tissue culture study and *in vivo* conditions are understood to be much more complicated than a single cell line being exposed to a single stimulus. Also of note, authors in many proteomics papers categorize proteins found as belonging to certain sub-groups of proteins. This implies the functions of proteins identified are well understood. But multiple functions are usually ascribed to given proteins, and this list tends to expand with time. The same protein, or different isoforms of a protein, may perform varying functions in different tissues. Thus, assigning a functional role to a given protein identified in a tissue is somewhat arbitrary. For example, Wei lists Vimentin as being a cytoskeletal protein. Vimentin has been identified in this role, but the UniProt website (http://www.uniprot.org) lists multiple other functions, including double stranded RNA binding, glycoprotein binding, identical protein binding, keratin filament binding, C-protein terminus binding, scaffold protein binding, a structural constituent of cytoskeleton, and a structural constituent of eye lens. Thus vimentin's exact function within this cell line under this stimulus is uncertain.

1.3.3 Jágr et al. dentin and dental pulp studies

Two further papers I will review are from Jágr et al (Eckhardt et al., 2014; Jágr et al., 2012). They used extracted wisdom teeth, taking five teeth from five donors, ages 22-23. The technique they used to separate dentin from pulp, enamel and cementum was to first scrape away the cementum layer and soft tissue remnants with an iron spatula, and then to remove the coronal part of the tooth by cutting away the crown beneath the cemento-enamel junction. Following these procedures, the tooth roots were crushed in a jaw vice. Finally pulp and dentin fragments were separated manually with cotton pliers and frozen in liquid nitrogen. The first study is an analysis of the dentin proteome, the second of the pulp proteome. These proteomes

are related since dentin is understood to by synthesized by pulp. As Eckhardt states in the second paper in this series, their objective was not only to “create a. . .list of proteins present in human dental pulp tissue,” but also to “study the proteins in the pulp-dentin complex.” Several other studies have been done on dentin proteomics (Chun et al., 2011; Park et al., 2009) ,but Jágr et al’s is the most comprehensive. Finally, in this chapter I will briefly discuss a review paper (Jágr et al., 2014).

1.3.3.1 Jágr et al. Dentin proteomics

The first of their papers examines the dentin proteome (Jágr et al., 2012). Frozen dentin fragments were pulverized into powder, which were extracted with GuHCl, demineralized with EDTA, and finally subjected to gel separation. Jágr used 2D gel separation of dentin proteins where previous dentin proteome studies had used 1D gel separation. This contributed to improved annotation of proteins through more peptide identifications. Gel protein analysis is advantageous in that it produces some quantitative information where my research, to be discussed later, is purely qualitative, *e.g.* showing the protein was present but giving no indication of its relative proportion within the total tissue protein. Jágr et al. are also aware that proteins with high pH values are not separated well on gels, nor are proteins with high or low molecular weight. Following the 2D gel separation technique protein dots were removed, dissolved, trypsin digest was performed followed by LC-MS/MS analysis.

1.3.3.2 The Dentin proteome is relatively small

Dentin is potentially an advantageous tissue to do proteomics on since the number of proteins in dentin may be less than in most soft tissues. The mineralized tissue component of dentin is non-vital so it can be expected that dentin organic matrix is fairly consistent, made up from a limited number of proteins laid down as a lattice during the initial stages of dentin synthesis.

Importantly, these proteins probably have minimal turnover, due to tissue non-vitality, whereas the vital dental pulp proteome in all likelihood changes dynamically, both during development and in response to environmental changes caused by caries, trauma etc. Jágr et al. compare their results with the two other dentin proteomics studies cited above and stated the top 10 most abundant proteins found in dentin were “approximately” the same in all three studies. Despite this somewhat vague conclusion this is a primary goal of proteomics research: to find correlation between different studies of the same tissue. They were able to find 289 proteins but 20 were removed from the total because they were likely contaminant proteins, keratins etc. Keratin could have come from skin, despite careful gloved handling of tissues, in sufficient quantities to appear on experimental gels. An issue with Jágr et al’s study, and many proteomic studies, is purity of sample. Experimental teeth were crushed and the soft tissue components removed, but because odontoblast processes extend into dentin, pulpal soft tissue is an integral component of dentin hard tissue and very difficult to remove. Also, the separation of soft and hard tissue components with tissue pliers may be where many of these putative contaminants originated in the sample, since smaller fragments would have been difficult to see and remove. For example, the authors identified three enamel proteins—ANXA1, ANXA2 and ANXA5—that are probably contaminants from enamel that became accidentally incorporated into the sample. This is always an issue with proteomics, proteins identified in tissue samples cannot be specifically tagged to cells or tissue regions unless radiolabeled antibodies to individual proteins are made, which can later be localized using tissue histology. For example, immune response proteins were identified in this study, suggesting dentin is a more biologically active tissue than is thought, but these proteins could also have come from fragments of dental pulp tissue in the original dentin sample.

1.3.3.3 Eckhard et al., 2014 the dental pulp proteome

The second half of this study examined the dental pulp tissue from the original crushed wisdom tooth roots (Eckhardt et al., 2014). The manner in which pulp samples were obtained, crushing the roots in a vice and removing the hard tissue fragments, could have created considerable contamination of sample with dentin, cementum and even enamel fragments. Moreover these procedures would have involved significant trauma to pulp tissue and encouraged natural proteolysis, possibly reducing the final yield of proteins, *e.g.* caspase activity associated with apoptosis.

Eckhardt et al. performed a good analysis of the proteins they found in pulp and compare these to the proteins previously found in their dentin samples. They also include a thorough review of dental pulp literature and drew attention to examples of specific proteins found in both their and other studies. For example, they highlighted vimentin and nestin (which we also found in our 2015 study) (Eckhard et al., 2015; Eckhardt et al., 2014).

1.3.3.4 Eckhardt et al., 2014 data analysis

A criticism of this research is that the LC-MS/MS analysis is not up to current standards. They used Mascot thresholds for determining the true hits, a Mascot score ≥ 20 to identify a peptide and ≥ 60 to identify a protein. The current standard is to start with the search software, *e.g.* Mascot, but then input this search engine data into higher level software, *e.g.* the TPP, to assess the number and quality of mascot hits for specific peptides in determining FDR for both peptides and proteins (in our study we used FDR ≤ 1 % for both). This is not wrong per se but does not conform to current HPP standards (Deutsch et al., 2016a). Using their data analysis protocols they found 342 pulp proteins.

1.3.3.5 Practical applications of proteomics research

In their introduction, Eckhardt et al., made an effort to relate their proteomic research to clinical application for regeneration and repair of the pulp dentin complex (Eckhardt et al., 2014). The 'why' and 'where will this take us' are valid questions in proteomic research, but clinical applications are still largely conjectural and, at this stage, proteomics is more molecular anatomy than becoming applied molecular science. We need to understand the anatomy before we can understand function and the anatomy is very complicated as I discuss in this thesis. I do not feel we are on the verge of discovering some critical protein that will revascularize the pulp-dentin complex in regenerative endodontic therapy and the authors mislead the reader somewhat in suggesting they were.

Eckhardt et al. show a Venn diagram (Fig. 2) illustrating 37 proteins shared by pulp and dentin tissues which are not present in the plasma. They concluded "these proteins might be candidates to participate in the unique pulp-dentin complex and thus have potential in future regenerative approaches." This is a misrepresentation of their data. Any of the proteins in the pulp tissue could be important for regeneration, since it is the pulp which produces the dentin. Dentin contains growth factors that stimulate the pulp to differentiate and lay down dentin (Hargreaves et al., 2012). But the authors seem to infer that a Venn overlap indicates an anatomical relationship where the proteins are working side by side to create new dentin.

Eckhardt et al. made a brief attempt to justify their claim that their research is important to improve regenerative therapy in discussing a cell culture study comparing the proteomes of bovine dental pulp stem cells, mesenchymal stem cells, and periodontal ligament stem cells, finding 5 proteins that were upregulated in the dental stem cells compared with the other two stem cell populations (Mrozik et al., 2010). Eckhardt et al. identified two of these in their research: ubiquitin carboxyl-terminal hyrdrolase isozyme L1 and Rho GDP-dissociation inhibitor 1. However, G6PD is a basic enzyme in the pentose phosphate pathway, active in all cells,

Eckhardt et al., did not discuss how this protein could be important for regenerative therapy. These experiments are significant efforts to better understand proteomics, how protein expression changes across related cell types and within cell types under changing conditions. However, I feel it is important to not overstate the short term impact of proteomics—which is still firmly in the realm of basic science—or proteomics may lose credibility.

Jágr et al. also published a broader review in “Proteomics of Human Teeth and Saliva” which discusses pulp and dentin in some depth and their relationships with other tooth components and the importance of saliva (Jágr et al., 2014). I mention this paper because the authors discuss caries, the commonest disease of teeth, and highlight how much remains unknown about why some individuals are caries resistant while others are caries susceptible. Jágr et al., propose this may well be due to protein antibacterial agents or proteins which protect tooth surfaces in some way. These could be components of saliva or dentinal fluid. Elucidating other protein mechanisms of caries resistance or caries susceptibility could lead to clinical applications or at least to better understanding of cariology and epidemiology. Individual variation in the ability of saliva to buffer bacterial acids which cause caries is well understood, but there could be additional pulpal elements, potentially proteins, which also confer caries resistance.

1.3.3.6 Organic matrix of dentin

The role the organic matrix of dentin plays in initiating and regulating dentin mineralization is understood on a conceptual level but not in detail. For example, Jágr et al. discuss how some matrix proteins are thought to act as nucleators, attracting hydroxyapatite crystal formation around them, while other proteins act as inhibitors to mineralization, possibly maintaining patency of dentinal tubules (Jágr et al., 2014). Clinical endodontics shows many examples of how mineralization can go awry, for example root canal calcification, formation of pulp stones,

sclerotic dentin. Dentin and root canals are highly variable structures. They go on to highlight many pathological processes within dentin already understood to involve problems with proteins. Collagens are the major protein component of dentin, comprising 85-90 % of its organic structure. Collagen type I is a simple protein made up of X-gly-X repeats, but it's combinations with other forms of collagen and 3-D molecular structure quickly make it a very complicated lattice upon which hydroxyapatite can be laid down. Collagen types III, V, VI, XI, XII are all understood to be components of dentin organic matrix. Minor protein components of dentine may be just as important in affecting its biological properties. For example, defects in dentin sialophosphoprotein, the most abundant non-collagenous protein found in dentin, are associated with dentin dysplasia type II and dentinogenesis imperfecta types II and III. Dentin sialophosphoprotein is thought to stimulate mineralization of dentin but dentin sialophosphoprotein is also found in non-mineralizing tissues where it must perform other functions (Jágr et al., 2014).

1.3.3.7 Dentin participation in the immune response?

Jágr et al. 2014 hypothesize dentin protein components of the immune response may contribute to caries prevention. But the evidence level is weak to make this supposition. Can the immune response of the dentin-pulp complex prevent caries? Caries can progress slower or faster in different individuals and this may be related to pulpal immune response, but it would be difficult to design a study which standardized patient bacterial flora, diet and oral hygiene—known strong risk factors, so that significant dentin-pulp complex immune protein effects could be observed. Jágr et al. discuss all of the dental pulp's cellular constituents, their role in dentin regeneration, immune response, and tooth sensation and how the neural component of dental pulp is vital to function in many ways. For example, denervated teeth have been shown to have diminished survival, confirming important cooperation between odontoblasts and/or other pulpal

cells, and nerve fibers. In addition to fibroblasts, the dental pulp stroma interior to the odontoblast layer is made up of many immune cells: dendritic cells, macrophages, lymphocytes, endothelial cells, and mesenchymal cells including dental pulp stem cells, which are one of the potential stem cell populations to allow pulp regeneration/ revascularization therapy.

1.3.4 Eckhard et al., 2015

1.3.4.1 Currently the definitive study in pulp proteomics

The largest pulp proteomics study to date, was published by our group in 2015 (Eckhard et al., 2015). We found 4332 proteins at an FDR ≤ 1 %, making it by far the most extensive pulp proteomics study to date. This was a proteomics oriented study, aimed at identifying the largest number of proteins possible with high confidence using multiple search engines and the TPP, closely adhering to HPP mass spectrometry guidelines for the standardization of data interpretation in protein identification worldwide (Deutsch et al., 2016a). Most of the techniques I employed in my research protocols were developed in this research so the body of my thesis is in many ways part two of this 2015 study.

1.3.4.2 Missing proteins

Eckhard et al. discuss the problem of persistent missing proteins. After numerous high quality, large scale proteomics studies, more than 12 % of predicted proteins remain undetected, likely due to one of six possible factors. Missing proteins remain unfound because they either: 1) appear only in organs or regions difficult to study, 2) are expressed only during certain developmental stages (*e.g..* prenatally), 3) are expressed only under unusual stress conditions at which times it is difficult to take tissue samples, 4) are present in very small numbers and so below detectability with current technology, 5) lack tryptic cleavage sites to produce peptides of manageable size for LC-MS/MS detection (*e.g.* peptides which are too long to be processed

predictably), or 6) share sequence homology with other proteins. This last possibility is an interesting dilemma where the same tryptic peptide may originate from several different proteins. In such situations, the TPP counts the redundant peptide as contributing to confirming all potential source proteins, but additional non-redundant peptides are required to confirm that each specific protein is present. Thus, if a 'missing' protein has only tryptic cleavage sites which produce peptides shared with other proteins, and lacks a specific non-redundant peptide which can originate only from the missing protein, then it will not be positively Identified with proteomics using trypsin digest peptides, even though it's redundant peptides are present in the samples.

1.3.4.3 Patterns of peptide N-terminus modifications

Eckhard et al. performed a thorough analysis of modifications to identified peptides which show patterns applicable to human proteomics in general, for example they found that where the N-terminus was an Alanine, Serine, or Threonine residue it was almost always acetylated, whereas for Glycine, Lysine and Valine residues N-termini were typically free. Eckhard et al. discuss the complex biology of dental pulp, being comprised of fibroblasts, plasma proteins and other proteins derived from red and white blood cells, *e.g.* cathepsin K from neutrophil lysosomes. Dentin sialophosphoprotein was also identified in the pulp samples and attributed this to the presence of odontoblasts. The paper discusses how ideally proteins should be identified with several high quality peptide spectra matches and for the best confirmation such peptide spectra are compared to the spectra of synthetically derived reference peptides. The analysis of peptides identified is very detailed and reveals many interesting products, for example proteins were identified with unprocessed signal peptides. These signal peptides indicate the protein is marked to be secreted. So these proteins were identified somewhere

between translation and secretion (when the signal peptide is usually removed by signal peptidases)(Eckhard et al., 2015).

1.3.4.4 Novel cleavages identified by TOPFINDER tool

This study shows the power of our preTAILS shotgun combined with TAILS analysis examining semi-ArgC peptide cleavage products in expanding the proteome of dental pulp. We discovered many internal N-termini indicating novel cleavages, many of which we were able to map with the web tool TOPFINDER developed by our group. This is a data analysis tool that contains information on existing evidence for novel termini, so our N-terminal peptides could be confirmed as previously identified, or if not they could be added to the database. This is an important collaborative tool into post-translational modifications; individual protein products which undergo enzymatic cleavage to form multiple proteins with independent biologic activity.

1.4 Research goals

My research was fairly large in scope, examining the dental pulps of 26 wisdom teeth from 9 donors. The primary goal of my study was to separate the central pulp tissue from the odontoblast cell layer to produce an odontoblast enriched sample and a pulpal stroma sample. Then by using techniques developed in our earlier study (Eckhard et al., 2015), I hoped to demonstrate differences in the proteomes of these two tissues, reflecting their different functional roles in the dentin-pulp complex. My secondary goals were to continue our earlier research by confirming and further annotating the pulp stroma proteome and, hopefully, identify some proteins classified as ‘missing’ within the neXtProt database (http://www.nextprot.org). I also wanted to take advantage of the fact that our donors age ranged from 15 to 39 years, a period during which wisdom teeth complete their development, meaning they shift from the late developmental period of rapid dentin formation and root completion, to the adult period of dentin

maintenance. By comparing the younger pulps, i.e. ≤ 20 years, to the mature pulps, i.e. > 20 years, I hoped to find proteomic differences that reflected this shift. This would give me multiple data sets to compare, pulp stroma to odontoblast, young to mature, young stroma to old stroma etc. By subdividing and comparing the proteomes of these different groups I hoped to gain insight into the biology of dental pulp through patterns and/or differences I identified.

Chapter 2: Body of Thesis

2.1 Introduction

The human dental pulp organ is the terminal developmental stage of the embryologic dental papilla from which the mesenchymal tooth elements, pulp and dentin, originate. As such the dental pulp is responsible for the various developmental stages of dentin and then dentin maintenance throughout life. Upon tooth eruption, the dental pulp establishes neural communication with the trigeminal ganglion, which allows tooth sensation, for example to temperature, pressure, pH, and osmotic forces. The dental pulp also responds to various events and stresses that may occur in the life of the tooth, *e.g.* caries, trauma, occlusal disturbances (Hargreaves et al., 2012; Kumar, 2011). Thus, knowledge of the pulp proteome should lead to improved understanding of pulpal function at a molecular level, and this has potential for clinical translation *e.g.* to reduce chronic dentinal sensitivity, or to encourage pulpal regeneration.

Cell protein expression dynamically changes depending on the current needs of the cell and the cell's role within tissue. Furthermore, different regions of an organ are specialized to perform specific functions. Thus, the odontoblast layer is responsible for producing and maintaining dentin, synthesizing proteins in common with other mineralized tissues, and unique proteins with roles in mineralization, most notably the dentin-specific dentin sialophosphoprotein (MacDougall et al., 1997), whereas the pulpal stroma is responsible for supplying nutrients and removing metabolites from the odontoblast cell layer and also maintaining the nerve and blood supply to the pulp. No proteomic analysis has been performed to date on odontoblasts, due in part to their low cell numbers and unique monolayer distribution in the pulp chamber. Moreover, few proteomics studies have been performed on pulp or dentin and these have mostly utilized 2-D gel-based proteomic analysis. Thus, the odontoblast proteome is unknown and our previous

reverse-phase liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of human dental pulp represents the largest data set of any dental tissue to date (Eckhard et al., 2015). Here we developed a technique to extract odontoblast cell layer protein for proteomic analysis from the limited numbers of cells present in the human odontoblast layer. We compared this proteome with paired analyses of healthy dental pulps from prophylactic 3rd molar extractions.

2.2 Materials and methods

2.2.1 Pulp harvest procedure and protein extraction

Informed consent was obtained from patients according to protocols approved by the University of British Columbia Clinical Research Ethics Board. Human dental pulps were collected from healthy, caries-free 3rd molars of two female and seven male patients, ranging from 15 to 39 years of age (Table 1). All extractions were prophylactic. Immediately after extraction, the teeth were partially sectioned vertically or axially with a high-speed dental turbine under water spray, to minimize heating or traumatizing pulp tissue, and split in two with a dental elevator exposing the pulp. Using sterile curettes and barbed broaches, tissue from the pulp chambers and coronal thirds of root canals was harvested and placed in 250 µL of protein denaturant solution (8 M GuHCl) and transported on dry ice. After weighing the pulp samples, 8 M GuHCl was added to 500 µL and the samples were stored at -80 °C. For sample preparation, the pulps were homogenized using an Ultra-Turrax (IKA Works, Inc.) in 8 M GuHCl and further extracted for 90 min at 22 °C—conditions under which pulp proteins, including proteases, are denatured and inactivated. Dental pulp protein was collected by chloroform/methanol precipitation and dissolved in 500 µL of 8 M GuHCl, as described previously (Eckhard et al., 2015). The protein concentration was determined using the Bradford assay (Bio-Rad) with bovine serum albumin as a standard. To extract protein from the odontoblast cell layer after removal of the pulp stroma

we employed a modified crucible technique that we had pioneered previously (Tjäderhane et al., 1998) and had used to show that human odontoblasts remain firmly attached to the dentin and so can be cultured for several days upon addition of culture medium. Following a similar procedure, the four decoronated and depulped 3rd molars, consisting of the root trunk apical to the cementoenamel junction and the attached tooth roots, were vertically mounted in beading wax to create a closed system where the canal foramina were occluded. The pulp chambers and root canal systems were filled with ~100 µL 8 M GuHCl denaturant solution and the odontoblast cell layer protein, with potentially some nonmineralized dentin matrix proteins, was extracted overnight at 22 ºC (Fig. 1A). Following two 50 µL 8 M GuHCl washes, the denaturant solution and washes were pooled for each donor, and processed as described above for the pulp stroma samples.

2.2.2 Histological analysis

The pulp from tooth #48, donor 5 was fixed in 4 % paraformaldehyde in PBS overnight at 4 °C. The sample was subsequently washed four times with PBS, dehydrated through a graded series of ethanol washes, and cleared with xylene. The sample was then embedded in paraffin blocks and cut into serial sections of 12 µm thickness. Hematoxylin and eosin (HE) staining and digital image recording were performed by Wax-it Histology Services at the University of British Columbia to confirm the amount of the odontoblast cell layer remaining on typical pulp samples (Fig. 1B).

2.2.3 TAILS N-terminomics

Protein N-termini enrichment by TAILS was performed as follows (Fig. 2) (Kleifeld et al., 2011b): samples were reduced in 5 mM dithiothreitol (30 min, 65 °C) to allow cysteine carbamidomethylation using 15 mM iodoacetamide (45 min in the dark, room temperature).

Excess blocking reagent was quenched by adding 10 mM dithiothreitol (30 min, room temperature). The pH was then adjusted to 6.5, to allow reductive dimethylation of primary amines, with 40 mM heavy formaldehyde ($^{13}CD_2O$ in D_2O; Cambridge Isotopes) and 20 mM sodium cyanoborohydride (overnight at 37 °C). Importantly, blocking of primary amines was performed at the whole protein level before trypsin digest. Thus, protein alpha amines of free protein N-termini were labeled, thereby allowing for their enrichment and discrimination by TAILS from internal tryptic peptides. Following overnight incubation, 20 mM heavy formaldehyde and 20 mM cyanoborohydride were readded (2 h, 37 °C) to ensure complete amine labeling. Reactions were then quenched using 100 mM Tris-HCl, pH 6.8 (30 min, 37 °C), and reagents and salts removed by chloroform/methanol protein precipitation (Wessel and Flügge, 1984). Polypeptide pellets were dissolved in 50 mM NaOH, adjusted to pH 7.5 with 100 mM 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid, 250 µL, diluted 1:1 with HPLC-grade water, and digested with mass spectrometry-grade trypsin (Trypsin Gold, Promega) at 100:1 (w/w) protein:enzyme ratio. Following overnight incubation at 37 °C, digestion efficiency was confirmed with SDS-PAGE. 10 % of the tryptic digest was analyzed by shotgun proteomics "preTAILS" analyses (~50 µg protein). To separate trypsin-generated internal and C-terminal peptides and to allow for enrichment of the N-terminome, the TAILS samples were adjusted to pH 6.5 and water-soluble HPG-ALD polymer (http://flintbox.com/public/project/1948/) was added (5x peptide mass w/w) in 20 mM sodium cyanoborohydride, pH 6.8 (37 °C, overnight) to covalently link the tryptic peptides via their free primary alpha amines to the polymer. The coupling reaction was quenched with 100 mM tris(hydroxymethyl)aminomethane buffer, pH 6.8 (30 min, 37 °C), and both the naturally blocked (*e.g.* acetylated or pyro-Glu) and experimentally labeled N-terminal peptides (dimethylated N-terminal peptides) were collected by ultra-filtration (Amicon Ultra-0.5, MWCO 10 kDa). The internal tryptic and C-terminal peptides covalently bound to the polymer remained trapped on the filter. All blocked N-terminal peptides comprising

the N-terminome (TAILS) and shotgun (preTAILS) samples were desalted using C18 STAGE-tips (Rappsilber et al., 2007) and frozen in liquid nitrogen until analysis by LC-MS/MS according to (Eckhard et al., 2015).

2.2.4 Mass spectrometry

Peptide samples after C18 STAGE-tip purification were analyzed using an Accurate Mass G6550A quadrupole time-of-flight (Q-TOF) mass spectrometer (Agilent), coupled on-line to a nanoflow HPLC (Agilent 1200 Series) with a Chip Cube nanospray ionization interface (Agilent). A high capacity HPLC-Chip (Agilent) with 160 nL enrichment column and a 0.075 mm x 150 mm analytical column (Zorbax 300SB-C18, 5 µm) was used. Each sample was loaded on the enrichment column at a flow rate of 4 µL/min (buffer A; 0.1 % formic acid) and with a 4 µL injection flush volume. After that, a 110.2 min gradient was established at 300 nL/min; first from 0 % to 5 % buffer B (99.9 % acetonitrile, 0.1 % formic acid) over 2 min, then from 5 % to 45 % buffer B over 78 min, then increased to 60 % over 10 minutes, and brought in one step (0.1 min) to 95 % buffer B, held at 95 % for 20 min, and then reduced to 3 % buffer B again (0.1 min) to recondition the column for the next analysis. Peptides were ionized by electrospray ionization at 1.8 kV, and MS-analysis was performed in positive polarity with precursor ions detected between 300 and 2000 *m/z*. The top three ions per scan were selected for collision induced dissociation (CID) using a narrow exclusion window of 1.3 atomic mass units and at a MS/MS scan rate of two spectra per second. The collision energy was adjusted automatically depending on the charge state of the parent ions, and precursor ions were then excluded for 30 s from further CID. The LC-MS/MS system was controlled by Mass Hunter version B.02.01 (Agilent).

2.2.5 Data analysis

Agilent LC-MS/MS raw data were converted to mgf and mzXML files using MSConvertGUI (Chambers et al., 2012) so that the acquired spectra could be matched to peptide sequences in the human UniProt protein database (http://www.uniprot.org, release 2013_10) using four dedicated search engines, Mascot v2.4 (Perkins et al., 1999), X! Tandem CYCLONE TPP 2011.12.01.1 (Craig and Beavis, 2004), MS-GF+ v10072 (Kim and Pevzner, 2014), and Comet 2015.01 rev 0 (Eng et al., 2013). Search criteria were set at a semi-ArgC cleavage pattern using the same parameters as in our earlier study (Eckhard et al., 2015). Identified peptide-spectrum-matches (PSMs) were statistically evaluated using PeptideProphet (Keller et al., 2002), and combined using iProphet (Shteynberg et al., 2011) within the Trans-Proteomic Pipeline (TPP) v4.8.0 PHILAE (Deutsch et al., 2015); a 1 % false discovery rate (FDR) cut-off was applied. Peptides were assigned to proteins using the ProteinProphet (Nesvizhskii et al., 2003) module within TPP, and protein probability was set to ≥ 0.95, equivalent to a FDR of ≤ 1 %. Where peptides matched multiple protein sequences, ProteinProphet chose one protein entry as representative.

2.2.6 Bioinformatics analysis

For a peptide to be classified as a valid N-terminus it required both a blocked N-terminus and a C-terminal arginine. Note that while trypsin normally cleaves C-terminal of both arginine and lysine residues, lysines in the TAILS workflow were dimethylated at the protein level and so were protected from trypsin cleavage. Blocked true N-termini were classified as those with either: 1) an N-terminus carrying a heavy dimethyl label, indicative of a free amino terminus before trypsin digest, or 2) an acetylated N-terminus or pyro-glutamate N-terminus, both of which are enzymatically mediated modified N-termini. Two other blocked N-terminal peptides were selected for by the TAILS protocol, but do not necessarily represent protein N-termini as

they mainly result from spontaneous N-terminal cyclization of glutamine (Gln->pyro-Gln) or carbamidomethylated cysteine (Cys -> pyro-Cys). These cyclizations are known side-reactions of tryptic peptides under the applied experimental conditions, but are useful to improve protein identification since they are valid predominantly internal peptides. Protein identifications of so-called “missing” proteins were based on neXtProt classes PE2-PE5 (www.nextprot.org, release 2014-09-19) and rechecked against release 2017-04-12. Thus, due to the risk of missed assignment of internal tryptic peptides commencing with a cysteine or glutamine as N-termini, this stringency of analysis excludes potential *bona fide* natural N-termini commencing with Gln or Cys from being further considered.

2.3 Results

2.3.1 Protein extract sample summary

The homogeneity of the central pulp stroma and its separation from the majority of the odontoblast cell layer was confirmed by histology (Fig. 1B,C). As the pulp tissue sample contained proportionally a very small number of remnant odontoblasts, the proteomic data needs be interpreted with this minor caveat. A total of 39 LC-MS/MS analyses were performed: 10 of the odontoblast cell layer and 29 of pulp stroma (Table 1).

2.3.2 Proteomic workflow and statistical analysis of mass spectra data

Categorization of the N-terminome confirmed high enrichment of N-terminal peptides by TAILS from the internal semi-Arg(C) peptides (Fig. 2) and the figure also shows the relative proportions of the various N-terminal peptides. Figure 3 provides an overview of the numbers of peptides and proteins identified at the various steps in the bioinformatics workflow. Thus, the PSM matches by each of the 4 search engines in the pulp stroma (12,063 nonredundant peptides) and odontoblast cell layer (4,888 nonredundant peptides) before statistical modeling using the

TPP are shown in Figure 3A,D. The PSM identifications were then statistically verified using PeptideProphet for the preTAILS shotgun and TAILS N-terminomic analyses and combined

Table 1. Pulp Protein Sample Data

Donor Age/ Sex $n = 9$		Pulp stroma					Odontoblast Cell Layer			
		# of pulps	Total tissue weight (mg)	Number of LC-MS/MS Data Sets		Unique Peptides / Total	# of pulps	Number of LC-MS/MS Data Sets		Unique Peptides / Total
				Pre-TAILS $n = 11$	TAILS $n = 18$			Pre-TAILS $n = 5$	TAILS $n = 5$	
1	18 ♂	1	16.9	2	2	2,571 / 8,399				
2	18 ♀	4	89.5	1	1	3,882 / 12,054	1	1	1	2,941 / 10,046
3	27 ♀	3	25.9	1	1	4,434 / 12,359				
4	30 ♂	2	8.18	1	1	2,639 / 9,892	2	2	2	2,361 / 7,667
5	19 ♂	4	119	2	5	4,332 / 37,436				
6	39 ♂	1	9.34	1	1	2,755 / 10,351	1	2	2	2,303 / 7,609
7	19 ♂	4	99	1	1	2,062 / 5,921				
8	18 ♂	3	233	1	3	3,381 / 14,531				
9	15 ♂	4	424	1	3	3,169 / 14.175				

Table 1. Tooth donor sex, age, and corresponding number of acquired LC-MS/MS shotgun preTAILS and TAILS data sets of donor dental pulp and odontoblast cell layer protein. Numbers of unique peptides and total peptides are the sum of PreTAILS shotgun + TAILS peptides from each donor sample identified by all of the four search engines (Mascot, X! Tandem, MS-GF+, and Comet) and having a PeptideProphet output with an FDR ≤ 1 %.

using iProphet for the pulp stroma (total of 11,314 peptides; 2-way Venn diagrams) and odontoblast cell layer (4,648) (Fig. 3A,D; 2-way Venn diagrams). From positional information, the total numbers of protein N-termini were calculated after using iProphet to combine the data sets (Fig. 3B,E; 2-way Venn diagrams), which were assigned to parent proteins from direct interpolation of the peptide data (Fig. 3C,F). Finally, true protein identifications were assigned by ProteinProphet at an FDR ≤ 1 % (Fig. 4A-C). TAILS is highly sensitive. Thus, despite the odontoblast cell layer proteome being prepared from a very small number of dentin-adherent odontoblasts and detached tubules after removal of the pulp stroma, we identified one third as many odontoblast proteins (895) as pulp stroma proteins (2,423) (Fig. 4C).

2.3.3 Protein identifications

The identification of 2,974 pulp proteins here (Fig. 4A), when combined with our earlier study (Eckhard et al., 2015), have greatly expanded the annotation of the pulp proteome by 5,097 unique proteins (Fig. 4B). This represents a major contribution to the total known pulp protein numbers identified by mass spectrometry (5,190), as revealed by comparing our combined annotation with two earlier studies from other groups (Fig. 4A,B). However, these previous studies were made using 2D gel proteomics performed with one search engine alone and without statistical modeling at a FDR < 5 % (Eckhardt et al., 2014; Pääkkönen et al., 2005). Note, that due to the increased stringency applied in the identification of peptides and proteins by statistical modeling in our multistep bioinformatics pipeline, the total numbers of peptides and proteins confidently identified was reduced from that initially obtained using search engines alone (*see* Fig. 3A,C,D,F).

Comparison of our odontoblast cell layer proteome data with a previous dentin proteome report (Jágr et al., 2012) reveals similarities and differences between the proteins found in the odontoblast cell layer and dentin (Fig. 4C). We also compared the odontoblast cell layer

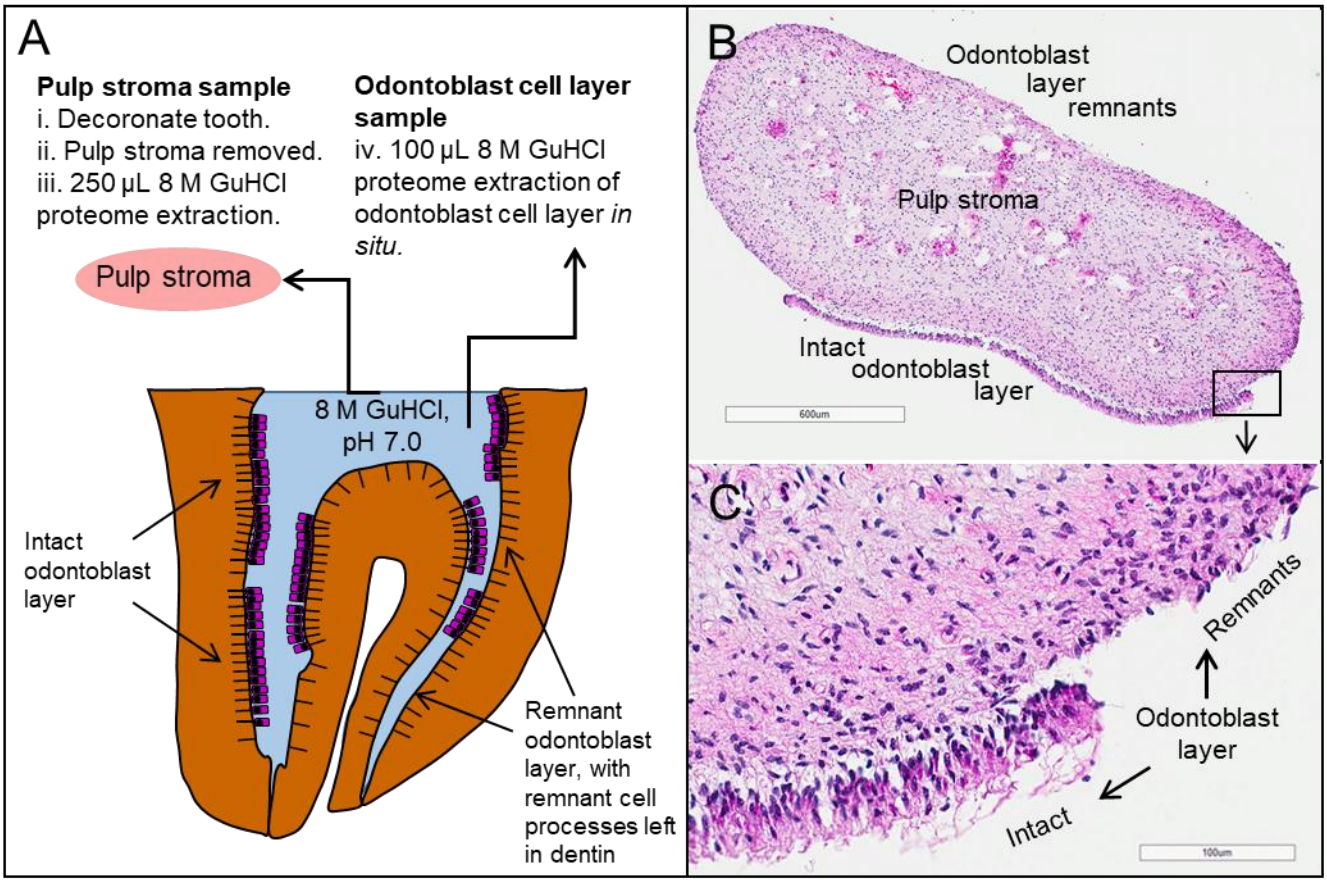


Figure 1. Protein extraction and histological confirmation

(**A**) Protein extraction steps (i to iv) applied to the dental pulp and odontoblast cell layer of third molars in this study. (**B**) H & E stain of axial section of pulp tissue from distal root of tooth #48, donor 5. The majority of the odontoblast layer remained attached to the dentinal walls, but in some areas remnants of odontoblast layer were found attached to the pulp tissue. Bar, 600 µm. (**C**) High power detail of boxed area in B. Bar, 100 µm.

proteomes of donors younger/older than 20 years, showing differences, but with the caveat that undersampling may have skewed this comparison (Fig. 4D). Appendix Figure 6 shows the total peptide spectra matches for the four search engines. Appendix Figure 7 shows the breakdown of naturally occurring N-termini, compared with neo-N-termini that are the result of unknown proteolytic processing or incidental cleavages. Almost two thirds of N-termini are classified as neo-N-termini in both the odontoblast and pulpal stromal samples, indicating proteolytic events that are as yet unexplained. A similar pattern was seen with N-termini in our earlier study (Eckhard et al., 2015). Appendix Table 1 lists the 211 proteins found in the odontoblast sample (Fig. 4C) that were not identified in the pulpal stroma or dentin. Appendix Tables 3, 4, 5 list proteins found in the young pulp odontoblast sample *versus* those found in the mature pulp odontoblast sample, and those proteins found in the odontoblast samples of both young and mature pulps (Fig. 4D). Uploaded to ProteomeXchange are the complete peptide and protein lists and all mass spectrometry data from each experiment and mass spectrometry analysis under the PXD identifier <PXD006557>

2.3.4 Comparison with other tissue matrisomes

We compared the extracellular matrix proteins found in the odontoblast cell layer and the dental pulp with tissue distributions of other extracellular matrix proteins in the so called “matrisome” (Naba et al., 2016) (Appendix Tables 3 and 4). Dentin is a unique human connective tissue and one that has dentin specific proteins, such as dentin sialophosphoprotein (MacDougall et al., 1997), as well as proteins common to mineralized bone, *e.g.* osteonectin, osteocalcin, osteopontin, and dentin matrix protein 1. NeXtProt and Peptide Atlas do not list dentin sialophosphoprotein as having been identified by mass spectrometry. Here, we identified the highly charged dentin sialophosphoprotein in the odontoblast cell layer by both N-terminal

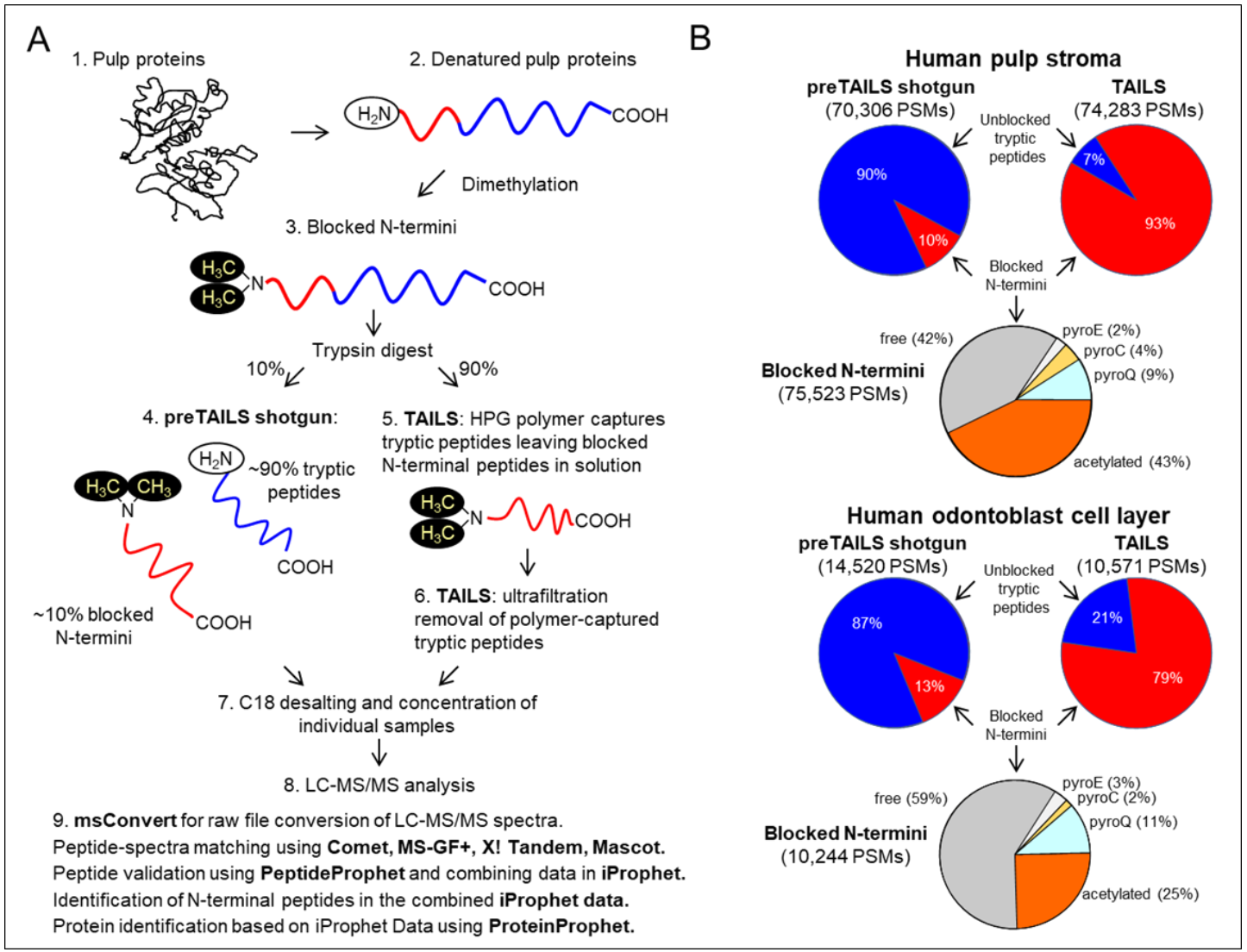


Figure 2. TAILS workflow and Peptide Classification

Figure 2. (**A**) TAILS workflow. **1**. Tissue samples were extracted in 8 M GuHCl denaturant buffer. **2**. Denatured protein (in blue) schematic shows the N-terminal peptide in red. **3**. N-terminal α-amines of peptides and ε-amino groups of lysine side chains were dimethylated with isotopically labelled formaldehyde. The denatured proteins were trypsinized, cleaving C-terminal to arginine residues only as the lysines were blocked. **4**. ~10 % of the sample was directed for preTAILS shotgun proteomic analysis. **5**. To enrich for protein N-terminal peptides, the HPG-ALD polymer was used to covalently bind and remove the internal tryptic peptides. **6**. Upon subsequent ultrafiltration, the blocked N-terminal peptides were collected in the unbound fraction whereas the internal tryptic peptides bound to the polymer were retained on the filter. **7**. preTAILS shotgun proteome and TAILS N-terminome samples. **8**. Samples submitted for LC-MS/MS analysis. **9**. Bioinformatic analyses of the aquired peptide fragmentation spectra. (**B**) Classification of peptides identified by preTAILS shotgun and TAILS N-terminome analyses showing the high enrichment of N-terminal peptides by TAILS (red) from the internal tryptic and C-terminal peptides (blue). PSM, Peptide Spectra Match, whereby LC-MS/MS spectra are matched by search engine software to the mass-to-charge-ratio (*m/z*) and the fragmentation pattern of a peptide sequence in the Uniprot database by the four search engines shown. The pie chart shows the relative proportions of the five main types of N-terminal peptide identified: peptides with experimental conjugated dimethylation (denoted as 'free', unblocked natural N-termini), natural acetylation (predominantly at positions 1 and 2 of the protein chain), or the three common N-terminal cyclizations of glutamate (pyroE), glutamine (pyroQ) or cysteine (pyroC).

peptides and shotgun tryptic peptides. Confirming our recent report (Eckhard et al., 2015), we also identified dentin sialophosphoprotein in dental pulp by N-terminal and internal tryptic peptides, where dentin sialophosphoprotein most likely derives from the remnant odontoblasts in the dental pulp sample. Osteonectin, osteocalcin, osteopontin, dentin matrix protein 1 were not identified in the odontoblast cell layer, possibly reflecting completed dentin formation in the portion of the donor tooth analyzed at the age sampled.

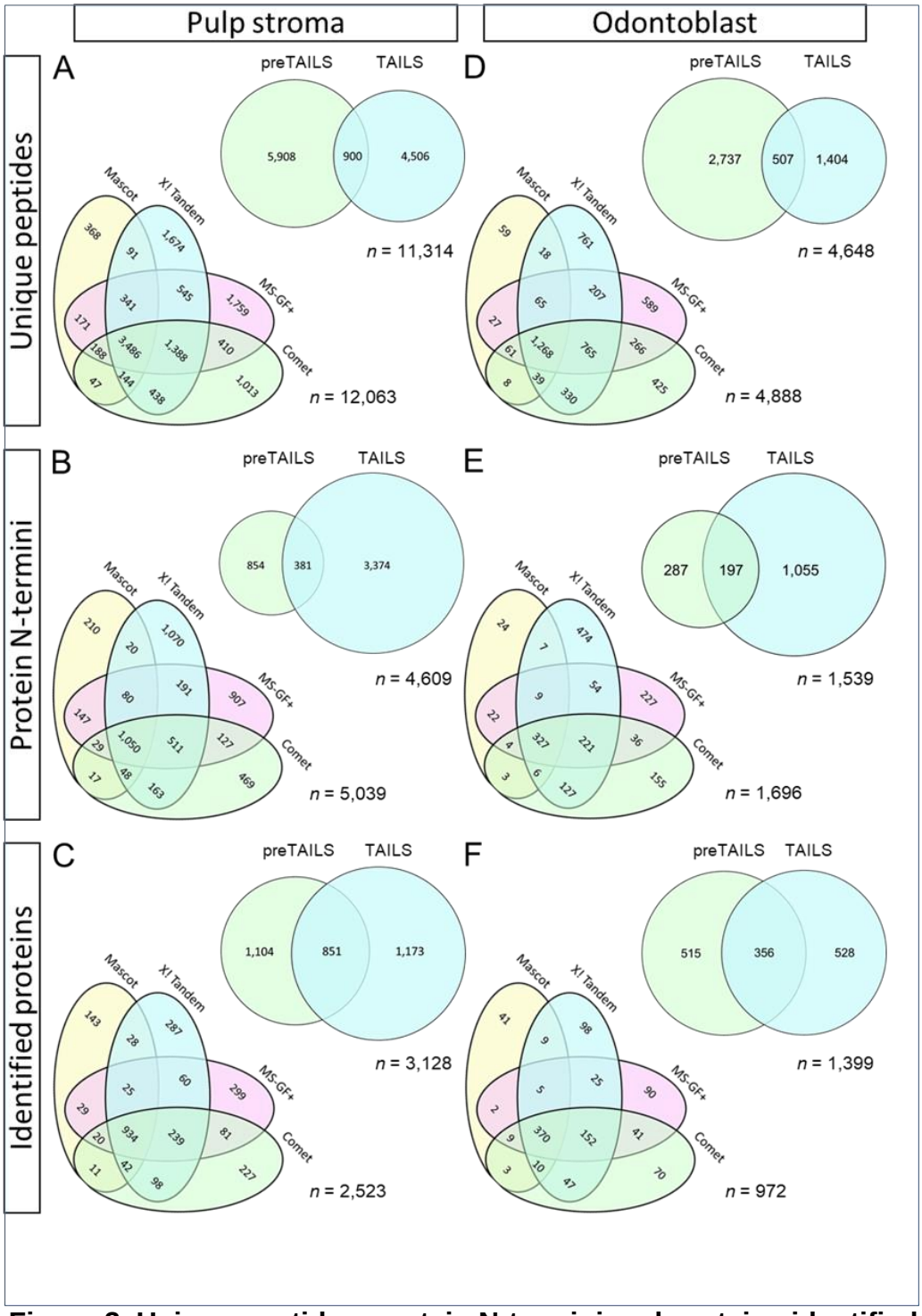


Figure 3. Unique peptides, protein N-termini and proteins identified

Figure 3. Compares unique peptides, protein N-termini, and proteins identified in odontoblast cell layer and pulpal stroma extracts. The 4-way Venn diagrams show the peptide and protein identifications filtered by PeptideProphet to a FDR ≤ 1 % from the 4 search engines; the 2-way Venn diagrams show the unique peptides, protein N-termini, and proteins identified by the preTAILS and TAILS analyses and analyzed by iProphet to a FDR ≤ 1 %. The number of proteins shown in the 2- and 4-way Venn diagrams is an estimate extrapolated from the peptide data with an ~ FDR ≤ 5 % at this initial step in the bioinformatic pipeline we used.

2.3.5 Proteomic evidence of four 'missing' proteins

We found proteomic evidence for 4 of the current 3,134 proteins listed as "missing proteins" in the neXtProt knowledge base (Fig. 5), the official knowledge base of the Human Proteome Project of the Human Proteome Organization (HUPO) (Deutsch et al., 2016). Missing proteins are defined as those without protein existence (PE) at the protein level (PE1), whereas PE2 proteins have only been detected by transcript evidence alone. Figure 5 shows four high confidence, previously unidentified (PE2) or only recently identified (PE1), proteins we found in the dental pulp samples. The high confidence, FDR ≤ 1 %, TAILS peptides are shown in red, whereas peptides found by preTAILS shotgun analyses are shown in blue. Biological replicates, i.e. finding the same peptide in different donors, increase the probability that a protein has been identified, as do technical replicates, i.e. finding the peptide multiple times in the same donor (Deutsch et al., 2016). Thus, the effectiveness of an orthogonal technique, such as TAILS, is revealed in increasing proteome coverage and identifying unique peptides and proteins, which were not found in shotgun proteomics approaches (Fig. 5). The four "found" missing protein candidates are: uncharacterized protein C14orf105 (Q9NVL8), existence is

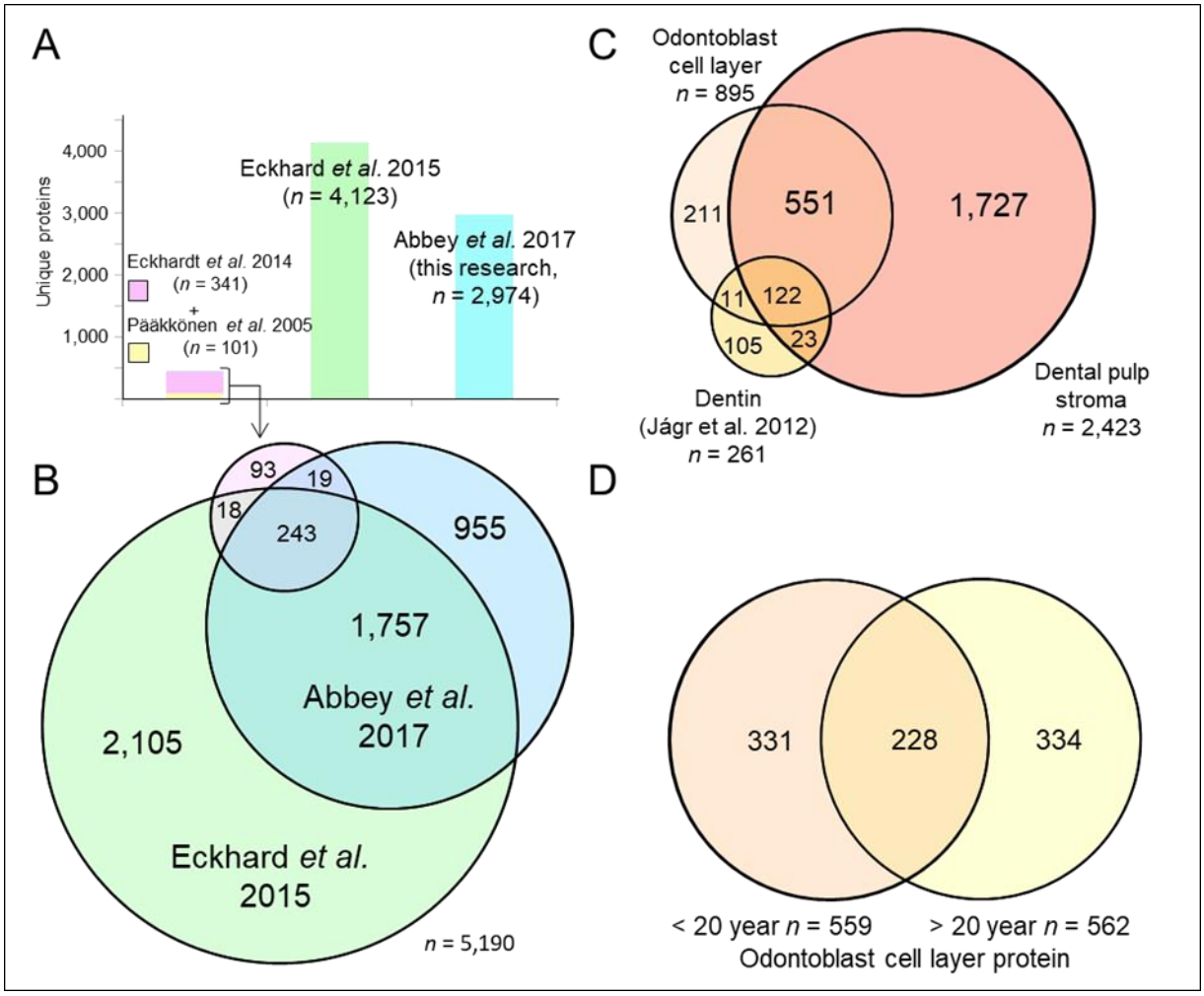


Figure 4. Comparison of dental pulp proteomic studies

Figure 4. (**A**) Comparison of our present analyses with the 3 previous dental pulp proteomics studies in the literature. The left column (pink/yellow) is the sum of two 2D gel studies (452 proteins). The green and blue bars are protein numbers from our recent (Eckhard *et al.* 2015) and present study after ProteinProphet analysis, FDR ≤ 1 %. In A and B we used ProteinProphet to identify proteins with an FDR ≤ 1 % using all 39 odontoblast (10) + pulpal stroma (29) data sets combined, which gave 2,974 unique proteins for dental pulp. In comparison, when the odontoblast cell layer and pulpal stroma data sets were assessed individually (10 *vs.* 29 data sets) fewer proteins are identified (*e.g.* panel C) as less peptide evidence is used to generate the statistical model in each of the separate analyses. (**B**) Overlap in total proteins identified by the 4 studies. The present study identified 955 proteins not previously identified in dental pulp. (**C**) Venn diagram showing a comparison of the odontoblast cell layer proteome with dental pulp stroma and a proteomic analysis of *dentin* by Jágr *et al.* 2012. When the 10 data sets from the odontoblast cell layer (≤ 20 years and > 20 years) were combined for ProteinProphet analysis (FDR ≤ 1 %) 895 proteins were identified. (**D**) Comparison of the odontoblast proteome of pulps from donors younger than age 20 with those of donors older than 20 after ProteinProphet analyses (FDR ≤ 1 %). Each data set for the two age groups were considered separately (2 *vs.* 8 data sets) generating a total of 893 unique proteins identified.

based on evidence at the transcript level (PE2); axonemal dynein heavy chain 6 (Q9C0G6), existence is based on evidence at the protein level (PE1); leucine-rich repeat-containing protein 9 (Q6ZRR7), existence is based on evidence at the transcript level (PE2); and WD repeat-containing protein KIAA1875 (A6NE52), an entry whose protein(s) existence is based on evidence at the protein level (PE1), but for which there is no functional information available. Of these dynein heavy chain 6 is noteworthy as this protein is considered to be a force generating protein of respiratory cilia with ATPase activity involving microtubules and so may play transport roles in the long odontoblast process extension in dentinal tubules.

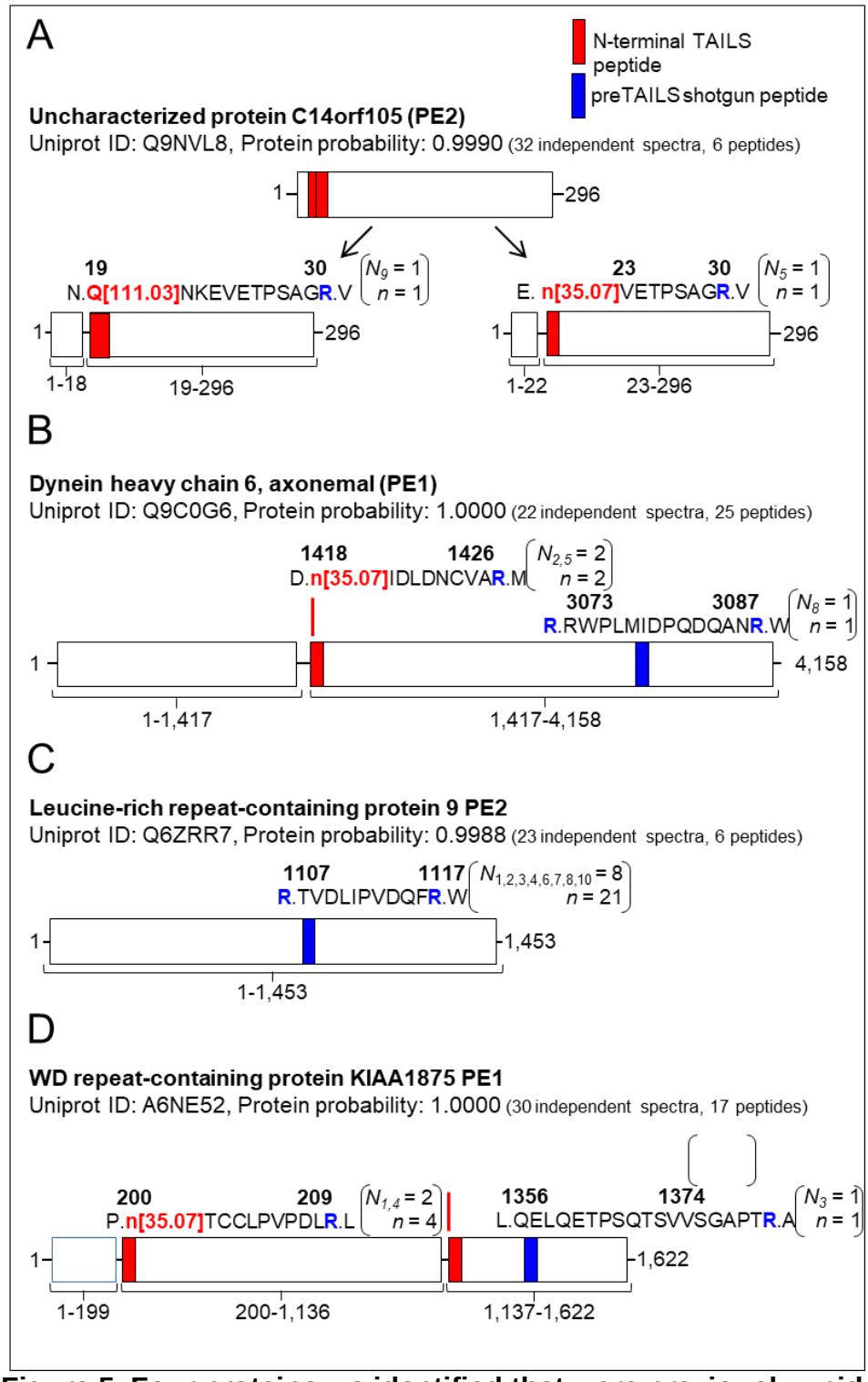


Figure 5. Four proteins we identified that were previously unidentified (neXtProt)

Figure 5. Four proteins we identified that had not been previously identified by proteomic analyses. The relative position of the high confidence N-terminal (red) and internal tryptic (blue) peptides identified with a FDR ≤ 1 %. *N*, number of biological replicates; subscript, donor identifier numbers (*see* Table 1) in which the peptide was identified; *n*, number of technical replicates in which the peptide was identified. Numbers indicate amino acid residue position. PE1, neXtProt protein existence 1, where the protein has been identified at the protein level (*e.g.* by antibodies), but in the examples shown, not previously by mass spectrometry; PE2, neXtProt protein existence 2, where the protein has been identified at the transcript level only. (**A**) Protein C14orf105 (PE2 neXtProt), two different nested N-terminal peptides were identified starting at position 19 or position 23 and both extending to Arg30. (**B**) Axonemal dynein heavy chain 6, (PE1 neXtProt), identified with two high confidence peptides found 3 times in 2 donors. (**C**) Leucine-rich repeat-containing protein 9 (PE2 neXtProt) was identified by a single peptide, but this was found 21 times in 8 donors, giving high confidence of the protein identification. (**D**) WD repeat-containing protein KIAA1875 (PE1 neXtProt), 4 high confidence peptides were found in 3 donors. Note: only the highest confidence peptides (FDR ≤ 1 %) are shown. **R** (in blue), arginine.

2.4 Discussion

2.4.1 Progress in pulpal proteomics

Despite the limited amount of odontoblast cell layer sample, we generated high quality spectra sufficient to identify nearly 900 human odontoblast cell layer proteins, the first such analysis of odontoblasts. Moreover, our present proteome analysis of human dental pulp in combination with our previous study (Eckhard et al., 2015) has identified in total 5,097 human dental pulp proteins (FDR <1 %), and represents the largest data set of human pulp stroma proteome to date. Unlike pulp stromal fibroblasts, odontoblasts have a specialized role producing and maintaining the mineralized organic matrix of human dentin, yet, odontoblast cell function is incompletely understood. For example, how odontoblasts participate in pulp sensation or

secrete dentinal fluid into dentinal tubules remain open questions. Indeed, the odontoblast process that extends into dentinal tubules has no biologic equivalent (Hargreaves et al., 2012). We found that the odontoblast layer had a distinct proteomic composition in comparison with both the adjacent dentin and pulp stromal tissue. This confirmed functional differences between these two areas of the dental pulp. In addition to fibroblasts, resident immune and stem cells, the pulp stroma also contains cells of other tissues, *e.g.* blood vessels and nerve axons. Thus, these cells also contribute proteins to the dental pulp proteome in addition to extracellular matrix proteins and proteins secreted or shed from the pulp fibroblasts, as listed in the accompanying appendices and in the uploaded publically accessible proteome data sets in the ProteomeXchange database.

2.4.2 Odontoblasts and tooth development

The primary role of the odontoblasts and pulp in childhood is tooth development, i.e. formation, eruption and continued root growth, whereas, in adulthood, the role shifts to dentin maintenance. The donor ages (15-39) here enabled limited observation of differences in the pulp proteome across several developmental stages of the 3rd molar teeth we collected. Developmental timing is variable, but the clinical crown of 3rd molars is usually completed between ages 12-16, erupting between age 17-21, with root development completed by 18-25 years (Nelson, 2015). Therefore, we divided our samples into two groups, ≤ 20 years and > 20 years. In the first group, with an actual donor age of 15 to 19, active developmental processes of the root apical region were taking place, with continued rapid deposition of secondary dentin in the clinical crown. However, in the root trunk and coronal one third of the root from where we extracted the odontoblast cell layer proteins, the majority of dentin formation was complete. In the second group, donor ages 27, 30 and 39, all tooth development processes were complete. Although our study had a relatively small sample size, proteins found in the age subgroups

examined may direct further study as to their possible roles in tooth development (Appendix Table 2).

2.4.3 Mineralization proteins and donor age

Dentin sialophosphoprotein, the most abundant noncollagenous protein in dentin extracellular matrix and important in dentin mineralization, was identified only in the younger pulps, indicating active mineralization in these teeth *versus* older teeth, as expected. However, other mineralized extracellular matrix proteins were not identified in the limited amount of human odontoblast cell layer sampled in the root trunk and upper third of the root, *e.g.* dentin matrix acidic phosphoprotein I, osteonectin, osteocalcin, and osteopontin (Hargreaves et al., 2012). Some proteins are recalcitrant to mass spectrometry due to an unfavorable distribution of acidic residues, which do not favor ionization, or of the basic residues that are cleaved by trypsin in proteomic workflows. Alternatively, the absence of evidence for these proteins may be due to the resting state of the odontoblast cell layer present in the coronal one third of the root canals extracted, which may not have been expressing these proteins after root formation here was completed. Although age-related changes in the proteins of the odontoblast cell layer were evident, further analyses at appropriate donor age and developmental stage may be required to identify other mineralized extracellular matrix proteins in odontoblasts by mass spectrometry. In comparison, the peptide and protein identifications comparing younger and older donors for the pulpal stroma was highly overlapped revealing little to no significant proteome changes with age (data not shown, Abbey et al. 2017, unpublished).

2.4.4 Key findings from my research

Appendix Table 1 lists the 211 proteins that were found exclusively in the odontoblast cell layer when compared with the pulp stroma and dentin proteomes. Of these, we identified proteins that

have been associated with neural function, *e.g.* neuroplastin (Q9Y639) (Beesley et al., 2014), DMX-like protein 2 (Q8TDJ6) (Nagano et al., 2002), and advillin (Q8NEN9) (Marks et al., 1998), all of which were found only in the odontoblast sample; this may reflect odontoblast sensory roles or the close association between the odontoblast layer and nerve endings which cross over the cell free zone to terminate in the odontoblast layer (Hargreaves et al., 2012). Exclusive to the odontoblast tissue sample, we also found retinal guanylyl cyclase (Q02846), which has been associated with vision (Duda et al., 1999). This protein as well, as olfactory and taste receptor proteins we identified previously (Eckhard et al., 2015), may play roles in dentin sensation mediated by odontoblasts, neurons or both.

Odontoblasts also perform a secretory role, producing a constant distal flow of dentinal fluid that moves from the pulp down the dentinal tubules. Several ion transport and secretion-related proteins were found only in the odontoblast samples, for example: inositol 1,4,5-triphosphaste receptor type 1 (Q14643), which has been shown to regulate epithelial secretion of electrolytes in mice (Park et al., 2013); lactotransferrin (P02788), an iron binding anion-transport protein secreted in tears (Azkargorta et al., 2015) and breast milk (Giansanti et al., 2016); and salivary acidic proline-rich phosphoprotein 1/2 (P02810), a major component of human saliva (Wu et al., 2014)—any one of which may be components of, or involved with the secretion of, dentinal fluid.

Discovery studies such as this one inform on the biology of the odontoblast and pulp, and may suggest, for example, possible cellular roles in the timing of the formation of dentin matrix, mediating tooth sensation, and secretion of dentinal fluid. Future studies may also show how the pulp proteome responds and adjusts to different events in the lifecycle of a human tooth and clarify the contribution of individual proteins in pulp function and in pulp to odontoblast cell-cell signalling. Although proteomic analyses of dentin in the past have generated limited protein identifications, present ongoing analysis of human dentin using modern high accuracy mass

spectrometers promise to also increase molecular knowledge of the dentin proteome and potentially reveal new dentin-specific proteins, some or all of which are derived from odontoblasts. This will increase knowledge of how the pulpal stroma and odontoblast layer cooperate in the formation and biology of this unique mineralized tissue.

Chapter 3: Conclusion

3.1 Overview

I successfully extracted protein from the odontoblast cell layer remaining on the walls of the pulp chambers using our crucible technique to enrich for the odontoblast proteome. This is the first time a distinct proteome has been shown to exist for the odontoblast cell layer and the 222 unique dental pulp proteins found to exist only in this layer can guide future research in discrete odontoblast functions. I expanded the dental pulp proteome annotation by identifying four previously missing proteins that had not been detected before at the protein level. These findings are an important contribution to the Human Proteome Project. I also succeeded in expanding annotation of the pulpal stroma tissue by 955 unique proteins using similar techniques to those used in our earlier study (Eckhard et al., 2015). All of these findings expand our knowledge of protein function and human biology, and, more specifically, pave the way to better understanding of dental pulp biology.

3.2 Comparing our preTAILS shotgun + TAILS protocols to 2D gel separation

In total, I identified 2974 pulp proteins in the combined samples (preTAILS shotgun + TAILS for the odontoblast + pulp stroma samples). Combined with our 2015 paper, this produces the largest human dental pulp protein annotation to date, of 5097 distinct proteins. These large numbers of proteins, compared with earlier outdated 2D gel proteomics investigations, are due to our TAILS and preTAILS techniques examining the entire tissue protein extract while gel studies analyzed only the limited areas of gels where protein dots appeared. The gel separation technique is valid and effective, as discussed in chapter 1, and proteins were found in dentin through these studies, which I did not identify in the odontoblast cell layer, specifically osteonectin, osteocalcin, and dentin matrix protein I (Jágr et al., 2012). However, gel studies

have more limitations than advantages since proteins that are few in number will be lost by this technique. Our techniques rely completely on the liquid chromatography separation and MS identification of peptides. With 2D gel separation, one has the additional confirmation of a given protein due to molecular weight and charge migration behavior as well as LC-MS/MS identification. However, this is not always the case, since many proteins appear in gels in fragments due to enzyme mediated or incidental cleavage (Jágr et al., 2012).

3.3 Proteomics data analysis

Previous studies on pulp proteomics have primarily used threshold values built into the Mascot search engine to differentiate true peptides and proteins among the *m/z* ratios of MS spectra. We used several search engines: Mascot v2.4, X! TANDEM CYCLONE TPP 2011.12.01.1, MS-GF+ V10072, and Comet 2015.01 rev 0. All these search engines use different valid algorithms to match spectra to peptides, so more peptides will be identified if the spectra are reviewed by more search engines. Figure 3A, D shows how peptides were identified by search engines, showing that different search engines produce different results. A given spectra might be slightly under the threshold for positive identification for one search engine but above it with another. Thus, peptides and proteins identified by more than one search engine are identified with greater certainty; TPP calculates this by combining all search engine data with iProphet. Our LC-MS/MS data was put through stringent analysis by TPP to conform to HUPO standards of peptide and protein identification, which are being improved and standardized with time and were recently reviewed in the Journal of Proteome Research (Deutsch et al., 2016). One problem in comparing different proteomic studies is that different sets of data are interpreted in different ways, so that results are not directly comparable; hence, the standards initiative of the HPP and the protein identification guidelines are updated annually. Note that I strictly adhered to current HPP guidelines in this thesis work, using multiple search engines, processing their

results with PeptideProphet, combining these results with iProphet, and finally processing the iProphet data with ProteinProphet. This data processing workflow compiles all the spectra and their confidence levels to calculate net FDR for peptides and proteins. For example, for a given protein we might have two high quality spectra indicating peptides with an FDR ≤ 1 %, three others with a less perfect match having an FDR ≤ 5 % and five more with an FDR ≤ 15 %. For the best estimate of whether or not the protein is present the TPP considers all of these peptides in calculating a net FDR for the protein. In the study by Eckhardt et al. 2014, peptides were considered valid if they exceeded a Mascot threshold score of 20 (Mascot was the single search engine they used); this works reasonably well for peptides, but then proteins were confirmed if they exceeded a protein score of 60. This would also work fairly well in the 2D gel study since only a small number of proteins were present in the sample. The Mascot program however, is designed to make peptide spectra matches, not statistically match large numbers of peptides to large numbers of proteins. Our protocols removed false positives and more accurately determined proteins with an FDR ≤ 1 %. Using more search engines, I was able to identify more peptides, and by using a stringent software analysis I removed many questionable spectra and their matched peptides, to retain only peptides and proteins for which there was a high confidence level of being present, below FDR ≤ 1 %.

3.4 Pulp sample mass

Also of note is my sample size. Eckhardt et al. used 5 wisdom teeth, from which only the canal pulp was removed; thus, their samples did not contain pulp chamber pulp as in our research where we combined all sections of pulp tissue. They quoted the weight of only one pulp as being approximately 2.5 mg in weight. Our study examined the pulps of 26 teeth from 9 donors with a combined pulp weight of 1024 mg. Using greater sample size and doing more LC-MS/MS analyses (39 in total) was part of the reason I identified more peptides and annotated more

proteins. Also, my technique to remove the pulp tissue immediately after extraction and place the tissue in GuHCl buffer on dry ice proved effective at maintaining tissue protein integrity until samples were denatured and analyzed under experimental conditions. My high yield of proteins annotated shows cellular self-destructive processes of apoptosis were slowed or prevented. Finally, what is probably the main advantage of my approach is that the previous work by Eckhardt et al. used outdated 2D gel separation technology followed by individual tandem MS analyses for individual proteins (cut from their gels), while we used multiple LC-MS/MS analyses for individual donors with at least two from each donor (one preTAILS shotgun, one TAILS). Each of my LC-MS/MS analyses revealed 100s to 1000s of peptides, whereas in studies by Eckhardt et al., Jágr et al., their trypsinized dissolved gel segments would reveal only one or a few proteins that migrated together on the 2D gel separation. Thus, I used orthogonal techniques that provided high mass accuracy identification and I performed more experiments, allowing me to annotate a much higher number of proteins than were found in the previous 2D gel proteomics studies.

3.5 Practical applications of proteomic research

Proteomics is a relatively new dimension of biological research and scientists are eager to reach forward to potential practical applications of our expanding knowledge of specific proteins found in tissues. Such applications may still be some way in the future, however, due to: 1) the complexity of protein function; 2) the large numbers of proteins found in tissues; and 3) the degree of poorly understood proteolytic processing that has been shown to occur (Appendix Fig 7). Small changes in protein numbers or proportions probably result in significant biologic changes, but it is difficult to quantify and assess the relative importance of these. For example, earlier I discussed that Mrojik et al. 2015 found 5 proteins that were upregulated in dental stem cells compared with the two other stem cell populations; Eckhardt et al. identified two of these

up-regulated proteins in their own research; I found these same two proteins, plus two of the other three identified by Mrojik et al., (Tryptophan-tRNA synthetase, and G6PD) but I did not find Ribonucleosidase diphosphonate reductase M1 chain (RIR1_Bovin), listed in UniProt (http://www.uniprot.org) as a bovine herpes virus protein. Eckhardt et al. supports Mrojik et al.'s hypothesis that these proteins, up-regulated compared with two other *in vitro* stem cell cultures, are thus important for dentin regeneration. But G6PD is a basic enzyme in the pentose phosphate pathway, active in all cells. There is no rationale that adding more G6PD to a devitalized root canal system will encourage pulpal regeneration. Several pulp proteomics papers feature tables showing which categories of proteins were identified in their studies, *e.g.* immune function, structural constituent of cytoskeleton, catalytic activity (Eckhard et al., 2015; Eckhardt et al., 2014). This is useful from the perspective of reviewing categories of protein action, but it requires pigeonholing proteins into performing specific, discrete tasks, which is rarely the case. A protein such as collagen type I has a discrete function, it is a structural protein found throughout extracellular matrix tissue and is usually found tightly wound in a triple helical conformation, made possible by its glycine rich composition. However even type I collagen can also be found as gelatin when not in this triple helical form, and its function changes. Reviewing the UniProt and NeXtProt websites for information on the function of specific proteins often shows that a given protein has been identified by different studies involved in many different processes. Jágr et al's discussion of caries resistance due to proteins is interesting and I propose these could be secreted in dentinal fluid. To test such a hypothesis, first we would need a method to collect dentinal fluid from caries prone and caries susceptible individuals. The proteomes of these fluids could be compared and significant differences might be observed.

3.6 Large numbers of proteins to assess

The other problem we encounter as techniques improve is that we find more proteins and the body of data gets bigger. For example Eckhardt et al. identified 37 proteins in pulp and dentin that were not also found in a plasma proteome used for comparison. They discuss in some detail these 37 proteins and suggested they may be good candidates for future endodontic regenerative therapies (Eckhardt et al., 2014). Is this a reasonable possibility? In our study, we found 2,426 proteins in pulp tissue and 895 in the odontoblast cell layer. These proteins are summarized in the supplemental tables and online at ProteomeXchange with the PXD identifier <PXD006557>. I do not propose we found thousands of proteins that should be investigated as potential protein co-factors that will improve endodontic regenerative treatments. Due to space limitations we cannot even discuss the possible functions of all identified proteins. And my numbers are probably still far from the actual number expressed *in vivo* in tissues, thought to number in excess of 10,000. We were however successful in expanding the pulp proteome by nearly a 1000 proteins not found in our earlier study, and, more importantly, we were able to separate the odontoblast cell layer from the pulp stroma and characterize different proteomes in the two tissue samples (Eckhard et al., 2015). This produces important and novel information regarding the proteins expressed by odontoblasts, which are responsible for dentin production and mineralization, proteins that would either not be expressed or be expressed in smaller numbers by the fibroblast rich pulpal stroma. By annotating the proteins found in different tissues, we enlarge our understanding of these tissues and their biologic function.

3.7 Dentin and bone are related tissues

My research can guide future investigations into odontoblast function and the mechanics of dentin production, which may segue into the broader field of bone metabolism, since dentin and bone are closely related. One can consider the dentin-pulp complex as a closely related model

for osteoblast-bone metabolism. The odontoblast unicellular layer is well defined on the pulpal periphery against the pre-dentin border and as we showed, it is possible to analyze this layer. Osteoblasts on the other hand, are distributed throughout bone and separating the proteins produced only by osteoblasts would be more challenging.

3.8 Our samples are enriched, not pure for tissue types

A caveat for my data interpretation is that the pulp stroma sample includes a minor degree of contaminant odontoblasts that detached with the pulp and their cellular proteins will contribute to the pulp stroma data sets. Similarly, when I extracted protein from the odontoblast cell layer remaining on the dentin walls, potentially some protein from the nonmineralized dentin was also extracted. However, this is not considered a major issue as the predentin was originally derived from the odontoblast cell layer. Thus, odontoblast protein contaminants are probably present in the pulpal stroma sample, just as some pulpal stroma proteins are present in the odontoblast sample. Our research could be complemented by odontoblast cell culture studies to see how these proteomes differ from actual tissue extracts, and use these as controls to see how odontoblast protein synthesis could be influenced by different stimulants, *e.g.* TGF-β, BMP-1. However, recovering odontoblasts from the dentin is technically impossible due to the long odontoblast process embedded in the dentin. In any case, cells in culture behave differently to cells in situ and so proteomic analyses of detached cells will also differ.

3.9 Future directions: cell culture studies and western blot analysis

Odontoblast cell culture studies would also be useful to confirm the presence of certain proteins in this cell line, *e.g.* that they do not in fact originate from a nearby cell type. We identified several proteins that have been associated with neural function and sensation: neuroplastin (Q9Y639), DMX-like protein 2 (Q8TDJ6), advillin (Q8NEN9), and retinal guanylyl cyclase

(Q02846), all present only in the odontoblast cell layer sample. These may be involved in the process of odontoblast mediated dental sensation, or the proteins may originate from nerve axons also present in the odontoblast enriched tissue sample. Raising antibodies to these proteins and performing western blot analyses of odontoblast cell cultures or immunohistochemistry could be used confirm if these proteins are indeed synthesized by odontoblasts, potentially leading to improved understanding of how odontoblasts and nerve tissue interact to produce dental sensation.

3.10 Final comments

With an estimated 10,000 or more proteins in each cell of the body, much work remains to be done before we can understand fully when proteins are individually expressed, up- or down-regulated, how and why they are modified, and how they interact with one another to create biologic function and life. This research has taken a step forward and answered several questions: *e.g.* that the odontoblast cell layer has a proteome distinct from that of the pulp stroma, proteomic confirmation that dentin sialophosphoprotein is synthesized in the odontoblast layer; but has also highlighted other questions, *e.g* are the sensation related proteins we found in the odontoblast cell layer sample active in the odontoblast or do they derive from adjacent nerve fibers? And regardless of their location do they in fact function to produce sensation or do they perform other tasks? By mapping out the protein geography of the human body we may someday be able to travel far beyond the confines of the limited scientific understanding of protein function where we are now. Soon the world of proteomics will open our eyes to new understanding of how these fascinating molecules are created, modified and adapted to form and shape life.

Bibliography

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). An Overview of Gene Control.

Anderson, N.L., and Anderson, N.G. (1998). Proteome and proteomics: New technologies, new concepts, and new words. ELECTROPHORESIS *19*, 1853–1861.

Azkargorta, M., Soria, J., Ojeda, C., Guzmán, F., Acera, A., Iloro, I., Suárez, T., and Elortza, F. (2015). Human Basal Tear Peptidome Characterization by CID, HCD, and ETD Followed by in Silico and in Vitro Analyses for Antimicrobial Peptide Identification. J. Proteome Res. *14*, 2649–2658.

Beck-Sickinger, A.G., and Mörl, K. (2006). Posttranslational Modification of Proteins. Expanding Nature's Inventory. By Christopher T. Walsh. Angew. Chem. Int. Ed. *45*, 1020–1020.

Beesley, P.W., Herrera-Molina, R., Smalla, K.-H., and Seidenbecher, C. (2014). The Neuroplastin adhesion molecules: key regulators of neuronal plasticity and synaptic function. J. Neurochem. *131*, 268–283.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. *30*, 918–920.

Chun, S.Y., Lee, H.J., Choi, Y.A., Kim, K.M., Baek, S.H., Park, H.S., Kim, J.-Y., Ahn, J.-M., Cho, J.-Y., Cho, D.-W., et al. (2011). Analysis of the soluble human tooth proteome and its ability to induce dentin/tooth regeneration. Tissue Eng. Part A *17*, 181–191.

Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinforma. Oxf. Engl. *20*, 1466–1467.

Deutsch, E.W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., and Moritz, R.L. (2015). Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics Clin. Appl. 9(7-8):745-54

Deutsch, E.W., Overall, C.M., Van Eyk, J.E., Baker, M.S., Paik, Y.-K., Weintraub, S.T., Lane, L., Martens, L., Vandenbrouck, Y., Kusebauch, U., et al. (2016). Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. J. Proteome Res. *15*, 3961–3970.

Duda, T., Venkataraman, V., Goraczniak, R., Lange, C., Koch, K.W., and Sharma, R.K. (1999). Functional consequences of a rod outer segment membrane guanylate cyclase (ROS-GC1) gene mutation linked with Leber's congenital amaurosis. Biochemistry (Mosc.) *38*, 509–515.

Eckhard, U., Marino, G., Abbey, S.R., Tharmarajah, G., Matthew, I., and Overall, C.M. (2015). The Human Dental Pulp Proteome and N-Terminome: Levering the Unexplored Potential of

Semitryptic Peptides Enriched by TAILS to Identify Missing Proteins in the Human Proteome Project in Underexplored Tissues. J. Proteome Res. *14*, 3568–3582.

Eckhardt, A., Jágr, M., Pataridis, S., and Mikšík, I. (2014). Proteomic Analysis of Human Tooth Pulp: Proteomics of Human Tooth. J. Endod. *40*, 1961–1966.

Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics *13*, 22–24.

Giansanti, F., Panella, G., Leboffe, L., and Antonini, G. (2016). Lactoferrin from Milk: Nutraceutical and Pharmacological Properties. Pharmaceuticals *9*, 61.

Hargreaves, K.M., Goodis, H.E., and Tay, F.R. (2012). Seltzer and Bender's Dental Pulp (Quintessence Pub.).

Henry, M.A., and Hargreaves, K.M. (2007). Peripheral mechanisms of odontogenic pain. Dent. Clin. North Am. *51*, 19–44, v.

Jágr, M., Eckhardt, A., Pataridis, S., and Mikšík, I. (2012). Comprehensive proteomic analysis of human dentin. Eur. J. Oral Sci. *120*, 259–268.

Jágr, M., Eckhardt, A., Pataridis, S., Broukal, Z., Dušková, J., and Mikšík, I. (2014). Proteomics of human teeth and saliva. Physiol. Res. *63 Suppl 1*, S141-154.

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

Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. *5*, 5277.

Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R.K., Starr, A.E., Foster, L.J., Kizhakkedathu, J.N., and Overall, C.M. (2010). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat. Biotechnol. *28*, 281–288.

Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J.N., and Overall, C.M. (2011a). Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nat. Protoc. *6*, 1578–1611.

Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J.N., and Overall, C.M. (2011b). Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. Nat. Protoc. *6*, 1578–1611.

Kumar, G.S. (2011). Orban's Oral Histology and Embryology (Elsevier India).

MacDougall, M., Simmons, D., Luan, X., Nydegger, J., Feng, J., and Gu, T.T. (1997). Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. Dentin phosphoprotein DNA sequence determination. J. Biol. Chem. *272*, 835–842.

Marks, P.W., Arai, M., Bandura, J.L., and Kwiatkowski, D.J. (1998). Advillin (p92): a new member of the gelsolin/villin family of actin regulatory proteins. J. Cell Sci. *111 (Pt 15)*, 2129–2136.

Mrozik, K.M., Zilm, P.S., Bagley, C.J., Hack, S., Hoffmann, P., Gronthos, S., and Bartold, P.M. (2010). Proteomic characterization of mesenchymal stem cell-like populations derived from ovine periodontal ligament, dental pulp, and bone marrow: analysis of differentially expressed proteins. Stem Cells Dev. *19*, 1485–1499.

Naba, A., Clauser, K.R., Ding, H., Whittaker, C.A., Carr, S.A., and Hynes, R.O. (2016). The extracellular matrix: Tools and insights for the "omics" era. Matrix Biol. *49*, 10–24.

Nagano, F., Kawabe, H., Nakanishi, H., Shinohara, M., Deguchi-Tawarada, M., Takeuchi, M., Sasaki, T., and Takai, Y. (2002). Rabconnectin-3, a novel protein that binds both GDP/GTP exchange protein and GTPase-activating protein for Rab3 small G protein family. J. Biol. Chem. *277*, 9629–9632.

Nelson, S.J.D., MS (2015). Wheeler's Dental Anatomy, Physiology and Occlusion, 10e (St. Louis, Missouri: Saunders).

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658.

Pääkkönen, V., Ohlmeier, S., Bergmann, U., Larmas, M., Salo, T., and Tjäderhane, L. (2005). Analysis of gene and protein expression in healthy and carious tooth pulp with cDNA microarray and two-dimensional gel electrophoresis. Eur. J. Oral Sci. *113*, 369–379.

Paik, Y.-K., Jeong, S.-K., Omenn, G.S., Uhlen, M., Hanash, S., Cho, S.Y., Lee, H.-J., Na, K., Choi, E.-Y., Yan, F., et al. (2012). The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. Nat. Biotechnol. *30*, 221–223.

Park, E.-S., Cho, H.-S., Kwon, T.-G., Jang, S.-N., Lee, S.-H., An, C.-H., Shin, H.-I., Kim, J.-Y., and Cho, J.-Y. (2009). Proteomics analysis of human dentin reveals distinct protein expression profiles. J. Proteome Res. *8*, 1338–1346.

Park, S., Shcheynikov, N., Hong, J.H., Zheng, C., Suh, S.H., Kawaai, K., Ando, H., Mizutani, A., Abe, T., Kiyonari, H., et al. (2013). Irbit Mediates Synergy Between Ca2+ and cAMP Signaling Pathways During Epithelial Transport in Mice. Gastroenterology *145*, 232–241.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Sharon, D., Tilgner, H., Grubert, F., and Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. Nat. Biotechnol. *31*, 1009–1014.

Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol. Cell. Proteomics MCP *10*, M111.007690.

Tjäderhane, L., Salo, T., Larjava, H., Larmas, M., and Overall, C.M. (1998). A novel organ culture method to study the function of human odontoblasts in vitro: gelatinase expression by odontoblasts is differentially regulated by TGF-beta1. J. Dent. Res. *77*, 1486–1496.

Wei, X., Wu, L., Ling, J., Liu, L., Liu, S., Liu, W., Li, M., and Xiao, Y. (2008). Differentially expressed protein profile of human dental pulp cells in the early process of odontoblast-like differentiation in vitro. J. Endod. *34*, 1077–1084.

Wessel, D., and Flügge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. *138*, 141–143.

Wu, S., Brown, J.N., Tolić, N., Meng, D., Liu, X., Zhang, H., Zhao, R., Moore, R.J., Pevzner, P., Smith, R.D., et al. (2014). Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. PROTEOMICS *14*, 1211–1222.

Zhao, Y., and Jensen, O.N. (2009). Modification-specific proteomics: Strategies for characterization of post-translational modifications using enrichment techniques. Proteomics *9*, 4632–4641.

Appendix A Peptide spectra matches

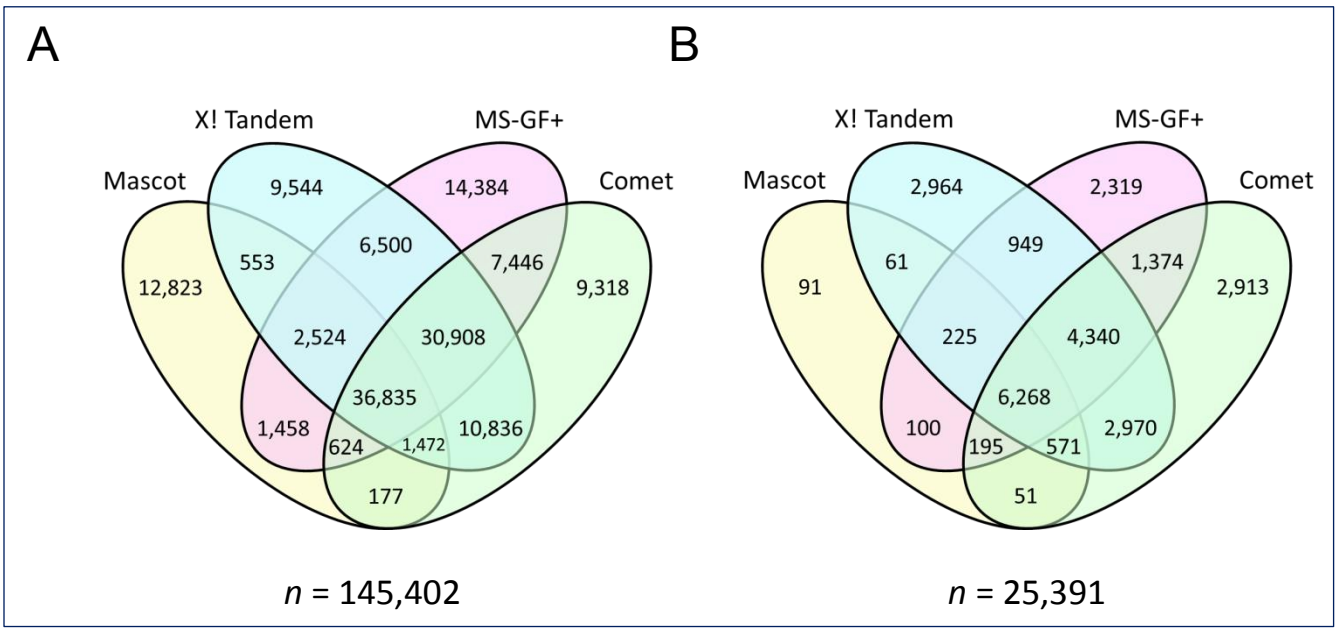
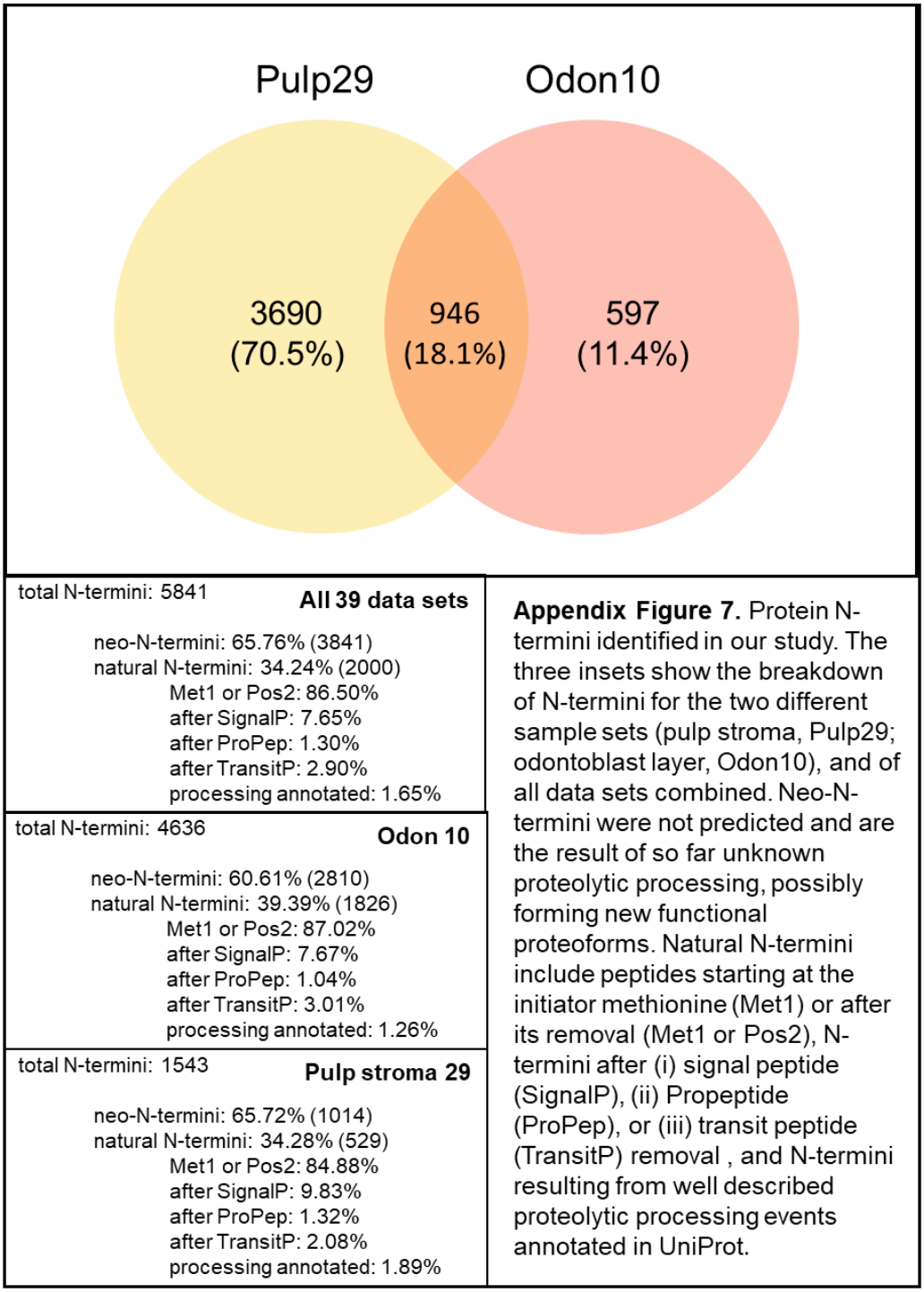


Figure 6. Peptide spectra matches found by the four search engines
(A) pulp stroma and (B) odontoblast cell layer samples. This data is the product of the raw LC-MS/MS spectra, translated by MSconvertGUI, and fed into search engines to identify peptide spectrum matches. Results were subsequently assessed and validated by PeptideProphet within the Trans-Proteomic Pipeline to ensure a ≤ 1 % false discovery rate (FDR).

Appendix B Natural and Neo-N-termini



Appendix Figure 7. Protein N-termini identified in our study. The three insets show the breakdown of N-termini for the two different sample sets (pulp stroma, Pulp29; odontoblast layer, Odon10), and of all data sets combined. Neo-N-termini were not predicted and are the result of so far unknown proteolytic processing, possibly forming new functional proteoforms. Natural N-termini include peptides starting at the initiator methionine (Met1) or after its removal (Met1 or Pos2), N-termini after (i) signal peptide (SignalP), (ii) Propeptide (ProPep), or (iii) transit peptide (TransitP) removal , and N-termini resulting from well described proteolytic processing events annotated in UniProt.

Figure 7. Numbers of Neo-N-Termini and Classification of Natural N-termini

Appendix C Odontoblast proteins

Table 2. Proteins found only in the odontoblast protein samples

211 proteins identified only in the odontoblast cell layer sample and not in stroma or dentine (from Jágr *et al* 2012). (Data also shown in Figure 4C).

Protein labeled Red is found only at transcript level PE2 according to neXtprot
Protein labeled **Red** bold was identified as a candidate missing protein, but did not satisy all criteria to be classified as a found missing protein

211	UniProt ID	Probability	Full protein name
1	Q9Y639	0.9933	Neuroplastin
2	Q9Y4D7	0.9962	Plexin-D1
3	Q9Y3D6	0.9894	Mitochondrial fission 1 protein OS=Homo sapiens GN=FIS1 PE=1 SV=2
4	Q9Y2Z4	0.9544	Tyrosine--tRNA ligase, mitochondrial OS=Homo sapiens GN=YARS2 PE=1 SV=2
5	Q9Y2J8	1.0000	Protein-arginine deiminase type-2 OS=Homo sapiens GN=PADI2 PE=1 SV=2
6	Q9Y2I6	0.9708	Ninein-like protein OS=Homo sapiens GN=NINL PE=1 SV=2
7	Q9Y2H0	0.9844	Disks large-associated protein 4
8	Q9Y2G9	0.9938	Protein strawberry notch homolog 2 OS=Homo sapiens GN=SBNO2 PE=2 SV=3
9	Q9Y2G4	0.9624	Ankyrin repeat domain-containing protein 6
10	Q9Y259	0.9923	Choline/ethanolamine kinase OS=Homo sapiens GN=CHKB PE=1 SV=3
11	Q9Y257	0.9937	Potassium channel subfamily K member 6 OS=Homo sapiens GN=KCNK6 PE=1 SV=1
12	Q9UQQ1	0.9901	N-acetylated-alpha-linked acidic dipeptidase-like protein
13	Q9UPW8	0.9524	Protein unc-13 homolog A OS=Homo sapiens GN=UNC13A PE=2 SV=4
14	Q9UPN3	0.9999	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 OS=Homo sapiens GN=MACF1 PE=1 SV=4
15	Q9UP83	0.9588	Conserved oligomeric Golgi complex subunit 5
16	Q9UNH5	0.9607	Dual specificity protein phosphatase CDC14A
17	Q9UNF0	0.9938	Protein kinase C and casein kinase substrate in neurons protein 2
18	Q9ULC6	0.9843	Protein-arginine deiminase type-1 OS=Homo sapiens GN=PADI1 PE=1 SV=2
19	Q9UKY1	0.9521	Zinc fingers and homeoboxes protein 1 OS=Homo sapiens GN=ZHX1 PE=1 SV=1
20	Q9UJC3	0.9739	Protein Hook homolog 1 OS=Homo sapiens GN=HOOK1 PE=1 SV=2
21	Q9UHX3	0.9682	Adhesion G protein-coupled receptor E2
22	Q9UBB9	0.9659	Tuftelin-interacting protein 11 OS=Homo sapiens GN=TFIP11 PE=1 SV=1
23	Q9P2J2	0.9757	Protein turtle homolog A
24	Q9P2E5	0.9763	Chondroitin sulfate glucuronyltransferase
25	Q9P1W3	0.9970	Transmembrane protein 63C OS=Homo sapiens GN=TMEM63C PE=2 SV=1
26	Q9P0X4	0.9791	Voltage-dependent T-type calcium channel subunit alpha-1I
27	Q9P0J1	0.9975	[Pyruvate dehydrogenase [acetyl-transferring]]-phosphatase 1, mitochondrial OS=Homo sapiens GN=PDP1 PE=1 SV=3
28	Q9NZ53	0.9924	Podocalyxin-like protein 2
29	Q9NY15	0.9823	Stabilin-1 OS=Homo sapiens GN=STAB1 PE=1 SV=3
30	Q9NVN3	0.9772	Synembryn-B
31	Q9NUU7	0.9713	ATP-dependent RNA helicase DDX19A
32	Q9NRD8	0.9651	Dual oxidase 2 OS=Homo sapiens GN=DUOX2 PE=1 SV=2
33	Q9HCL2	0.9875	Glycerol-3-phosphate acyltransferase 1, mitochondrial OS=Homo sapiens GN=GPAM PE=1 SV=3
34	Q9HA92	0.9685	Radical S-adenosyl methionine domain-containing protein 1, mitochondrial OS=Homo sapiens GN=RSAD1 PE=2 SV=2
35	Q9H9Y4	0.9967	GPN-loop GTPase 2 OS=Homo sapiens GN=GPN2 PE=2 SV=2
36	Q9H9V9	0.9538	JmjC domain-containing protein 4
37	Q9H9D4	0.9945	Zinc finger protein 408 OS=Homo sapiens GN=ZNF408 PE=1 SV=1
38	Q9H3R0	0.9686	Lysine-specific demethylase 4C
39	Q9H2G9	1.0000	Golgin-45 OS=Homo sapiens GN=BLZF1 PE=1 SV=2
40	Q9H2D6	0.9925	TRIO and F-actin-binding protein
41	Q9H299	0.9688	SH3 domain-binding glutamic acid-rich-like protein 3 OS=Homo sapiens GN=SH3BGRL3 PE=1 SV=1
42	Q9C0K0	0.9972	B-cell lymphoma/leukemia 11B
43	Q9C040	1.0000	Tripartite motif-containing protein 2
44	Q9BXP2	0.9682	Solute carrier family 12 member 9 OS=Homo sapiens GN=SLC12A9 PE=1 SV=1
45	Q96Q05	0.9696	Trafficking protein particle complex subunit 9
46	Q96PF1	0.9728	Protein-glutamine gamma-glutamyltransferase Z OS=Homo sapiens GN=TGM7 PE=2 SV=1
47	Q96NL6	0.9964	Sodium channel and clathrin linker 1 OS=Homo sapiens GN=SCLT1 PE=2 SV=2
48	Q96JH8	0.9997	Ras-associating and dilute domain-containing protein
49	Q96IT1	0.9633	Zinc finger protein 496
50	Q96FT7	0.9591	Acid-sensing ion channel 4
51	Q96EK7	0.9598	Constitutive coactivator of peroxisome proliferator-activated receptor gamma OS=Homo sapiens GN=FAM120B PE=1 SV=1
52	Q96EH3	0.9575	Mitochondrial assembly of ribosomal large subunit protein 1 OS=Homo sapiens GN=MALSU1 PE=1 SV=1
53	Q96DZ1	0.9977	Endoplasmic reticulum lectin 1
54	Q96CN9	0.9965	GRIP and coiled-coil domain-containing protein 1 OS=Homo sapiens GN=GCC1 PE=1 SV=1

	UniProt ID	Probability	Full protein name
55	Q96A99	0.9964	Pentraxin-4
56	Q969T7	0.9809	7-methylguanosine phosphate-specific 5'-nucleotidase
57	Q969S9	0.9903	Ribosome-releasing factor 2, mitochondrial
58	Q92994	0.9995	Transcription factor IIIB 90 kDa subunit
59	Q92985	1.0000	Interferon regulatory factor 7
60	Q92959	0.9631	Solute carrier organic anion transporter family member 2A1 OS=Homo sapiens GN=SLCO2A1 PE=1 SV=2
61	Q92805	0.9675	Golgin subfamily A member 1 OS=Homo sapiens GN=GOLGA1 PE=1 SV=3
62	Q92729	0.9766	Receptor-type tyrosine-protein phosphatase U
63	Q92574	0.9744	Hamartin
64	Q8WXH0	0.9851	Nesprin-2
65	Q8WWY3	0.9535	U4/U6 small nuclear ribonucleoprotein Prp31 OS=Homo sapiens GN=PRPF31 PE=1 SV=2
66	Q8TEU7	0.9555	Rap guanine nucleotide exchange factor 6
67	Q8TEA8	0.9933	D-tyrosyl-tRNA(Tyr) deacylase 1 OS=Homo sapiens GN=DTD1 PE=1 SV=2
68	Q8TE73	0.9517	Dynein heavy chain 5, axonemal OS=Homo sapiens GN=DNAH5 PE=1 SV=3
69	Q8TDV2	0.9599	Probable G-protein coupled receptor 148 OS=Homo sapiens GN=GPR148 PE=2 SV=2
70	**Q8TDJ6**	0.9775	DmX-like protein 2
71	Q8TD26	0.9517	Chromodomain-helicase-DNA-binding protein 6 OS=Homo sapiens GN=CHD6 PE=1 SV=4
72	Q8TC90	0.9817	Coiled-coil domain-containing glutamate-rich protein 1 OS=Homo sapiens GN=CCER1 PE=2 SV=1
73	Q8TAX7	0.9999	Mucin-7 OS=Homo sapiens GN=MUC7 PE=1 SV=2
74	Q8NEN9	0.9722	PDZ domain-containing protein 8 OS=Homo sapiens GN=PDZD8 PE=1 SV=1
75	Q8NE09	0.9753	Regulator of G-protein signaling 22
76	Q8ND04	0.9727	Protein SMG8
77	Q8NCN4	0.9938	E3 ubiquitin-protein ligase RNF169 OS=Homo sapiens GN=RNF169 PE=1 SV=2
78	Q8NBJ5	0.9832	Procollagen galactosyltransferase 1 OS=Homo sapiens GN=COLGALT1 PE=1 SV=1
79	Q8N3K9	0.9993	Cardiomyopathy-associated protein 5 OS=Homo sapiens GN=CMYA5 PE=1 SV=3
80	Q8N3C0	0.9740	Activating signal cointegrator 1 complex subunit 3 OS=Homo sapiens GN=ASCC3 PE=1 SV=3
81	Q8N2S1	0.9996	Latent-transforming growth factor beta-binding protein 4 OS=Homo sapiens GN=LTBP4 PE=1 SV=2
82	Q8IZD2	0.9640	Histone-lysine N-methyltransferase 2E
83	Q8IWV8	0.9944	E3 ubiquitin-protein ligase UBR2
84	Q8IVS8	0.9816	Glycerate kinase
85	Q86W26	0.9764	NACHT, LRR and PYD domains-containing protein 10 OS=Homo sapiens GN=NLRP10 PE=1 SV=1
86	Q86VL8	0.9538	Multidrug and toxin extrusion protein 2
87	Q86UT6	0.9975	NLR family member X1
88	Q86UP2	0.9933	Kinectin
89	Q86SQ4	1.0000	G-protein coupled receptor 126
90	Q7Z7L9	0.9881	Zinc finger and SCAN domain-containing protein 2
91	Q7Z6K5	0.9561	Isoform C15orf38-AP3S2 of UPF0552 protein C15orf38 OS=Homo sapiens GN=C15orf38
92	Q7Z6I6	0.9894	Rho GTPase-activating protein 30
93	Q7Z402	0.9927	Transmembrane channel-like protein 7 OS=Homo sapiens GN=TMC7 PE=2 SV=1
94	Q76LX8	0.9581	A disintegrin and metalloproteinase with thrombospondin motifs 13 OS=Homo sapiens GN=ADAMTS13 PE=1 SV=1
95	Q76I76	0.9970	Protein phosphatase Slingshot homolog 2 OS=Homo sapiens GN=SSH2 PE=1 SV=1
96	Q6ZSJ9	0.9986	Protein shisa-6 homolog
97	Q6ZN55	0.9933	Isoform 2 of Zinc finger protein 574 OS=Homo sapiens GN=ZNF574
98	Q6UWH4	0.9620	Protein FAM198B
99	Q6Q0C0	0.9516	E3 ubiquitin-protein ligase TRAF7 OS=Homo sapiens GN=TRAF7 PE=1 SV=1
100	Q6P1N0	0.9806	Coiled-coil and C2 domain-containing protein 1A
101	Q6NY19	0.9871	KN motif and ankyrin repeat domain-containing protein 3
102	**Q6NSI1**	0.9907	Putative ankyrin repeat domain-containing protein 26-like protein OS=Homo sapiens GN=ANKRD26P1 PE=5 SV=2
103	Q6EKJ0	0.9953	General transcription factor II-I repeat domain-containing protein 2B
104	Q63HM1	0.9994	Kynurenine formamidase
105	Q5U651	1.0000	Ras-interacting protein 1 OS=Homo sapiens GN=RASIP1 PE=1 SV=1
106	Q5T7M4	0.9969	Adipolin (Protein FAM132A OS=Homo sapiens GN=FAM132A PE=2 SV=2)
107	Q5T4B2	0.9995	Probable inactive glycosyltransferase 25 family member 3 OS=Homo sapiens GN=CERCAM PE=2 SV=1
108	Q5RHP9	0.9768	Glutamate-rich protein 3
109	Q5H9L4	0.9888	Transcription initiation factor TFIID subunit 7-like
110	Q58EX7	0.9995	Puratrophin-1
111	Q53GL7	0.9936	Poly [ADP-ribose] polymerase 10 OS=Homo sapiens GN=PARP10 PE=1 SV=2
112	Q53EQ6	0.9814	Tigger transposable element-derived protein 5
113	Q4LDE5	0.9883	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
114	Q496Y0	0.9550	LON peptidase N-terminal domain and RING finger protein 3
115	Q16760	0.9969	Diacylglycerol kinase delta
116	Q16538	**0.9619**	Probable G-protein coupled receptor 162
117	Q16363	0.9720	Laminin subunit alpha-4

	UniProt ID	Probability	Full protein name
118	Q15819	0.9933	Ubiquitin-conjugating enzyme E2 variant 2 OS=Homo sapiens GN=UBE2V2 PE=1 SV=4
119	Q15435	0.9758	Protein phosphatase 1 regulatory subunit 7
120	Q14790	0.9992	Caspase-8
121	Q14643	0.9551	Inositol 1,4,5-trisphosphate receptor type 1
122	Q14194	1.0000	Dihydropyrimidinase-related protein 1 OS=Homo sapiens GN=CRMP1 PE=1 SV=1
123	Q13618	0.9594	Cullin-3
124	Q13585	0.9570	Melatonin-related receptor OS=Homo sapiens GN=GPR50 PE=1 SV=3
125	Q13242	0.9816	Serine/arginine-rich splicing factor 9 OS=Homo sapiens GN=SRSF9 PE=1 SV=1
126	Q13085	0.9912	Acetyl-CoA carboxylase 1
127	Q08174	0.9676	Protocadherin-1
128	Q08043	0.9787	Alpha-actinin-3 OS=Homo sapiens GN=ACTN3 PE=1 SV=2
129	**Q02846**	0.9780	Retinal guanylyl cyclase 1 OS=Homo sapiens GN=GUCY2D PE=1 SV=2
130	P80511	0.9907	Protein S100-A12 OS=Homo sapiens GN=S100A12 PE=1 SV=2
131	P78562	1.0000	Phosphate-regulating neutral endopeptidase OS=Homo sapiens GN=PHEX PE=1 SV=1
132	P78415	0.9978	Iroquois-class homeodomain protein IRX-3 OS=Homo sapiens GN=IRX3 PE=2 SV=3
133	P69891	1.0000	Hemoglobin subunit gamma-1
134	P60763	0.9990	Ras-related C3 botulinum toxin substrate 3 OS=Homo sapiens GN=RAC3 PE=1 SV=1
135	P57721	0.9933	Poly(rC)-binding protein 3
136	P55283	0.9988	Cadherin-4
137	P53609	0.9869	Geranylgeranyl transferase type-1 subunit beta
138	P53420	0.9765	Collagen alpha-4(IV) chain OS=Homo sapiens GN=COL4A4 PE=1 SV=3
139	P49736	0.9722	DNA replication licensing factor MCM2 OS=Homo sapiens GN=MCM2 PE=1 SV=4
140	P49641	0.9944	Alpha-mannosidase 2x
141	P43355	0.9752	Melanoma-associated antigen 1 OS=Homo sapiens GN=MAGEA1 PE=1 SV=1
142	P43034	0.9540	Platelet-activating factor acetylhydrolase IB subunit alpha OS=Homo sapiens GN=PAFAH1B1 PE=1 SV=2
143	P42345	0.9988	Serine/threonine-protein kinase mTOR OS=Homo sapiens GN=MTOR PE=1 SV=1
144	P35125	0.9864	Ubiquitin carboxyl-terminal hydrolase 6
145	P33241	0.9822	Lymphocyte-specific protein 1 OS=Homo sapiens GN=LSP1 PE=1 SV=1
146	P28300	0.9933	Protein-lysine 6-oxidase OS=Homo sapiens GN=LOX PE=1 SV=2
147	P26447	0.9970	Protein S100-A4 OS=Homo sapiens GN=S100A4 PE=1 SV=1
148	P24158	0.9829	Myeloblastin OS=Homo sapiens GN=PRTN3 PE=1 SV=3
149	P22888	0.9991	Lutropin-choriogonadotropic hormone receptor
150	P21912	0.9953	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial OS=Homo sapiens GN=SDHB PE=1 SV=3
151	P21802	0.9568	Isoform 10 of Fibroblast growth factor receptor 2 OS=Homo sapiens GN=FGFR2
152	P20160	1.0000	Azurocidin OS=Homo sapiens GN=AZU1 PE=1 SV=3
153	P17858	1.0000	6-phosphofructokinase, liver type OS=Homo sapiens GN=PFKL PE=1 SV=6
154	P17844	1.0000	Probable ATP-dependent RNA helicase DDX5 OS=Homo sapiens GN=DDX5 PE=1 SV=1
155	P16157	0.9663	Ankyrin-1
156	P13796	0.9967	Plastin-2 OS=Homo sapiens GN=LCP1 PE=1 SV=6
157	P13727	0.9991	Bone marrow proteoglycan OS=Homo sapiens GN=PRG2 PE=1 SV=2
158	P13497	0.9860	Isoform BMP1-5 of Bone morphogenetic protein 1 OS=Homo sapiens GN=BMP1
159	P0CW18	0.9850	Serine protease 56 OS=Homo sapiens GN=PRSS56 PE=1 SV=1
160	P0CB43	0.9644	Protein HGH1 homolog
161	**P0C7T7**	0.9679	Putative uncharacterized protein FRMD6-AS1 OS=Homo sapiens GN=FRMD6-AS1 PE=5 SV=1
162	P0C0S5	1.0000	Histone H2A.Z
163	P09972	1.0000	Fructose-bisphosphate aldolase C OS=Homo sapiens GN=ALDOC PE=1 SV=2
164	P09917	0.9991	Arachidonate 5-lipoxygenase OS=Homo sapiens GN=ALOX5 PE=1 SV=2
165	P09488	0.9920	Glutathione S-transferase Mu 1
166	P09467	0.9933	Fructose-1,6-bisphosphatase 1 OS=Homo sapiens GN=FBP1 PE=1 SV=5
167	P09131	0.9596	P3 protein
168	P08887	0.9767	Interleukin-6 receptor subunit alpha
169	P08514	0.9942	Integrin alpha-IIb
170	P08311	1.0000	Cathepsin G OS=Homo sapiens GN=CTSG PE=1 SV=2
171	P08246	1.0000	Neutrophil elastase OS=Homo sapiens GN=ELANE PE=1 SV=1
172	P05164	1.0000	Myeloperoxidase
173	P05141	0.9976	ADP/ATP translocase 2 OS=Homo sapiens GN=SLC25A5 PE=1 SV=7
174	P05129	0.9986	Protein kinase C gamma type OS=Homo sapiens GN=PRKCG PE=1 SV=3
175	P04628	0.9701	Proto-oncogene Wnt-1 OS=Homo sapiens GN=WNT1 PE=1 SV=1
176	P02814	0.9933	Submaxillary gland androgen-regulated protein 3B OS=Homo sapiens GN=SMR3B PE=1 SV=2
177	P02810	0.9933	Salivary acidic proline-rich phosphoprotein 1/2 OS=Homo sapiens GN=PRH1 PE=1 SV=2
178	P02788	1.0000	Lactotransferrin
179	P02100	1.0000	Hemoglobin subunit epsilon OS=Homo sapiens GN=HBE1 PE=1 SV=2
180	P01877	0.9991	Ig alpha-2 chain C region OS=Homo sapiens GN=IGHA2 PE=1 SV=3
181	P01761	0.9927	Ig heavy chain V-I region SIE OS=Homo sapiens PE=1 SV=1

	UniProt ID	Probability	Full protein name
182	P01591	0.9933	Immunoglobulin J chain OS=Homo sapiens GN=IGJ PE=1 SV=4
183	P00966	0.9999	Argininosuccinate synthase OS=Homo sapiens GN=ASS1 PE=1 SV=2
184	P00748	1.0000	Coagulation factor XII OS=Homo sapiens GN=F12 PE=1 SV=3
185	O95816	0.9914	BAG family molecular chaperone regulator 2 OS=Homo sapiens GN=BAG2 PE=1 SV=1
186	**O94844**	0.9999	Rho-related BTB domain-containing protein 1 OS=Homo sapiens GN=RHOBTB1 PE=1 SV=2
187	O75821	0.9933	Eukaryotic translation initiation factor 3 subunit G OS=Homo sapiens GN=EIF3G PE=1 SV=2
188	O75718	0.9916	Cartilage-associated protein OS=Homo sapiens GN=CRTAP PE=1 SV=1
189	O75427	0.9965	Leucine-rich repeat and calponin homology domain-containing protein 4 OS=Homo sapiens GN=LRCH4 PE=1 SV=2
190	O75366	0.9533	Advillin
191	O75145	1.0000	Liprin-alpha-3
192	O75127	0.9883	Pentatricopeptide repeat-containing protein 1, mitochondrial OS=Homo sapiens GN=PTCD1 PE=1 SV=2
193	O60814	1.0000	Histone H2B type 1-K
194	O60271	1.0000	C-Jun-amino-terminal kinase-interacting protein 4
195	O60240	0.9972	Perilipin-1 OS=Homo sapiens GN=PLIN1 PE=1 SV=2
196	O43707	1.0000	Alpha-actinin-4
197	O43613	0.9624	Orexin receptor type 1 OS=Homo sapiens GN=HCRTR1 PE=2 SV=2
198	O43312	0.9944	Metastasis suppressor protein 1 OS=Homo sapiens GN=MTSS1 PE=1 SV=2
199	O15511	0.9920	Actin-related protein 2/3 complex subunit 5
200	O15439	0.9697	Multidrug resistance-associated protein 4
201	O15417	0.9999	Trinucleotide repeat-containing gene 18 protein OS=Homo sapiens GN=TNRC18 PE=1 SV=3
202	O14939	0.9994	Isoform PLD2B of Phospholipase D2 OS=Homo sapiens GN=PLD2
203	O14732	0.9865	Inositol monophosphatase 2
204	O00445	1.0000	Synaptotagmin-5 OS=Homo sapiens GN=SYT5 PE=2 SV=2
205	E7EW31	0.9805	Proline-rich basic protein 1 OS=Homo sapiens GN=PROB1 PE=2 SV=2
206	A8MW92	0.9996	PHD finger protein 20-like protein 1 OS=Homo sapiens GN=PHF20L1 PE=1 SV=2
207	**A8MTW9**	0.9916	Putative uncharacterized protein ENSP00000380674 OS=Homo sapiens PE=5 SV=2
208	A7MCY6	0.9688	TANK-binding kinase 1-binding protein 1 OS=Homo sapiens GN=TBKBP1 PE=1 SV=1
209	A7E2V4	0.9665	Zinc finger SWIM domain-containing protein 8
210	A6H8Y1	0.9726	Transcription factor TFIIIB component B'' homolog
211	A1KZ92	0.9947	Peroxidasin-like protein

Table 2: These proteins are likely to be highly expressed or only expressed by odontoblasts. Proteins are listed with their Uniprot ID, protein probability from TPP analysis of PSM data, and full protein name.

Appendix D Age differences in odontoblast proteins

Table 3. Young odontoblast proteins

Proteins found in the odontoblast cell layer samples from donors < 20 (*N* = 1, *n* = 2). (Data also shown in Figure 4D).			
331	**UniProt ID**	**Probability**	**Protein Description**
1	Q9Y6U3	1.0000	Adseverin OS=Homo sapiens GN=SCIN PE=1 SV=4
2	Q9Y6C2	1.0000	EMILIN-1 OS=Homo sapiens GN=EMILIN1 PE=1 SV=2
3	Q9Y639-1	0.9918	Neuroplastin, isoform 1
4	Q9Y5I2-3	0.9534	Protocadherin alpha-10, isoform 3
5	Q9Y4F1-2	0.9611	FERM, RhoGEF and pleckstrin domain-containing protein 1, isoform 2
6	Q9Y3B8	0.9918	Oligoribonuclease, mitochondrial OS=Homo sapiens GN=REXO2 PE=1 SV=3
7	Q9Y3A5	0.9910	Ribosome maturation protein SBDS OS=Homo sapiens GN=SBDS PE=1 SV=4
8	Q9Y2Z0-2	0.9918	Protein SGT1 homolog, isoform 2
9	Q9Y2J8	1.0000	Protein-arginine deiminase type-2 OS=Homo sapiens GN=PADI2 PE=1 SV=2
10	Q9Y2J2-2	0.9918	Band 4.1-like protein 3, isoform 2
11	Q9Y281	0.9918	Cofilin-2 OS=Homo sapiens GN=CFL2 PE=1 SV=1
12	Q9Y230	0.9918	RuvB-like 2 OS=Homo sapiens GN=RUVBL2 PE=1 SV=3
13	Q9UN36-3	0.9918	Protein NDRG2, isoform 3
14	Q9UKK9	0.9918	ADP-sugar pyrophosphatase OS=Homo sapiens GN=NUDT5 PE=1 SV=1
15	Q9UK22	0.9886	F-box only protein 2 OS=Homo sapiens GN=FBXO2 PE=1 SV=2
16	Q9UJW0-3	0.9846	Dynactin subunit 4, isoform 3
17	Q9UGM5-2	0.9918	Fetuin-B, isoform 2
18	Q9UBR2	0.9512	Cathepsin Z OS=Homo sapiens GN=CTSZ PE=1 SV=1
19	Q9UBQ7	1.0000	Glyoxylate reductase/hydroxypyruvate reductase OS=Homo sapiens GN=GRHPR PE=1 SV=1
20	Q9UBQ5	0.9910	Eukaryotic translation initiation factor 3 subunit K OS=Homo sapiens GN=EIF3K PE=1 SV=1
21	Q9P1W3	0.9991	Transmembrane protein 63C OS=Homo sapiens GN=TMEM63C PE=2 SV=1
22	Q9NZW4	1.0000	Dentin sialophosphoprotein OS=Homo sapiens GN=DSPP PE=1 SV=2
23	Q9NZJ9	0.9918	Diphosphoinositol polyphosphate phosphohydrolase 2 OS=Homo sapiens GN=NUDT4 PE=1 SV=2
24	Q9NZ53-2	0.9592	Podocalyxin-like protein 2, isoform 2
25	Q9NVS9	0.9735	Pyridoxine-5'-phosphate oxidase OS=Homo sapiens GN=PNPO PE=1 SV=1
26	Q9NVA2	1.0000	Septin-11 OS=Homo sapiens GN=SEPT11 PE=1 SV=3
27	Q9NUU7	0.9886	ATP-dependent RNA helicase DDX19A
28	Q9NUP9	0.9902	Protein lin-7 homolog C OS=Homo sapiens GN=LIN7C PE=1 SV=1
29	Q9NUL5-3	0.9918	Repressor of yield of DENV protein, isoform 3
30	Q9NR56-2	0.9894	Muscleblind-like protein 1, isoform 2
31	Q9NQC3-2	0.9918	Reticulon-4, isoform 2
32	Q9NPJ3	0.9918	Acyl-coenzyme A thioesterase 13 OS=Homo sapiens GN=ACOT13 PE=1 SV=1
33	Q9HB96	0.9915	Fanconi anemia group E protein OS=Homo sapiens GN=FANCE PE=1 SV=1
34	Q9H8L6	0.9886	Multimerin-2 OS=Homo sapiens GN=MMRN2 PE=1 SV=2
35	Q9H4A4	0.9918	Aminopeptidase B OS=Homo sapiens GN=RNPEP PE=1 SV=2
36	Q9H3P7	0.9782	Golgi resident protein GCP60 OS=Homo sapiens GN=ACBD3 PE=1 SV=4
37	Q9H0E2	1.0000	Toll-interacting protein OS=Homo sapiens GN=TOLLIP PE=1 SV=1
38	Q9C040-2	1.0000	Tripartite motif-containing protein 2, isoform 2
39	Q9BU02-2	0.9918	Thiamine-triphosphatase, isoform 2
40	Q9BS40	1.0000	Latexin OS=Homo sapiens GN=LXN PE=1 SV=2
41	Q9BRF8-2	0.9918	Serine/threonine-protein phosphatase CPPED1, isoform 2
42	Q99983	1.0000	Osteomodulin OS=Homo sapiens GN=OMD PE=1 SV=1
43	Q99832	0.9918	T-complex protein 1 subunit eta OS=Homo sapiens GN=CCT7 PE=1 SV=2
44	Q99715-4	0.9964	Collagen alpha-1(XII) chain, isoform 4
45	Q99623-2	0.9918	Prohibitin-2, isoform 2
46	Q99584	0.9918	Protein S100-A13 OS=Homo sapiens GN=S100A13 PE=1 SV=1
47	Q99439-2	0.9918	Calponin-2, isoform 2
48	Q99436	0.9918	Proteasome subunit beta type-7 OS=Homo sapiens GN=PSMB7 PE=1 SV=1
49	Q96MC2	0.9782	Dynein regulatory complex protein 1 OS=Homo sapiens GN=DRC1 PE=2 SV=2
50	Q96KP4	1.0000	Cytosolic non-specific dipeptidase OS=Homo sapiens GN=CNDP2 PE=1 SV=2

	UniProt ID	Probability	Protein Description
51	Q96HC4-2	0.9918	PDZ and LIM domain protein 5, isoform 2
52	Q96G03	0.9918	Phosphoglucomutase-2 OS=Homo sapiens GN=PGM2 PE=1 SV=4
53	Q96F85-2	0.9806	CB1 cannabinoid receptor-interacting protein 1, isoform 2
54	Q96EM0	0.9918	Trans-L-3-hydroxyproline dehydratase OS=Homo sapiens GN=L3HYPDH PE=1 SV=2
55	Q96CN9	0.9918	GRIP and coiled-coil domain-containing protein 1 OS=Homo sapiens GN=GCC1 PE=1 SV=1
56	Q96C86	0.9918	m7GpppX diphosphatase OS=Homo sapiens GN=DCPS PE=1 SV=2
57	Q96C19	0.9727	EF-hand domain-containing protein D2 OS=Homo sapiens GN=EFHD2 PE=1 SV=1
58	Q969T7-2	0.9846	7-methylguanosine phosphate-specific 5'-nucleotidase, isoform 2
59	Q92747	0.9918	Actin-related protein 2/3 complex subunit 1A OS=Homo sapiens GN=ARPC1A PE=1 SV=2
60	Q92729	0.9960	Receptor-type tyrosine-protein phosphatase U OS=Homo sapiens GN=PTPRU PE=1 SV=2
61	Q8WVM8	0.9999	Sec1 family domain-containing protein 1 OS=Homo sapiens GN=SCFD1 PE=1 SV=4
62	Q8TEA8	0.9918	D-tyrosyl-tRNA(Tyr) deacylase 1 OS=Homo sapiens GN=DTD1 PE=1 SV=2
63	Q8TC59-2	0.9683	Piwi-like protein 2, isoform 2
64	Q8TAX7	0.9997	Mucin-7 OS=Homo sapiens GN=MUC7 PE=1 SV=2
65	Q8NE09-2	0.9790	Regulator of G-protein signaling 22, isoform 2
66	Q8NBS9-2	0.9918	Thioredoxin domain-containing protein 5, isoform 2
67	Q8NA70	0.9657	Protein FAM47B OS=Homo sapiens GN=FAM47B PE=2 SV=2
68	Q8N4P3-2	0.9918	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase MESH1, isoform 2
69	Q8IW45-2	0.9918	ATP-dependent (S)-NAD(P)H-hydrate dehydratase, isoform 2
70	Q86UP2-2	0.9918	Kinectin, isoform 2
71	Q6UWY5	1.0000	Olfactomedin-like protein 1 OS=Homo sapiens GN=OLFML1 PE=1 SV=2
72	Q6QNY1	0.9918	Biogenesis of lysosome-related organelles complex 1 subunit 2 OS=Homo sapiens GN=BLOC1S2 PE=1 SV=1
73	Q6NZI2	1.0000	Polymerase I and transcript release factor OS=Homo sapiens GN=PTRF PE=1 SV=1
74	Q5VWZ2-2	0.9918	Lysophospholipase-like protein 1, isoform 2
75	Q5SSJ5-2	1.0000	Heterochromatin protein 1-binding protein 3, isoform 2
76	Q496Y0-2	0.9822	LON peptidase N-terminal domain and RING finger protein 3, isoform 2
77	Q2TB90-2	0.9778	Putative hexokinase HKDC1, isoform 2
78	Q2TAA2	0.9918	Isoamyl acetate-hydrolyzing esterase 1 homolog OS=Homo sapiens GN=IAH1 PE=1 SV=1
79	Q16658	1.0000	Fascin OS=Homo sapiens GN=FSCN1 PE=1 SV=3
80	Q16555	1.0000	Dihydropyrimidinase-related protein 2 OS=Homo sapiens GN=DPYSL2 PE=1 SV=1
81	Q16401-2	0.9910	26S proteasome non-ATPase regulatory subunit 5, isoform 2
82	Q16186	0.9918	Proteasomal ubiquitin receptor ADRM1 OS=Homo sapiens GN=ADRM1 PE=1 SV=2
83	Q15819	0.9918	Ubiquitin-conjugating enzyme E2 variant 2 OS=Homo sapiens GN=UBE2V2 PE=1 SV=4
84	Q15436	0.9918	Protein transport protein Sec23A OS=Homo sapiens GN=SEC23A PE=1 SV=2
85	Q15435-2	0.9774	Protein phosphatase 1 regulatory subunit 7, isoform 2
86	Q15427	0.9918	Splicing factor 3B subunit 4 OS=Homo sapiens GN=SF3B4 PE=1 SV=1
87	Q15185	1.0000	Prostaglandin E synthase 3 OS=Homo sapiens GN=PTGES3 PE=1 SV=1
88	Q15181	0.9918	Inorganic pyrophosphatase OS=Homo sapiens GN=PPA1 PE=1 SV=2
89	Q15063-2	1.0000	Periostin, isoform 2
90	Q15005	0.9918	Signal peptidase complex subunit 2 OS=Homo sapiens GN=SPCS2 PE=1 SV=3
91	Q14697-2	0.9918	Neutral alpha-glucosidase AB, isoform 2
92	Q14517	0.9934	Protocadherin Fat 1 OS=Homo sapiens GN=FAT1 PE=1 SV=2
93	Q14204	0.9918	Cytoplasmic dynein 1 heavy chain 1 OS=Homo sapiens GN=DYNC1H1 PE=1 SV=5
94	Q14203-2	0.9918	Dynactin subunit 1, isoform p135
95	Q14103-2	1.0000	Heterogeneous nuclear ribonucleoprotein D0, isoform 2
96	Q14019	1.0000	Coactosin-like protein OS=Homo sapiens GN=COTL1 PE=1 SV=3
97	Q13642-1	0.9836	Four and a half LIM domains protein 1, isoform 1
98	Q13618-3	0.9806	Cullin-3, isoform 3
99	Q13445	0.9918	Transmembrane emp24 domain-containing protein 1 OS=Homo sapiens GN=TMED1 PE=1 SV=1
100	Q13418	0.9918	Integrin-linked protein kinase OS=Homo sapiens GN=ILK PE=1 SV=2
101	Q13409-2	0.9894	Cytoplasmic dynein 1 intermediate chain 2, isoform 2B
102	Q13404-8	0.9918	Ubiquitin-conjugating enzyme E2 variant 1, isoform 6
103	Q13283	0.9990	Ras GTPase-activating protein-binding protein 1 OS=Homo sapiens GN=G3BP1 PE=1 SV=1
104	Q13242	0.9634	Serine/arginine-rich splicing factor 9 OS=Homo sapiens GN=SRSF9 PE=1 SV=1
105	Q12874	0.9918	Splicing factor 3A subunit 3 OS=Homo sapiens GN=SF3A3 PE=1 SV=1
106	Q09028-3	0.9918	Histone-binding protein RBBP4, isoform 3
107	Q08257	1.0000	Quinone oxidoreductase OS=Homo sapiens GN=CRYZ PE=1 SV=1
108	Q07157-2	0.9918	Tight junction protein ZO-1, isoform short

	UniProt ID	Probability	Protein Description
109	Q04837	0.9918	Single-stranded DNA-binding protein, mitochondrial OS=Homo sapiens GN=SSBP1 PE=1 SV=1
110	Q04760-2	0.9918	Lactoylglutathione lyase, isoform 2
111	Q03154-2	0.9902	Aminoacylase-1, isoform 2
112	Q02846	0.9657	Retinal guanylyl cyclase 1 OS=Homo sapiens GN=GUCY2D PE=1 SV=2
113	Q02543	0.9918	60S ribosomal protein L18a OS=Homo sapiens GN=RPL18A PE=1 SV=2
114	Q01995	0.9894	Transgelin OS=Homo sapiens GN=TAGLN PE=1 SV=4
115	Q00688	0.9886	Peptidyl-prolyl cis-trans isomerase FKBP3 OS=Homo sapiens GN=FKBP3 PE=1 SV=1
116	Q00577	1.0000	Transcriptional activator protein Pur-alpha OS=Homo sapiens GN=PURA PE=1 SV=2
117	P78562	1.0000	Phosphate-regulating neutral endopeptidase OS=Homo sapiens GN=PHEX PE=1 SV=1
118	P78417-3	0.9918	Glutathione S-transferase omega-1, isoform 3
119	P68363	1.0000	Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1
120	P63244	0.9918	Guanine nucleotide-binding protein subunit beta-2-like 1 OS=Homo sapiens GN=GNB2L1 PE=1 SV=3
121	P63220	0.9918	40S ribosomal protein S21 OS=Homo sapiens GN=RPS21 PE=1 SV=1
122	P62913	0.9918	60S ribosomal protein L11 OS=Homo sapiens GN=RPL11 PE=1 SV=2
123	P62879	1.0000	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2 OS=Homo sapiens GN=GNB2 PE=1 SV=3
124	P62873	1.0000	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 OS=Homo sapiens GN=GNB1 PE=1 SV=3
125	P62826	1.0000	GTP-binding nuclear protein Ran OS=Homo sapiens GN=RAN PE=1 SV=3
126	P62280	0.9894	40S ribosomal protein S11 OS=Homo sapiens GN=RPS11 PE=1 SV=3
127	P62195	1.0000	26S protease regulatory subunit 8 OS=Homo sapiens GN=PSMC5 PE=1 SV=1
128	P61163	0.9992	Alpha-centractin OS=Homo sapiens GN=ACTR1A PE=1 SV=1
129	P61160-2	1.0000	Actin-related protein 2, isoform 2
130	P61086-2	0.9910	Ubiquitin-conjugating enzyme E2 K, isoform 2
131	P60520	0.9790	Gamma-aminobutyric acid receptor-associated protein-like 2 OS=Homo sapiens GN=GABARAPL2 PE=1 SV=1
132	P59998	1.0000	Actin-related protein 2/3 complex subunit 4 OS=Homo sapiens GN=ARPC4 PE=1 SV=3
133	P55795	1.0000	Heterogeneous nuclear ribonucleoprotein H2 OS=Homo sapiens GN=HNRNPH2 PE=1 SV=1
134	P55283-2	0.9918	Cadherin-4, isoform 2
135	P55011-3	0.9918	Solute carrier family 12 member 2, isoform 2
136	P54920	0.9918	Alpha-soluble NSF attachment protein OS=Homo sapiens GN=NAPA PE=1 SV=3
137	P54819-2	0.9918	Adenylate kinase 2, mitochondrial, isoform 2
138	P54132	0.9918	Bloom syndrome protein OS=Homo sapiens GN=BLM PE=1 SV=1
139	P54098	0.9816	DNA polymerase subunit gamma-1 OS=Homo sapiens GN=POLG PE=1 SV=1
140	P52907	0.9989	F-actin-capping protein subunit alpha-1 OS=Homo sapiens GN=CAPZA1 PE=1 SV=3
141	P52272-2	0.9918	Heterogeneous nuclear ribonucleoprotein M, isoform 2
142	P51991	1.0000	Heterogeneous nuclear ribonucleoprotein A3 OS=Homo sapiens GN=HNRNPA3 PE=1 SV=2
143	P50991-2	0.9894	T-complex protein 1 subunit delta, isoform 2
144	P50990	0.9918	T-complex protein 1 subunit theta OS=Homo sapiens GN=CCT8 PE=1 SV=4
145	P50897	1.0000	Palmitoyl-protein thioesterase 1 OS=Homo sapiens GN=PPT1 PE=1 SV=1
146	P50453	1.0000	Serpin B9 OS=Homo sapiens GN=SERPINB9 PE=1 SV=1
147	P50395-2	1.0000	Rab GDP dissociation inhibitor beta, isoform 2
148	P49914	0.9918	5-formyltetrahydrofolate cyclo-ligase OS=Homo sapiens GN=MTHFS PE=1 SV=2
149	P49773	1.0000	Histidine triad nucleotide-binding protein 1 OS=Homo sapiens GN=HINT1 PE=1 SV=2
150	P49755	0.9918	Transmembrane emp24 domain-containing protein 10 OS=Homo sapiens GN=TMED10 PE=1 SV=2
151	P49720	0.9989	Proteasome subunit beta type-3 OS=Homo sapiens GN=PSMB3 PE=1 SV=2
152	P49588	0.9918	Alanine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=AARS PE=1 SV=2
153	P48739-2	0.9878	Phosphatidylinositol transfer protein beta isoform, isoform 2
154	P48637-2	1.0000	Glutathione synthetase, isoform 2
155	P48059-2	0.9918	Isoform 2 of LIM and senescent cell antigen-like-containing domain protein 1 OS=Homo sapiens GN=LIMS1
156	P47914	0.9634	60S ribosomal protein L29 OS=Homo sapiens GN=RPL29 PE=1 SV=2
157	P47897	0.9995	Glutamine--tRNA ligase OS=Homo sapiens GN=QARS PE=1 SV=1
158	P43652	1.0000	Afamin OS=Homo sapiens GN=AFM PE=1 SV=1
159	P43243	0.9918	Matrin-3 OS=Homo sapiens GN=MATR3 PE=1 SV=2
160	P43034	0.9689	Platelet-activating factor acetylhydrolase IB subunit alpha OS=Homo sapiens GN=PAFAH1B1 PE=1 SV=2
161	P42566	0.9918	Epidermal growth factor receptor substrate 15 OS=Homo sapiens GN=EPS15 PE=1 SV=2
162	P41250	0.9918	Glycine--tRNA ligase OS=Homo sapiens GN=GARS PE=1 SV=3
163	P41222	0.9918	Prostaglandin-H2 D-isomerase OS=Homo sapiens GN=PTGDS PE=1 SV=1
164	P40939	0.9918	Trifunctional enzyme subunit alpha, mitochondrial OS=Homo sapiens GN=HADHA PE=1 SV=2
165	P39023	0.9758	60S ribosomal protein L3 OS=Homo sapiens GN=RPL3 PE=1 SV=2
166	P38646	0.9918	Stress-70 protein, mitochondrial OS=Homo sapiens GN=HSPA9 PE=1 SV=2

	UniProt ID	Probability	Protein Description
167	P38606	1.0000	V-type proton ATPase catalytic subunit A OS=Homo sapiens GN=ATP6V1A PE=1 SV=2
168	P37802	0.9918	Transgelin-2 OS=Homo sapiens GN=TAGLN2 PE=1 SV=3
169	P36957	0.9918	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial OS=Homo sa
170	P35555	0.9894	Fibrillin-1 OS=Homo sapiens GN=FBN1 PE=1 SV=3
171	P35244	0.9918	Replication protein A 14 kDa subunit OS=Homo sapiens GN=RPA3 PE=1 SV=1
172	P35237	1.0000	Serpin B6 OS=Homo sapiens GN=SERPINB6 PE=1 SV=3
173	P32969	0.9894	60S ribosomal protein L9 OS=Homo sapiens GN=RPL9 PE=1 SV=1
174	P31949	0.9918	Protein S100-A11 OS=Homo sapiens GN=S100A11 PE=1 SV=2
175	P31948	0.9918	Stress-induced-phosphoprotein 1 OS=Homo sapiens GN=STIP1 PE=1 SV=1
176	P31946-2	1.0000	Isoform Short of 14-3-3 protein beta/alpha OS=Homo sapiens GN=YWHAB
177	P31942-2	0.9918	Heterogeneous nuclear ribonucleoprotein H3, isoform 2
178	P31939	1.0000	Bifunctional purine biosynthesis protein PURH OS=Homo sapiens GN=ATIC PE=1 SV=3
179	P30626-2	1.0000	Sorcin, isoform 2
180	P30086	1.0000	Phosphatidylethanolamine-binding protein 1 OS=Homo sapiens GN=PEBP1 PE=1 SV=3
181	P30084	0.9918	Enoyl-CoA hydratase, mitochondrial OS=Homo sapiens GN=ECHS1 PE=1 SV=4
182	P29762	1.0000	Cellular retinoic acid-binding protein 1 OS=Homo sapiens GN=CRABP1 PE=2 SV=2
183	P29218	0.9999	Inositol monophosphatase 1 OS=Homo sapiens GN=IMPA1 PE=1 SV=1
184	P28838-2	1.0000	Isoform 2 of Cytosol aminopeptidase OS=Homo sapiens GN=LAP3
185	P28074	1.0000	Proteasome subunit beta type-5 OS=Homo sapiens GN=PSMB5 PE=1 SV=3
186	P27816-2	1.0000	Microtubule-associated protein 4, isoform 2
187	P27797	0.9918	Calreticulin OS=Homo sapiens GN=CALR PE=1 SV=1
188	P27348	1.0000	14-3-3 protein theta OS=Homo sapiens GN=YWHAQ PE=1 SV=1
189	P26599-2	1.0000	Polypyrimidine tract-binding protein 1, isoform 2
190	P26022	1.0000	Pentraxin-related protein PTX3 OS=Homo sapiens GN=PTX3 PE=1 SV=3
191	P25789	0.9918	Proteasome subunit alpha type-4 OS=Homo sapiens GN=PSMA4 PE=1 SV=1
192	P25786-2	0.9918	Proteasome subunit alpha type-1, isoform long
193	P25705	1.0000	ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
194	P25686-2	0.9918	DnaJ homolog subfamily B member 2, isoform 2
195	P24752	0.9672	Acetyl-CoA acetyltransferase, mitochondrial OS=Homo sapiens GN=ACAT1 PE=1 SV=1
196	P23396	1.0000	40S ribosomal protein S3 OS=Homo sapiens GN=RPS3 PE=1 SV=2
197	P22626	1.0000	Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2
198	P22352	0.9918	Glutathione peroxidase 3 OS=Homo sapiens GN=GPX3 PE=1 SV=2
199	P22061-2	1.0000	Protein-L-isoaspartate(D-aspartate) O-methyltransferase, isoform 2
200	P21912	0.9918	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial OS=Homo sapiens GN=SDHB PE=1 SV=3
201	P21796	0.9918	Voltage-dependent anion-selective channel protein 1 OS=Homo sapiens GN=VDAC1 PE=1 SV=2
202	P21266	0.9918	Glutathione S-transferase Mu 3 OS=Homo sapiens GN=GSTM3 PE=1 SV=3
203	P20929	0.9795	Nebulin OS=Homo sapiens GN=NEB PE=1 SV=4
204	P20711-2	0.9733	Isoform 2 of Aromatic-L-amino-acid decarboxylase OS=Homo sapiens GN=DDC
205	P20073-2	1.0000	Annexin A7, isoform 2
206	P19652	1.0000	Alpha-1-acid glycoprotein 2 OS=Homo sapiens GN=ORM2 PE=1 SV=2
207	P18583-10	0.9918	Protein SON, isoform J
208	P17987	1.0000	T-complex protein 1 subunit alpha OS=Homo sapiens GN=TCP1 PE=1 SV=1
209	P17980	0.9918	26S protease regulatory subunit 6A OS=Homo sapiens GN=PSMC3 PE=1 SV=3
210	P17931	0.9918	Galectin-3 OS=Homo sapiens GN=LGALS3 PE=1 SV=5
211	P17858	1.0000	6-phosphofructokinase, liver type OS=Homo sapiens GN=PFKL PE=1 SV=6
212	P17844	1.0000	Probable ATP-dependent RNA helicase DDX5 OS=Homo sapiens GN=DDX5 PE=1 SV=1
213	P17174	0.9878	Aspartate aminotransferase, cytoplasmic OS=Homo sapiens GN=GOT1 PE=1 SV=3
214	P16930	0.9918	Fumarylacetoacetase OS=Homo sapiens GN=FAH PE=1 SV=2
215	P16455	0.9918	Methylated-DNA--protein-cysteine methyltransferase OS=Homo sapiens GN=MGMT PE=1 SV=1
216	P16152	1.0000	Carbonyl reductase [NADPH] 1 OS=Homo sapiens GN=CBR1 PE=1 SV=3
217	P15153	0.9918	Ras-related C3 botulinum toxin substrate 2
218	P15121	1.0000	Aldose reductase OS=Homo sapiens GN=AKR1B1 PE=1 SV=3
219	P14868	0.9918	Aspartate--tRNA ligase, cytoplasmic OS=Homo sapiens GN=DARS PE=1 SV=2
220	P14866	0.9918	Heterogeneous nuclear ribonucleoprotein L OS=Homo sapiens GN=HNRNPL PE=1 SV=2
221	P14618	1.0000	Pyruvate kinase PKM OS=Homo sapiens GN=PKM PE=1 SV=4
222	P14550	1.0000	Alcohol dehydrogenase [NADP(+)] OS=Homo sapiens GN=AKR1A1 PE=1 SV=3
223	P14174	1.0000	Macrophage migration inhibitory factor OS=Homo sapiens GN=MIF PE=1 SV=4
224	P13942-2	0.9790	Collagen alpha-2(XI) chain, isoform 2

	UniProt ID	Probability	Protein Description
225	P13804	1.0000	Electron transfer flavoprotein subunit alpha, mitochondrial OS=Homo sapiens GN=ETFA PE=1 SV=1
226	P13693	0.9918	Translationally-controlled tumor protein OS=Homo sapiens GN=TPT1 PE=1 SV=1
227	P13667	0.9918	Protein disulfide-isomerase A4 OS=Homo sapiens GN=PDIA4 PE=1 SV=2
228	P13473-2	0.9910	Lysosome-associated membrane glycoprotein 2, isoform LAMP-2B
229	P11766	1.0000	Alcohol dehydrogenase class-3 OS=Homo sapiens GN=ADH5 PE=1 SV=4
230	P11498	0.9918	Pyruvate carboxylase, mitochondrial OS=Homo sapiens GN=PC PE=1 SV=2
231	P11413-2	0.9812	Glucose-6-phosphate 1-dehydrogenase, isoform long
232	P11233	0.9918	Ras-related protein Ral-A
233	P11142-2	1.0000	Heat shock cognate 71 kDa protein, isoform 2
234	P10644	0.9918	cAMP-dependent protein kinase type I-alpha regulatory subunit OS=Homo sapiens GN=PRKAR1A PE=1 SV=1
235	P0CG47	1.0000	Polyubiquitin-B
236	P0C869-6	0.9918	Cytosolic phospholipase A2 beta, isoform 5
237	P0C7T7	0.9588	Putative uncharacterized protein FRMD6-AS1 OS=Homo sapiens GN=FRMD6-AS1 PE=5 SV=1
238	P0C0L4	1.0000	Complement C4-A
239	P09972	1.0000	Fructose-bisphosphate aldolase C OS=Homo sapiens GN=ALDOC PE=1 SV=2
240	P09874	0.9918	Poly [ADP-ribose] polymerase 1 OS=Homo sapiens GN=PARP1 PE=1 SV=4
241	P09493-3	1.0000	Tropomyosin alpha-1 chain, isoform 3
242	P09488	0.9918	Glutathione S-transferase Mu 1
243	P09467	0.9918	Fructose-1,6-bisphosphatase 1 OS=Homo sapiens GN=FBP1 PE=1 SV=5
244	P08697-2	1.0000	Alpha-2-antiplasmin, isoform 2
245	P08603	1.0000	Complement factor H OS=Homo sapiens GN=CFH PE=1 SV=4
246	P08134	0.9918	Rho-related GTP-binding protein RhoC
247	P08133	1.0000	Annexin A6 OS=Homo sapiens GN=ANXA6 PE=1 SV=3
248	P0DMV9	1.0000	**Heat shock 70 kDa protein 1B**
249	P07858	0.9665	Cathepsin B OS=Homo sapiens GN=CTSB PE=1 SV=3
250	P07738	1.0000	Bisphosphoglycerate mutase OS=Homo sapiens GN=BPGM PE=1 SV=2
251	P07108-6	0.9918	Acyl-CoA-binding protein, isoform 6
252	P06744-2	1.0000	Glucose-6-phosphate isomerase, isoform 2
253	P06737-2	1.0000	Glycogen phosphorylase, liver form, isoform 2
254	P06733	1.0000	Alpha-enolase OS=Homo sapiens GN=ENO1 PE=1 SV=2
255	P05186-2	0.9918	**Alkaline phosphatase, tissue-nonspecific isozyme, isoform 2**
256	P05090	0.9918	Apolipoprotein D OS=Homo sapiens GN=APOD PE=1 SV=1
257	P04843	0.9918	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapiens GN=RPN1 PE=1 SV=1
258	P04632	1.0000	Calpain small subunit 1 OS=Homo sapiens GN=CAPNS1 PE=1 SV=1
259	P04438	0.9910	Ig heavy chain V-II region SESS OS=Homo sapiens PE=2 SV=1
260	P04430	0.9918	Ig kappa chain V-I region BAN OS=Homo sapiens PE=1 SV=1
261	P04424-2	0.9719	Argininosuccinate lyase, isoform 2
262	P04217	1.0000	Alpha-1B-glycoprotein OS=Homo sapiens GN=A1BG PE=1 SV=4
263	P04208	0.9998	Ig lambda chain V-I region WAH OS=Homo sapiens PE=1 SV=1
264	P04207	0.9996	Ig kappa chain V-III region CLL OS=Homo sapiens PE=1 SV=2
265	P04080	0.9918	Cystatin-B OS=Homo sapiens GN=CSTB PE=1 SV=2
266	P02814	0.9918	Submaxillary gland androgen-regulated protein 3B OS=Homo sapiens GN=SMR3B PE=1 SV=2
267	P02810	0.9918	Salivary acidic proline-rich phosphoprotein 1/2 OS=Homo sapiens GN=PRH1 PE=1 SV=2
268	P02792	1.0000	Ferritin light chain OS=Homo sapiens GN=FTL PE=1 SV=2
269	P02774-3	1.0000	Vitamin D-binding protein, isoform 3
270	P02763	0.9989	Alpha-1-acid glycoprotein 1 OS=Homo sapiens GN=ORM1 PE=1 SV=1
271	P02753	1.0000	Retinol-binding protein 4 OS=Homo sapiens GN=RBP4 PE=1 SV=3
272	P02649	0.9918	Apolipoprotein E OS=Homo sapiens GN=APOE PE=1 SV=1
273	P01861	0.9996	Ig gamma-4 chain C region OS=Homo sapiens GN=IGHG4 PE=1 SV=1
274	P01859	1.0000	Ig gamma-2 chain C region OS=Homo sapiens GN=IGHG2 PE=1 SV=2
275	P01857	1.0000	Ig gamma-1 chain C region OS=Homo sapiens GN=IGHG1 PE=1 SV=1
276	P01781	1.0000	Ig heavy chain V-III region GAL OS=Homo sapiens PE=1 SV=1
277	P01762	0.9918	Ig heavy chain V-III region TRO OS=Homo sapiens PE=1 SV=1
278	P01761	0.9918	Ig heavy chain V-I region SIE OS=Homo sapiens PE=1 SV=1
279	P01719	0.9918	Ig lambda chain V-V region DEL OS=Homo sapiens PE=1 SV=1
280	P01624	0.9996	Ig kappa chain V-III region POM OS=Homo sapiens PE=1 SV=1
281	P01623	1.0000	Ig kappa chain V-III region WOL OS=Homo sapiens PE=1 SV=1
282	P01616	0.9918	Ig kappa chain V-II region MIL OS=Homo sapiens PE=1 SV=1

	UniProt ID	Probability	Protein Description
283	P01612	0.9918	Ig kappa chain V-I region Mev OS=Homo sapiens PE=1 SV=1
284	P01023	1.0000	Alpha-2-macroglobulin OS=Homo sapiens GN=A2M PE=1 SV=3
285	P01009	1.0000	Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 PE=1 SV=3
286	P00966	1.0000	Argininosuccinate synthase OS=Homo sapiens GN=ASS1 PE=1 SV=2
287	P00918	1.0000	Carbonic anhydrase 2 OS=Homo sapiens GN=CA2 PE=1 SV=2
288	P00748	1.0000	Coagulation factor XII OS=Homo sapiens GN=F12 PE=1 SV=3
289	P00558	1.0000	Phosphoglycerate kinase 1 OS=Homo sapiens GN=PGK1 PE=1 SV=3
290	P00505	0.9918	Aspartate aminotransferase, mitochondrial OS=Homo sapiens GN=GOT2 PE=1 SV=3
291	O95865	1.0000	N(G),N(G)-dimethylarginine dimethylaminohydrolase 2 OS=Homo sapiens GN=DDAH2 PE=1 SV=1
292	O95861-2	0.9918	3'(2'),5'-bisphosphate nucleotidase 1, isoform 2
293	O95816	0.9910	BAG family molecular chaperone regulator 2 OS=Homo sapiens GN=BAG2 PE=1 SV=1
294	O95777	0.9918	N-alpha-acetyltransferase 38, NatC auxiliary subunit OS=Homo sapiens GN=NAA38 PE=1 SV=3
295	O95372	0.9918	Acyl-protein thioesterase 2 OS=Homo sapiens GN=LYPLA2 PE=1 SV=1
296	O94979-2	0.9910	Protein transport protein Sec31A, isoform 2
297	O94903	0.9918	Proline synthase co-transcribed bacterial homolog protein OS=Homo sapiens GN=PROSC PE=1 SV=1
298	O94760	0.9918	N(G),N(G)-dimethylarginine dimethylaminohydrolase 1 OS=Homo sapiens GN=DDAH1 PE=1 SV=3
299	O75935-3	1.0000	Dynactin subunit 3, isoform 3
300	O75915	0.9918	PRA1 family protein 3 OS=Homo sapiens GN=ARL6IP5 PE=1 SV=1
301	O75874	1.0000	Isocitrate dehydrogenase [NADP] cytoplasmic OS=Homo sapiens GN=IDH1 PE=1 SV=2
302	O75746	0.9918	Calcium-binding mitochondrial carrier protein Aralar1 OS=Homo sapiens GN=SLC25A12 PE=1 SV=2
303	O75396	0.9649	Vesicle-trafficking protein SEC22b OS=Homo sapiens GN=SEC22B PE=1 SV=4
304	O75369-2	1.0000	Filamin-B,isoform 2
305	O75367-2	1.0000	Core histone macro-H2A.1, isoform 1
306	O75131	0.9846	Copine-3
307	O60664-2	0.9782	Perilipin-3, isoform 2
308	O60271-2	1.0000	C-Jun-amino-terminal kinase-interacting protein 4, isoform 2
309	O60234	0.9918	Glia maturation factor gamma OS=Homo sapiens GN=GMFG PE=1 SV=1
310	O43707-2	1.0000	Alpha-actinin-4, Isoform ACTN4ISO
311	O43598-2	0.9910	2'-deoxynucleoside 5'-phosphate N-hydrolase 1, isoform 2
312	O43488	1.0000	Aflatoxin B1 aldehyde reductase member 2 OS=Homo sapiens GN=AKR7A2 PE=1 SV=3
313	O43390-2	0.9854	Heterogeneous nuclear ribonucleoprotein R, isoform 2
314	O15511-2	0.9918	Actin-related protein 2/3 complex subunit 5, isoform 2
315	O15428	0.9918	Putative PIN1-like protein
316	O15144	1.0000	Actin-related protein 2/3 complex subunit 2 OS=Homo sapiens GN=ARPC2 PE=1 SV=1
317	O15061-2	1.0000	Synemin
318	O14980	0.9918	Exportin-1 OS=Homo sapiens GN=XPO1 PE=1 SV=1
319	O14979-2	0.9918	Heterogeneous nuclear ribonucleoprotein D-like, isoform 2
320	O14907	0.9918	Tax1-binding protein 3 OS=Homo sapiens GN=TAX1BP3 PE=1 SV=2
321	O14818-4	0.9997	Proteasome subunit alpha type-7, isoform 3
322	O14744	0.9918	Protein arginine N-methyltransferase 5 OS=Homo sapiens GN=PRMT5 PE=1 SV=4
323	O14576-2	0.9822	Cytoplasmic dynein 1 intermediate chain 1, isoform 2
324	O00629	0.9918	Importin subunit alpha-3 OS=Homo sapiens GN=KPNA4 PE=1 SV=1
325	O00299	1.0000	Chloride intracellular channel protein 1 OS=Homo sapiens GN=CLIC1 PE=1 SV=4
326	O00231-2	0.9918	26S proteasome non-ATPase regulatory subunit 11, isoform 2
327	O00159-3	0.9836	Unconventional myosin-Ic, isoform 3
328	O00148	1.0000	ATP-dependent RNA helicase DDX39A
329	B4E2M5	0.9918	Ankyrin repeat domain-containing protein 66 OS=Homo sapiens GN=ANKRD66 PE=2 SV=2
330	A6NHG4	0.9918	D-dopachrome decarboxylase-like protein
331	A6NDU8	0.9504	UPF0600 protein C5orf51 OS=Homo sapiens GN=C5orf51 PE=1 SV=1

Table 4. Mature odontoblast proteins

Proteins found only in odontoblast cell layer samples from donors > 20 (N = 2, n = 8). (Data also shown in Figure 4D).			
	UniProt ID	Probability	Protein Description
1	Q9Y6U3-2	1.0000	Adseverin, isoform 2
2	Q9Y4Y9	0.9921	U6 snRNA-associated Sm-like protein LSm5 OS=Homo sapiens GN=LSM5 PE=1 SV=3
3	Q9Y4D7-2	0.9707	Plexin-D1, isoform 2
4	Q9Y3D6	0.9895	Mitochondrial fission 1 protein OS=Homo sapiens GN=FIS1 PE=1 SV=2
5	Q9Y2Z4	0.9502	Tyrosine--tRNA ligase, mitochondrial OS=Homo sapiens GN=YARS2 PE=1 SV=2
6	Q9Y2H0-1	0.9866	Disks large-associated protein 4, isoform 2
7	Q9Y2G9	0.9926	Protein strawberry notch homolog 2 OS=Homo sapiens GN=SBNO2 PE=2 SV=3
8	Q9Y2E4	0.9837	Disco-interacting protein 2 homolog C OS=Homo sapiens GN=DIP2C PE=2 SV=2
9	Q9Y259	0.9727	Choline/ethanolamine kinase OS=Homo sapiens GN=CHKB PE=1 SV=3
10	Q9Y257	0.9968	Potassium channel subfamily K member 6 OS=Homo sapiens GN=KCNK6 PE=1 SV=1
11	Q9UQQ1-2	0.9885	N-acetylated-alpha-linked acidic dipeptidase-like protein, isoform 2
12	Q9UQP3	1.0000	Tenascin-N OS=Homo sapiens GN=TNN PE=1 SV=2
13	Q9UQ80	0.9635	Proliferation-associated protein 2G4 OS=Homo sapiens GN=PA2G4 PE=1 SV=3
14	Q9UQ16-2	0.9984	Dynamin-3, isoform 2
15	Q9UPW8	0.9829	Protein unc-13 homolog A OS=Homo sapiens GN=UNC13A PE=2 SV=4
16	Q9UPV0-2	0.9999	Centrosomal protein of 164 kDa, isoform 2
17	Q9UPN3-2	0.9590	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5, isoform 2
18	Q9UP83-2	0.9635	Conserved oligomeric Golgi complex subunit 5, isoform 2
19	Q9UNS2	0.9937	COP9 signalosome complex subunit 3 OS=Homo sapiens GN=COPS3 PE=1 SV=3
20	Q9UNH5-2	0.9647	Dual specificity protein phosphatase CDC14A, isoform 2
21	Q9UNF0-2	0.9947	Protein kinase C and casein kinase substrate in neurons protein 2, isoform 2
22	Q9UMS4	0.9937	Pre-mRNA-processing factor 19 OS=Homo sapiens GN=PRPF19 PE=1 SV=1
23	Q9ULC6	0.9924	Protein-arginine deiminase type-1 OS=Homo sapiens GN=PADI1 PE=1 SV=2
24	Q9UL46	0.9997	Proteasome activator complex subunit 2 OS=Homo sapiens GN=PSME2 PE=1 SV=4
25	Q9UKY1	0.9801	Zinc fingers and homeoboxes protein 1 OS=Homo sapiens GN=ZHX1 PE=1 SV=1
26	Q9UJ70-2	0.9838	N-acetyl-D-glucosamine kinase, isoform 2
27	Q9UHX3-2	0.9601	Adhesion G protein-coupled receptor E2, isoform 2
28	Q9UHD8-5	0.9994	Isoform 5 of Septin-9 OS=Homo sapiens GN=SEPT9
29	Q9UFH2-2	0.9991	Dynein heavy chain 17, axonemal, isoform 2
30	Q9P2P6	0.9840	StAR-related lipid transfer protein 9 OS=Homo sapiens GN=STARD9 PE=1 SV=3
31	Q9P2J2-2	0.9608	Protein turtle homolog A, isoform 2
32	Q9P2E9-2	1.0000	Ribosome-binding protein 1, isoform 1
33	Q9P2E5-2	0.9852	Chondroitin sulfate glucuronyltransferase, isoform 2
34	Q9P265	0.9997	Disco-interacting protein 2 homolog B OS=Homo sapiens GN=DIP2B PE=1 SV=3
35	Q9P0X4-4	0.9870	Voltage-dependent T-type calcium channel subunit alpha-1I, isoform 4
36	Q9P0J1	0.9993	[Pyruvate dehydrogenase [acetyl-transferring]]-phosphatase 1, mitochondrial OS=Homo sapiens GN=PDP1 PE=1 SV=3
37	Q9NZJ4-2	0.9953	Sacsin, isoform 2
38	Q9NZB8-6	0.9621	Molybdenum cofactor biosynthesis protein 1, isoform 2
39	Q9NY15	0.9658	Stabilin-1 OS=Homo sapiens GN=STAB1 PE=1 SV=3
40	Q9NVN3-1	0.9688	Synembryn-B, isoform 1
41	Q9NVA2-2	0.9998	Septin-11, isoform 2
42	Q9NRX4-2	0.9921	14 kDa phosphohistidine phosphatase, isoform 2
43	Q9NR99	0.9655	Matrix-remodeling-associated protein 5 OS=Homo sapiens GN=MXRA5 PE=2 SV=3
44	Q9HCL2	0.9749	Glycerol-3-phosphate acyltransferase 1, mitochondrial OS=Homo sapiens GN=GPAM PE=1 SV=3
45	Q9HCK4-2	0.9502	Roundabout homolog 2, isoform 2
46	Q9HA92	0.9918	Radical S-adenosyl methionine domain-containing protein 1, mitochondrial OS=Homo sapiens GN=RSAD1 PE=2 SV=2
47	Q9H9Y4	0.9954	GPN-loop GTPase 2 OS=Homo sapiens GN=GPN2 PE=2 SV=2
48	Q9H9D4	0.9935	Zinc finger protein 408 OS=Homo sapiens GN=ZNF408 PE=1 SV=1
49	Q9H3T3-3	0.9998	Semaphorin-6B, isoform 2
50	Q9H3R0-2	0.9698	Lysine-specific demethylase 4C, isoform 2
51	Q9H2G9	0.9999	Golgin-45 OS=Homo sapiens GN=BLZF1 PE=1 SV=2
52	Q9H2D6-2	0.9644	TRIO and F-actin-binding protein, isoform 3
53	Q9H299	0.9670	SH3 domain-binding glutamic acid-rich-like protein 3 OS=Homo sapiens GN=SH3BGRL3 PE=1 SV=1
54	Q9H251-2	0.9587	Cadherin-23, isoform 2

	UniProt ID	Probability	Protein Description
55	Q9GZX5	0.9999	Zinc finger protein 350 OS=Homo sapiens GN=ZNF350 PE=1 SV=3
56	Q9C0K0-2	0.9984	B-cell lymphoma/leukemia 11B, isoform 2
57	Q9C0H9-2	0.9999	SRC kinase signaling inhibitor 1, isoform 2
58	Q9BZZ2-2	0.9990	Sialoadhesin, isoform 2
59	Q9BYB0-2	0.9603	SH3 and multiple ankyrin repeat domains protein 3
60	Q9BXN1	1.0000	Asporin OS=Homo sapiens GN=ASPN PE=1 SV=2
61	Q9BX69	0.9574	Caspase recruitment domain-containing protein 6 OS=Homo sapiens GN=CARD6 PE=2 SV=2
62	Q9BV73-2	0.9987	Centrosome-associated protein CEP250, isoform 2
63	Q9BPX3	0.9909	Condensin complex subunit 3 OS=Homo sapiens GN=NCAPG PE=1 SV=1
64	Q99715	1.0000	Collagen alpha-1(XII) chain OS=Homo sapiens GN=COL12A1 PE=1 SV=2
65	Q96S55-2	0.9638	ATPase WRNIP1, isoform 2
66	Q96RW7-2	0.9998	Hemicentin-1, isoform 2
67	Q96PF1	0.9716	Protein-glutamine gamma-glutamyltransferase Z OS=Homo sapiens GN=TGM7 PE=2 SV=1
68	Q96P70	0.9947	Importin-9 OS=Homo sapiens GN=IPO9 PE=1 SV=3
69	Q96NL6	0.9960	Sodium channel and clathrin linker 1 OS=Homo sapiens GN=SCLT1 PE=2 SV=2
70	Q96M86	0.9631	Dynein heavy chain domain-containing protein 1 OS=Homo sapiens GN=DNHD1 PE=2 SV=2
71	Q96KP4-2	0.9947	Cytosolic non-specific dipeptidase, isoform 2
72	Q96JQ0	1.0000	Protocadherin-16 OS=Homo sapiens GN=DCHS1 PE=2 SV=1
73	Q96JH8-1	0.9991	Ras-associating and dilute domain-containing protein, isoform 1
74	Q96IT1-2	0.9867	Zinc finger protein 496, isoform 2
75	Q96HQ2-2	0.9890	CDKN2AIP N-terminal-like protein, isoform 2
76	Q96HB5-2	0.9881	Coiled-coil domain-containing protein 120, isoform 2
77	Q96FT7-2	0.9632	Acid-sensing ion channel 4, isoform 2
78	Q96EH3	0.9541	Mitochondrial assembly of ribosomal large subunit protein 1 OS=Homo sapiens GN=MALSU1 PE=1 SV=1
79	Q96DZ1-2	0.9979	Endoplasmic reticulum lectin 1, isoform 2
80	Q96DT5	0.9984	Dynein heavy chain 11, axonemal OS=Homo sapiens GN=DNAH11 PE=1 SV=3
81	Q96AX9-10	0.9936	E3 ubiquitin-protein ligase MIB2, isoform 10
82	Q96AC6	0.9579	Kinesin-like protein KIFC2 OS=Homo sapiens GN=KIFC2 PE=2 SV=1
83	Q96A99-2	0.9999	Pentraxin-4, isoform 1
84	Q969S9-2	0.9966	Ribosome-releasing factor 2, mitochondrial, isoform 2
85	Q92994-3	0.9980	Transcription factor IIIB 90 kDa subunit, isoform 3
86	Q92985-2	1.0000	Interferon regulatory factor 7, Isoform B
87	Q92959	0.9650	Solute carrier organic anion transporter family member 2A1 OS=Homo sapiens GN=SLCO2A1 PE=1 SV=2
88	Q92917	0.9628	G patch domain and KOW motifs-containing protein OS=Homo sapiens GN=GPKOW PE=1 SV=2
89	Q92817	0.9671	Envoplakin OS=Homo sapiens GN=EVPL PE=1 SV=3
90	Q92805	0.9670	Golgin subfamily A member 1 OS=Homo sapiens GN=GOLGA1 PE=1 SV=3
91	Q8WZ42-1	1.0000	Titin, isoform 1
92	Q8WXI7	0.9975	Mucin-16 OS=Homo sapiens GN=MUC16 PE=1 SV=2
93	Q8WXH0-2	0.9794	Nesprin-2, isoform 2
94	Q8WUA2	0.9937	Peptidyl-prolyl cis-trans isomerase-like 4 OS=Homo sapiens GN=PPIL4 PE=1 SV=1
95	Q8TEY5-2	0.9991	Cyclic AMP-responsive element-binding protein 3-like protein 4, isoform 2
96	Q8TDY2-2	0.9957	RB1-inducible coiled-coil protein 1, isoform 2
97	Q8TDX9-2	0.9970	Polycystic kidney disease protein 1-like 1, isoform 2
98	Q8TC90	0.9560	Coiled-coil domain-containing glutamate-rich protein 1 OS=Homo sapiens GN=CCER1 PE=2 SV=1
99	Q8TBE0-2	0.9710	Bromo adjacent homology domain-containing 1 protein, isoform 2
100	Q8NI27-2	0.9680	THO complex subunit 2, isoform 2
101	Q8NG06	0.9610	Tripartite motif-containing protein 58 OS=Homo sapiens GN=TRIM58 PE=2 SV=2
102	Q8NFU4	0.9855	Follicular dendritic cell secreted peptide OS=Homo sapiens GN=FDCSP PE=1 SV=1
103	Q8NF91-4	0.9997	Nesprin-1, isoform 4
104	Q8NDA2-3	0.9999	Hemicentin-2
105	Q8ND90	0.9595	Paraneoplastic antigen Ma1 OS=Homo sapiens GN=PNMA1 PE=1 SV=2
106	Q8ND04-2	0.9938	Protein SMG8, isoform 2
107	Q8NCN4	0.9881	E3 ubiquitin-protein ligase RNF169 OS=Homo sapiens GN=RNF169 PE=1 SV=2
108	Q8NBJ5	0.9923	Procollagen galactosyltransferase 1 OS=Homo sapiens GN=COLGALT1 PE=1 SV=1
109	Q8N7J2-2	0.9814	APC membrane recruitment protein 2, isoform 2
110	Q8N6Y0	0.9713	Usher syndrome type-1C protein-binding protein 1 OS=Homo sapiens GN=USHBP1 PE=1 SV=1
111	Q8N3C0	0.9961	Activating signal cointegrator 1 complex subunit 3 OS=Homo sapiens GN=ASCC3 PE=1 SV=3
112	Q8N2S1	0.9956	Latent-transforming growth factor beta-binding protein 4 OS=Homo sapiens GN=LTBP4 PE=1 SV=2
113	Q8N2C7-2	0.9693	Protein unc-80 homolog, isoform 2
114	Q8N108-16	0.9879	Mesoderm induction early response protein 1, isoform 6
115	Q8IZD2-2	0.9738	Histone-lysine N-methyltransferase 2E, isoform 2
116	Q8IWY9-1	0.9976	Codanin-1, isoform 1
117	Q8IWV8-4	0.9917	E3 ubiquitin-protein ligase UBR2, isoform 4

	UniProt ID	Probability	Protein Description
118	Q8IWA0	0.9978	WD repeat-containing protein 75 OS=Homo sapiens GN=WDR75 PE=1 SV=1
119	Q8IVS8-2	0.9756	Glycerate kinase, isoform 2
120	Q8IVL1-10	0.9652	Neuron navigator 2, isoform 10
121	Q8IVF4	0.9931	Dynein heavy chain 10, axonemal OS=Homo sapiens GN=DNAH10 PE=1 SV=4
122	Q8IUX7	1.0000	Adipocyte enhancer-binding protein 1 OS=Homo sapiens GN=AEBP1 PE=1 SV=1
123	Q86YV0-2	0.9919	RAS protein activator like-3, isoform 2
124	Q86XR2-2	0.9920	Niban-like protein 2, isoform 2
125	Q86VL8-3	0.9640	Multidrug and toxin extrusion protein 2, isoform 3
126	Q86UT6-2	0.9991	NLR family member X1, isoform 2
127	Q7Z7L9-3	0.9885	Zinc finger and SCAN domain-containing protein 2, isoform 3
128	Q7Z6M4	0.9989	mTERF domain-containing protein 2 OS=Homo sapiens GN=MTERFD2 PE=1 SV=3
129	Q7Z6I6-2	0.9905	Rho GTPase-activating protein 30, isoform 2
130	Q7Z4H8	0.9704	KDEL motif-containing protein 2 OS=Homo sapiens GN=KDELC2 PE=1 SV=2
131	Q7Z3K6-2	0.9739	Mesoderm induction early response protein 3, isoform 2
132	Q7KZF4	1.0000	Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens GN=SND1 PE=1 SV=1
133	Q76I76	0.9961	Protein phosphatase Slingshot homolog 2 OS=Homo sapiens GN=SSH2 PE=1 SV=1
134	Q71U36-2	1.0000	Tubulin alpha-1A chain, isoform 2
135	Q6ZSJ9-2	0.9991	Protein shisa-6 homolog, isoform 2
136	Q6ZRR7	0.9844	Leucine-rich repeat-containing protein 9 OS=Homo sapiens GN=LRRC9 PE=2 SV=2
137	Q6UVK1	0.9619	Chondroitin sulfate proteoglycan 4 OS=Homo sapiens GN=CSPG4 PE=1 SV=2
138	Q6PIJ6	0.9825	F-box only protein 38 OS=Homo sapiens GN=FBXO38 PE=1 SV=3
139	Q6P1N0-2	0.9839	Coiled-coil and C2 domain-containing protein 1A, isoform 2
140	Q6P1J9	0.9905	Parafibromin OS=Homo sapiens GN=CDC73 PE=1 SV=1
141	Q68DN1	0.9844	Uncharacterized protein C2orf16 OS=Homo sapiens GN=C2orf16 PE=2 SV=3
142	Q63HM1-2	0.9994	Kynurenine formamidase, isoform 2
143	Q5VUA4-2	0.9982	Zinc finger protein 318, isoform 2
144	Q5VTR2	0.9700	E3 ubiquitin-protein ligase BRE1A OS=Homo sapiens GN=RNF20 PE=1 SV=2
145	Q5VST9-2	1.0000	Obscurin, isoform 2
146	Q5U651	1.0000	Ras-interacting protein 1 OS=Homo sapiens GN=RASIP1 PE=1 SV=1
147	Q5T4B2	0.9921	Probable inactive glycosyltransferase 25 family member 3 OS=Homo sapiens GN=CERCAM PE=2 SV=1
148	Q5T011-5	0.9813	Protein SZT2, isoform 3
149	Q5SNT2	0.9703	Transmembrane protein 201 OS=Homo sapiens GN=TMEM201 PE=1 SV=1
150	Q5RHP9-2	0.9695	Glutamate-rich protein 3, isoform 2
151	Q5JT82	0.9974	Krueppel-like factor 17 OS=Homo sapiens GN=KLF17 PE=1 SV=1
152	Q58EX7-2	0.9987	Puratrophin-1, isoform 2
153	Q53GL7	0.9929	Poly [ADP-ribose] polymerase 10 OS=Homo sapiens GN=PARP10 PE=1 SV=2
154	Q53EQ6-2	0.9810	Tigger transposable element-derived protein 5, isoform 2
155	Q4LDE5-2	0.9753	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1, isoform 2
156	Q27J81-2	0.9925	Inverted formin-2, isoform 2
157	Q16822	0.9973	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=3
158	Q16760-2	0.9949	Diacylglycerol kinase delta, isoform 1
159	Q16555-2	1.0000	Dihydropyrimidinase-related protein 2, isoform 2
160	Q16363-2	0.9777	Laminin subunit alpha-4, isoform 2
161	Q15878-2	0.9987	Voltage-dependent R-type calcium channel subunit alpha-1E, isoform 2
162	Q15772-1	0.9997	Striated muscle preferentially expressed protein kinase, isoform 1
163	Q15751	0.9959	Probable E3 ubiquitin-protein ligase HERC1 OS=Homo sapiens GN=HERC1 PE=1 SV=2
164	Q15233-2	0.9992	Non-POU domain-containing octamer-binding protein, isoform 2
165	Q15063	1.0000	Periostin OS=Homo sapiens GN=POSTN PE=1 SV=2
166	Q15019	0.9999	Septin-2 OS=Homo sapiens GN=SEPT2 PE=1 SV=1
167	Q14974	1.0000	Importin subunit beta-1 OS=Homo sapiens GN=KPNB1 PE=1 SV=2
168	Q14790-2	0.9991	Caspase-8, isoform 2
169	Q14764	0.9761	Major vault protein OS=Homo sapiens GN=MVP PE=1 SV=4
170	Q14315-2	0.9926	Filamin-C, isoform 2
171	Q14191	0.9968	Werner syndrome ATP-dependent helicase OS=Homo sapiens GN=WRN PE=1 SV=2
172	Q14011	1.0000	Cold-inducible RNA-binding protein OS=Homo sapiens GN=CIRBP PE=1 SV=1
173	Q13585	0.9866	Melatonin-related receptor OS=Homo sapiens GN=GPR50 PE=1 SV=3
174	Q13370	0.9982	cGMP-inhibited 3',5'-cyclic phosphodiesterase B OS=Homo sapiens GN=PDE3B PE=1 SV=2
175	Q13228-2	1.0000	Selenium-binding protein 1, isoform 2
176	Q13191-2	0.9534	E3 ubiquitin-protein ligase CBL-B, isoform Truncated 1
177	Q13085-2	0.9847	Acetyl-CoA carboxylase 1, isoform 2
178	Q12931	0.9999	Heat shock protein 75 kDa, mitochondrial OS=Homo sapiens GN=TRAP1 PE=1 SV=3
179	Q12765-2	0.9954	Secernin-1, isoform 2
180	Q08043	0.9864	Alpha-actinin-3 OS=Homo sapiens GN=ACTN3 PE=1 SV=2

	UniProt ID	Probability	Protein Description
181	Q07065	1.0000	Cytoskeleton-associated protein 4 OS=Homo sapiens GN=CKAP4 PE=1 SV=2
182	Q05707-2	1.0000	Collagen alpha-1(XIV) chain, isoform 2
183	Q04323-2	0.9981	UBX domain-containing protein 1, isoform 2
184	Q02388-2	0.9749	Collagen alpha-1(VII) chain, isoform 2
185	Q01082-2	0.9738	Spectrin beta chain, non-erythrocytic 1, isoform short
186	Q00587-2	0.9778	Cdc42 effector protein 1, isoform 2
187	Q00341	0.9971	Vigilin OS=Homo sapiens GN=HDLBP PE=1 SV=2
188	P98161-2	0.9874	Polycystin-1, isoform 2
189	P98160	1.0000	Basement membrane-specific heparan sulfate proteoglycan core protein OS=Homo sapiens GN=HSPG2 PE=1 SV=4
190	P80511	0.9916	Protein S100-A12 OS=Homo sapiens GN=S100A12 PE=1 SV=2
191	P78527-2	0.9998	DNA-dependent protein kinase catalytic subunit, isoform 2
192	P78509-2	0.9819	Reelin, isoform 2
193	P78415	0.9963	Iroquois-class homeodomain protein IRX-3 OS=Homo sapiens GN=IRX3 PE=2 SV=3
194	P69891	1.0000	Hemoglobin subunit gamma-1
195	P62917	1.0000	60S ribosomal protein L8 OS=Homo sapiens GN=RPL8 PE=1 SV=2
196	P62306	0.9947	Small nuclear ribonucleoprotein F OS=Homo sapiens GN=SNRPF PE=1 SV=1
197	P62269	0.9958	40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=3
198	P61626	0.9989	Lysozyme C OS=Homo sapiens GN=LYZ PE=1 SV=1
199	P60983	0.9997	Glia maturation factor beta OS=Homo sapiens GN=GMFB PE=1 SV=2
200	P56945-2	0.9833	Breast cancer anti-estrogen resistance protein 1, isoform 2
201	P55735	0.9716	Protein SEC13 homolog OS=Homo sapiens GN=SEC13 PE=1 SV=3
202	P54727-2	0.9987	UV excision repair protein RAD23 homolog B, isoform 2
203	P53609-2	0.9707	Geranylgeranyl transferase type-1 subunit beta, isoform 2
204	P51991-2	0.9947	Heterogeneous nuclear ribonucleoprotein A3, isoform 2
205	P51608-2	0.9947	Isoform B of Methyl-CpG-binding protein 2 OS=Homo sapiens GN=MECP2
206	P49736	0.9566	DNA replication licensing factor MCM2 OS=Homo sapiens GN=MCM2 PE=1 SV=4
207	P49641-1	0.9842	Alpha-mannosidase 2x, isoform 1
208	P46063	0.9947	ATP-dependent DNA helicase Q1 OS=Homo sapiens GN=RECQL PE=1 SV=3
209	P43355	0.9741	Melanoma-associated antigen 1 OS=Homo sapiens GN=MAGEA1 PE=1 SV=1
210	P41091	0.9926	Eukaryotic translation initiation factor 2 subunit 3, isoform 2
211	P40692	0.9997	DNA mismatch repair protein Mlh1 OS=Homo sapiens GN=MLH1 PE=1 SV=1
212	P37837	1.0000	Transaldolase OS=Homo sapiens GN=TALDO1 PE=1 SV=2
213	P36578	0.9755	60S ribosomal protein L4 OS=Homo sapiens GN=RPL4 PE=1 SV=5
214	P36542-2	0.9591	Isoform Heart of ATP synthase subunit gamma, mitochondrial OS=Homo sapiens GN=ATP5C1
215	P35232	0.9926	Prohibitin OS=Homo sapiens GN=PHB PE=1 SV=1
216	P35125-2	0.9879	Ubiquitin carboxyl-terminal hydrolase 6, isoform 2
217	P33241	0.9843	Lymphocyte-specific protein 1 OS=Homo sapiens GN=LSP1 PE=1 SV=1
218	P31943	0.9563	Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens GN=HNRNPH1 PE=1 SV=4
219	P30153	0.9958	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform OS=Homo sapiens GN=PPP2R1A PE=1 SV=4
220	P30050	0.9942	60S ribosomal protein L12 OS=Homo sapiens GN=RPL12 PE=1 SV=1
221	P30044-2	0.9947	Peroxiredoxin-5, mitochondrial, isoform Cytoplasmic + peroxisomal
222	P28300	0.9947	Protein-lysine 6-oxidase OS=Homo sapiens GN=LOX PE=1 SV=2
223	P27695	0.9964	DNA-(apurinic or apyrimidinic site) lyase OS=Homo sapiens GN=APEX1 PE=1 SV=2
224	P27348	0.9947	14-3-3 protein theta OS=Homo sapiens GN=YWHAQ PE=1 SV=1
225	P26447	0.9950	Protein S100-A4 OS=Homo sapiens GN=S100A4 PE=1 SV=1
226	P25705-2	1.0000	ATP synthase subunit alpha, mitochondrial, isoform 2
227	P25398	1.0000	40S ribosomal protein S12 OS=Homo sapiens GN=RPS12 PE=1 SV=3
228	P24158	0.9833	Myeloblastin OS=Homo sapiens GN=PRTN3 PE=1 SV=3
229	P23142	1.0000	Fibulin-1 OS=Homo sapiens GN=FBLN1 PE=1 SV=4
230	P22626-2	1.0000	Heterogeneous nuclear ribonucleoproteins A2/B1, isoform A2
231	P21817-2	0.9991	Ryanodine receptor 1, isoform 2
232	P20908	0.9984	Collagen alpha-1(V) chain OS=Homo sapiens GN=COL5A1 PE=1 SV=3
233	P20160	1.0000	Azurocidin OS=Homo sapiens GN=AZU1 PE=1 SV=3
234	P18135	1.0000	Ig kappa chain V-III region HAH
235	P16157-10	0.9812	Ankyrin-1, isoform Er9
236	P14625	1.0000	Endoplasmin OS=Homo sapiens GN=HSP90B1 PE=1 SV=1
237	P14618-2	1.0000	Pyruvate kinase PKM, isoform M1
238	P14314-2	0.9997	Glucosidase 2 subunit beta, isoform 2
239	P13861	0.9947	cAMP-dependent protein kinase type II-alpha regulatory subunit OS=Homo sapiens GN=PRKAR2A PE=1 SV=2
240	P13796	0.9979	Plastin-2 OS=Homo sapiens GN=LCP1 PE=1 SV=6
241	P13727	0.9993	Bone marrow proteoglycan OS=Homo sapiens GN=PRG2 PE=1 SV=2
242	P13497-4	0.9615	Isoform BMP1-5 of Bone morphogenetic protein 1 OS=Homo sapiens GN=BMP1
243	P12270	0.9979	Nucleoprotein TPR OS=Homo sapiens GN=TPR PE=1 SV=3

	UniProt ID	Probability	Protein Description
244	P12107-2	0.9960	Collagen alpha-1(XI) chain, isoform B
245	P11277-2	1.0000	Spectrin beta chain, erythrocytic, isoform 2
246	P11216	0.9997	Glycogen phosphorylase, brain form OS=Homo sapiens GN=PYGB PE=1 SV=5
247	P11142	1.0000	Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
248	P11137-3	0.9609	**Microtubule-associated protein 2, isoform 3**
249	P11047	1.0000	Laminin subunit gamma-1 OS=Homo sapiens GN=LAMC1 PE=1 SV=3
250	P10619	0.9722	Lysosomal protective protein OS=Homo sapiens GN=CTSA PE=1 SV=2
251	P0CB43	0.9551	Protein HGH1 homolog, isoform 2
252	P0C0L4-2	1.0000	Complement C4-A, isoform 2
253	P09960-2	0.9916	Leukotriene A-4 hydrolase, isoform 2
254	P09936	0.9999	Ubiquitin carboxyl-terminal hydrolase isozyme L1 OS=Homo sapiens GN=UCHL1 PE=1 SV=2
255	P09917	0.9988	**Arachidonate 5-lipoxygenase OS=Homo sapiens GN=ALOX5 PE=1 SV=2**
256	P09104	0.9998	Gamma-enolase OS=Homo sapiens GN=ENO2 PE=1 SV=3
257	P08621-2	0.9802	U1 small nuclear ribonucleoprotein 70 kDa, isoform 2
258	P08603-2	1.0000	Complement factor H, isoform 2
259	P08572	0.9999	Collagen alpha-2(IV) chain OS=Homo sapiens GN=COL4A2 PE=1 SV=4
260	P08514-2	0.9976	Integrin alpha-IIb, isoform 2
261	P08311	1.0000	Cathepsin G OS=Homo sapiens GN=CTSG PE=1 SV=2
262	P08246	1.0000	Neutrophil elastase OS=Homo sapiens GN=ELANE PE=1 SV=1
263	P08133-2	1.0000	Annexin A6, isoform 2
264	P07998	0.9527	Ribonuclease pancreatic OS=Homo sapiens GN=RNASE1 PE=1 SV=4
265	P07332-2	0.9819	Tyrosine-protein kinase Fes/Fps, isoform 2
266	P06753-2	1.0000	Tropomyosin alpha-3 chain
267	P06312	0.9998	Ig kappa chain V-IV region
268	P06310	1.0000	Ig kappa chain V-II region RPMI 6410 OS=Homo sapiens PE=4 SV=1
269	P06309	0.9999	Ig kappa chain V-II region GM607 (Fragment) OS=Homo sapiens PE=4 SV=1
270	P05997	0.9684	Collagen alpha-2(V) chain OS=Homo sapiens GN=COL5A2 PE=1 SV=3
271	P05186-3	1.0000	Alkaline phosphatase, tissue-nonspecific isozyme, isoform 3
272	P05164-3	1.0000	Myeloperoxidase, isoform H7
273	P05129	0.9983	Protein kinase C gamma type OS=Homo sapiens GN=PRKCG PE=1 SV=3
274	P05109	1.0000	Protein S100-A8 OS=Homo sapiens GN=S100A8 PE=1 SV=1
275	P04628	0.9905	Proto-oncogene Wnt-1 OS=Homo sapiens GN=WNT1 PE=1 SV=1
276	P04217-2	0.9999	Alpha-1B-glycoprotein, isoform 2
277	P02788-2	0.9999	Lactotransferrin, isoform DeltaLf
278	P02730	1.0000	Band 3 anion transport protein OS=Homo sapiens GN=SLC4A1 PE=1 SV=3
279	P02549-2	1.0000	Spectrin alpha chain, erythrocytic 1, isoform 2
280	P02458-1	0.9995	Collagen alpha-1(II) chain, isoform 1
281	P02100	1.0000	Hemoglobin subunit epsilon OS=Homo sapiens GN=HBE1 PE=1 SV=2
282	P02042	1.0000	Hemoglobin subunit delta OS=Homo sapiens GN=HBD PE=1 SV=2
283	P01877	0.9990	Ig alpha-2 chain C region OS=Homo sapiens GN=IGHA2 PE=1 SV=3
284	P01780	0.9942	Ig heavy chain V-III region JON OS=Homo sapiens PE=1 SV=1
285	P01775	0.9947	Ig heavy chain V-III region LAY OS=Homo sapiens PE=1 SV=1
286	P01764	0.9942	Ig heavy chain V-III region VH26 OS=Homo sapiens PE=1 SV=1
287	P01700	0.9947	Ig lambda chain V-I region HA OS=Homo sapiens PE=1 SV=1
288	P01699	0.9947	Ig lambda chain V-I region VOR, isoform 2
289	P01611	1.0000	Ig kappa chain V-I region Wes OS=Homo sapiens PE=1 SV=1
290	P01596	0.9999	Ig kappa chain V-I region CAR, isoform 2
291	P01591	0.9942	Immunoglobulin J chain OS=Homo sapiens GN=IGJ PE=1 SV=4
292	P01009-2	0.9993	Alpha-1-antitrypsin, isoform 2
293	P00568	0.9947	Adenylate kinase isoenzyme 1 OS=Homo sapiens GN=AK1 PE=1 SV=3
294	P00488	0.9932	Coagulation factor XIII A chain OS=Homo sapiens GN=F13A1 PE=1 SV=4
295	O95996-2	0.9969	Adenomatous polyposis coli protein 2, isoform 2
296	O95714	0.9872	E3 ubiquitin-protein ligase HERC2 OS=Homo sapiens GN=HERC2 PE=1 SV=2
297	O95602	0.9676	DNA-directed RNA polymerase I subunit RPA1 OS=Homo sapiens GN=POLR1A PE=1 SV=2
298	O95072-2	0.9954	Meiotic recombination protein REC8 homolog, isoform 2
299	O95071	0.9888	E3 ubiquitin-protein ligase UBR5 OS=Homo sapiens GN=UBR5 PE=1 SV=2
300	O94915	0.9823	Protein furry homolog-like OS=Homo sapiens GN=FRYL PE=1 SV=2
301	O94913	0.9569	Pre-mRNA cleavage complex 2 protein Pcf11 OS=Homo sapiens GN=PCF11 PE=1 SV=3
302	O76003	0.9999	Glutaredoxin-3 OS=Homo sapiens GN=GLRX3 PE=1 SV=2
303	O75970-2	1.0000	Multiple PDZ domain protein, isoform 2
304	O75821	0.9947	Eukaryotic translation initiation factor 3 subunit G OS=Homo sapiens GN=EIF3G PE=1 SV=2
305	O75718	0.9666	Cartilage-associated protein OS=Homo sapiens GN=CRTAP PE=1 SV=1
306	O75427	0.9997	Leucine-rich repeat and calponin homology domain-containing protein 4 OS=Homo sapiens GN=LRCH4 PE=1 SV=2

	UniProt ID	Probability	Protein Description
307	O75366-2	0.9513	Advillin, isoform 2
308	O75145-2	0.9999	Liprin-alpha-3, isoform 2
309	O60884	0.9916	DnaJ homolog subfamily A member 2 OS=Homo sapiens GN=DNAJA2 PE=1 SV=1
310	O60522-2	0.9750	Tudor domain-containing protein 6, isoform 2
311	O60506-2	0.9544	Heterogeneous nuclear ribonucleoprotein Q, isoform 2
312	O60240	0.9961	Perilipin-1 OS=Homo sapiens GN=PLIN1 PE=1 SV=2
313	O43861-2	0.9585	Probable phospholipid-transporting ATPase IIB, isoform 2
314	O43681	0.9958	ATPase ASNA1 OS=Homo sapiens GN=ASNA1 PE=1 SV=2
315	O43525	0.9586	Potassium voltage-gated channel subfamily KQT member 3 OS=Homo sapiens GN=KCNQ3 PE=1 SV=2
316	O43312	0.9692	Metastasis suppressor protein 1 OS=Homo sapiens GN=MTSS1 PE=1 SV=2
317	O15439-2	0.9905	Multidrug resistance-associated protein 4, isoform 2
318	O15417	0.9859	Trinucleotide repeat-containing gene 18 protein OS=Homo sapiens GN=TNRC18 PE=1 SV=3
319	O15354	0.9947	Probable G-protein coupled receptor 37 OS=Homo sapiens GN=GPR37 PE=1 SV=2
320	O15260-2	0.9726	Surfeit locus protein 4, isoform 2
321	O15049	0.9968	NEDD4-binding protein 3 OS=Homo sapiens GN=N4BP3 PE=1 SV=3
322	O14939-2	0.9975	Isoform PLD2B of Phospholipase D2 OS=Homo sapiens GN=PLD2
323	O14732-2	0.9792	Inositol monophosphatase 2, isoform 2
324	O00410	0.9849	Importin-5 OS=Homo sapiens GN=IPO5 PE=1 SV=4
325	O00338-2	0.9581	Sulfotransferase 1C2, isoform long
326	O00186	0.9883	Syntaxin-binding protein 3 OS=Homo sapiens GN=STXBP3 PE=1 SV=2
327	H7BZ55	0.9911	Putative ciliary rootlet coiled-coil protein-like 3 protein OS=Homo sapiens PE=5 SV=2
328	B2RTY4-2	0.9912	Unconventional myosin-Ixa, isoform 2
329	A8MTW9	0.9990	Putative uncharacterized protein ENSP00000380674 OS=Homo sapiens PE=5 SV=2
330	A7MCY6	0.9534	TANK-binding kinase 1-binding protein 1 OS=Homo sapiens GN=TBKBP1 PE=1 SV=1
331	A6NMB1	0.9926	Sialic acid-binding Ig-like lectin 16 OS=Homo sapiens GN=SIGLEC16 PE=2 SV=3
332	A6NKG5	0.9675	Retrotransposon-like protein 1 OS=Homo sapiens GN=RTL1 PE=2 SV=3
333	A6H8Y1-6	0.9648	Transcription factor TFIIIB component B'' homolog, isoform 6
334	A1KZ92-2	0.9837	Peroxidasin-like protein, isoform 2

Table 5. Odontoblast proteins found in both age groups

Proteins found in all odontoblast cell layer samples (N = 3, n = 10). (Data also shown in Figure 4D).

	UniProt ID	Probability	Protein Description
1	Q9Y6G5	0.9993	COMM domain-containing protein 10 OS=Homo sapiens GN=COMMD10 PE=1 SV=1
2	Q9Y5Z4	0.9918	Heme-binding protein 2 OS=Homo sapiens GN=HEBP2 PE=1 SV=1
3	Q9Y5X3	0.9918	Sorting nexin-5 OS=Homo sapiens GN=SNX5 PE=1 SV=1
4	Q9Y490	1.0000	Talin-1 OS=Homo sapiens GN=TLN1 PE=1 SV=3
5	Q9Y285	0.9918	Phenylalanine--tRNA ligase alpha subunit OS=Homo sapiens GN=FARSA PE=1 SV=3
6	Q9NRN5	1.0000	Olfactomedin-like protein 3 OS=Homo sapiens GN=OLFML3 PE=2 SV=1
7	Q9NPH2-2	0.9918	Inositol-3-phosphate synthase 1, isoform 2
8	Q99497	1.0000	Protein DJ-1 OS=Homo sapiens GN=PARK7 PE=1 SV=2
9	Q96IU4	0.9918	Alpha/beta hydrolase domain-containing protein 14B OS=Homo sapiens GN=ABHD14B PE=1 SV=1
10	Q96CX2	1.0000	BTB/POZ domain-containing protein KCTD12 OS=Homo sapiens GN=KCTD12 PE=1 SV=1
11	Q92530	0.9918	Proteasome inhibitor PI31 subunit OS=Homo sapiens GN=PSMF1 PE=1 SV=2
12	Q8N8S7-2	0.9852	Protein enabled homolog, isoform 2
13	Q8N1G4	0.9918	Leucine-rich repeat-containing protein 47 OS=Homo sapiens GN=LRRC47 PE=1 SV=1
14	Q86U42-2	0.9918	Polyadenylate-binding protein 2, isoform 2
15	Q86SQ4-2	0.9798	G-protein coupled receptor 126, isoform 2
16	Q16643	0.9918	Drebrin OS=Homo sapiens GN=DBN1 PE=1 SV=4
17	Q16181	1.0000	Septin-7 OS=Homo sapiens GN=SEPT7 PE=1 SV=2
18	Q15691	0.9918	Microtubule-associated protein RP/EB family member 1 OS=Homo sapiens GN=MAPRE1 PE=1 SV=3
19	Q15582	1.0000	Transforming growth factor-beta-induced protein ig-h3 OS=Homo sapiens GN=TGFBI PE=1 SV=1
20	Q15149-2	1.0000	Plectin, isoform 2
21	Q15113	1.0000	Procollagen C-endopeptidase enhancer 1 OS=Homo sapiens GN=PCOLCE PE=1 SV=2
22	Q15084-2	1.0000	Protein disulfide-isomerase A6, isoform 2
23	Q15056-2	0.9918	Eukaryotic translation initiation factor 4H, isoform short
24	Q14624-2	1.0000	Inter-alpha-trypsin inhibitor heavy chain H4, isoform 2

	UniProt ID	Probability	Protein Description
25	Q13813-2	1.0000	Spectrin alpha chain, non-erythrocytic 1, isoform 2
26	Q13561-2	1.0000	Dynactin subunit 2, isoform 2
27	Q09666	1.0000	Neuroblast differentiation-associated protein AHNAK OS=Homo sapiens GN=AHNAK PE=1 SV=2
28	Q07954	0.9870	Prolow-density lipoprotein receptor-related protein 1 OS=Homo sapiens GN=LRP1 PE=1 SV=2
29	Q07507	1.0000	Dermatopontin OS=Homo sapiens GN=DPT PE=2 SV=2
30	Q07021	0.9918	Complement component 1 Q subcomponent-binding protein, mitochondrial OS=Homo sapiens GN=C1QBP PE=1 SV=1
31	Q06830	1.0000	Peroxiredoxin-1 OS=Homo sapiens GN=PRDX1 PE=1 SV=1
32	Q06828	1.0000	Fibromodulin OS=Homo sapiens GN=FMOD PE=1 SV=2
33	Q05682-2	0.9918	Caldesmon, isoform 2
34	Q04446	1.0000	1,4-alpha-glucan-branching enzyme OS=Homo sapiens GN=GBE1 PE=1 SV=3
35	Q03252	1.0000	Lamin-B2 OS=Homo sapiens GN=LMNB2 PE=1 SV=3
36	Q02952-2	1.0000	A-kinase anchor protein 12, isoform 2
37	Q01518-2	1.0000	Adenylyl cyclase-associated protein 1, isoform 2
38	Q01469	0.9918	Fatty acid-binding protein, epidermal OS=Homo sapiens GN=FABP5 PE=1 SV=3
39	Q00839-2	1.0000	Heterogeneous nuclear ribonucleoprotein U, isoform short
40	Q00610-2	1.0000	Clathrin heavy chain 1, isoform 2
41	P84243	1.0000	Histone H3.3 OS=Homo sapiens GN=H3F3A PE=1 SV=2
42	P80748	1.0000	Ig lambda chain V-III region LOI OS=Homo sapiens PE=1 SV=1
43	P78371	1.0000	T-complex protein 1 subunit beta OS=Homo sapiens GN=CCT2 PE=1 SV=4
44	P69905	1.0000	Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
45	P68871	1.0000	Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2
46	P68402-2	0.9918	Platelet-activating factor acetylhydrolase IB subunit beta, isoform 2
47	P68104	1.0000	Elongation factor 1-alpha 1
48	P63241	1.0000	Eukaryotic translation initiation factor 5A-1
49	P63104	1.0000	14-3-3 protein zeta/delta OS=Homo sapiens GN=YWHAZ PE=1 SV=1
50	P62942	0.9918	Peptidyl-prolyl cis-trans isomerase FKBP1A OS=Homo sapiens GN=FKBP1A PE=1 SV=2
51	P62937	1.0000	Peptidyl-prolyl cis-trans isomerase A OS=Homo sapiens GN=PPIA PE=1 SV=2
52	P62805	1.0000	Histone H4 OS=Homo sapiens GN=HIST1H4A PE=1 SV=2
53	P62701	0.9999	40S ribosomal protein S4, X isoform OS=Homo sapiens GN=RPS4X PE=1 SV=2
54	P62258	1.0000	14-3-3 protein epsilon OS=Homo sapiens GN=YWHAE PE=1 SV=1
55	P62140	1.0000	Serine/threonine-protein phosphatase PP1-beta catalytic subunit OS=Homo sapiens GN=PPP1CB PE=1 SV=3
56	P61978-2	1.0000	Heterogeneous nuclear ribonucleoprotein K, isoform 2
57	P61313-2	0.9918	60S ribosomal protein L15, isoform 2
58	P61158	0.9918	Actin-related protein 3 OS=Homo sapiens GN=ACTR3 PE=1 SV=3
59	P61088	0.9830	Ubiquitin-conjugating enzyme E2 N OS=Homo sapiens GN=UBE2N PE=1 SV=1
60	P60981	1.0000	Destrin OS=Homo sapiens GN=DSTN PE=1 SV=3
61	P60660-2	1.0000	Myosin light polypeptide 6, isoform smooth muscle
62	P60174-1	1.0000	Triosephosphate isomerase, isoform 2
63	P57721-2	0.9918	Poly(rC)-binding protein 3, isoform 2
64	P55084	1.0000	Trifunctional enzyme subunit beta, mitochondrial OS=Homo sapiens GN=HADHB PE=1 SV=3
65	P55072	1.0000	Transitional endoplasmic reticulum ATPase OS=Homo sapiens GN=VCP PE=1 SV=4
66	P52565	1.0000	Rho GDP-dissociation inhibitor 1 OS=Homo sapiens GN=ARHGDIA PE=1 SV=3
67	P51884	1.0000	Lumican OS=Homo sapiens GN=LUM PE=1 SV=2
68	P51610-2	1.0000	Host cell factor 1, isoform 2
69	P50454	1.0000	Serpin H1 OS=Homo sapiens GN=SERPINH1 PE=1 SV=2
70	P49591	0.9902	Serine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=SARS PE=1 SV=3
71	P49189	1.0000	4-trimethylaminobutyraldehyde dehydrogenase OS=Homo sapiens GN=ALDH9A1 PE=1 SV=3
72	P48681	1.0000	Nestin OS=Homo sapiens GN=NES PE=1 SV=2
73	P48643	0.9902	T-complex protein 1 subunit epsilon OS=Homo sapiens GN=CCT5 PE=1 SV=1
74	P48047	0.9918	ATP synthase subunit O, mitochondrial OS=Homo sapiens GN=ATP5O PE=1 SV=1
75	P47756-2	1.0000	F-actin-capping protein subunit beta, isoform 2
76	P47755	1.0000	F-actin-capping protein subunit alpha-2 OS=Homo sapiens GN=CAPZA2 PE=1 SV=3
77	P46940	0.9918	Ras GTPase-activating-like protein IQGAP1 OS=Homo sapiens GN=IQGAP1 PE=1 SV=1
78	P46821	1.0000	Microtubule-associated protein 1B OS=Homo sapiens GN=MAP1B PE=1 SV=2
79	P46781	0.9910	40S ribosomal protein S9 OS=Homo sapiens GN=RPS9 PE=1 SV=3
80	P46108-2	1.0000	Adapter molecule crk, isoform Crk-I
81	P45974-2	0.9918	Ubiquitin carboxyl-terminal hydrolase 5, isoform short
82	P43686-2	1.0000	26S protease regulatory subunit 6B, isoform 2
83	P40926	0.9918	Malate dehydrogenase, mitochondrial OS=Homo sapiens GN=MDH2 PE=1 SV=3
84	P40429	0.9918	60S ribosomal protein L13a OS=Homo sapiens GN=RPL13A PE=1 SV=2
85	P40227-2	0.9918	T-complex protein 1 subunit zeta, isoform 2
86	P40121-2	1.0000	Macrophage-capping protein, isoform 2
87	P38919	0.9998	Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4

	UniProt ID	Probability	Protein Description
88	P36955	1.0000	Pigment epithelium-derived factor OS=Homo sapiens GN=SERPINF1 PE=1 SV=4
89	P35579	1.0000	Myosin-9 OS=Homo sapiens GN=MYH9 PE=1 SV=4
90	P32119	1.0000	Peroxiredoxin-2 OS=Homo sapiens GN=PRDX2 PE=1 SV=5
91	P31930	0.9918	Cytochrome b-c1 complex subunit 1, mitochondrial OS=Homo sapiens GN=UQCRC1 PE=1 SV=3
92	P31323	0.9989	cAMP-dependent protein kinase type II-beta regulatory subunit OS=Homo sapiens GN=PRKAR2B PE=1 SV=3
93	P31150	1.0000	Rab GDP dissociation inhibitor alpha OS=Homo sapiens GN=GDI1 PE=1 SV=2
94	P30740	0.9918	Leukocyte elastase inhibitor OS=Homo sapiens GN=SERPINB1 PE=1 SV=1
95	P30101	1.0000	Protein disulfide-isomerase A3 OS=Homo sapiens GN=PDIA3 PE=1 SV=4
96	P30049	1.0000	ATP synthase subunit delta, mitochondrial OS=Homo sapiens GN=ATP5D PE=1 SV=2
97	P30043	1.0000	Flavin reductase (NADPH) OS=Homo sapiens GN=BLVRB PE=1 SV=3
98	P30041	1.0000	Peroxiredoxin-6 OS=Homo sapiens GN=PRDX6 PE=1 SV=3
99	P29401-2	1.0000	Transketolase, isoform 2
100	P28482-2	0.9918	Mitogen-activated protein kinase 1, isoform 2
101	P27824	1.0000	Calnexin OS=Homo sapiens GN=CANX PE=1 SV=2
102	P27708	0.9918	CAD protein OS=Homo sapiens GN=CAD PE=1 SV=3
103	P27361-2	0.9918	Mitogen-activated protein kinase 3, isoform 2
104	P26641	0.9918	Elongation factor 1-gamma OS=Homo sapiens GN=EEF1G PE=1 SV=3
105	P26368-2	0.9918	Splicing factor U2AF 65 kDa subunit, isoform 2
106	P26038	1.0000	Moesin OS=Homo sapiens GN=MSN PE=1 SV=3
107	P25788-2	0.9918	Proteasome subunit alpha type-3, isoform 2
108	P25311	1.0000	Zinc-alpha-2-glycoprotein OS=Homo sapiens GN=AZGP1 PE=1 SV=2
109	P24821-4	1.0000	Tenascin, isoform 2
110	P23528	0.9918	Cofilin-1 OS=Homo sapiens GN=CFL1 PE=1 SV=3
111	P23284	0.9918	Peptidyl-prolyl cis-trans isomerase B OS=Homo sapiens GN=PPIB PE=1 SV=2
112	P23246-2	1.0000	Splicing factor, proline- and glutamine-rich, isoform short
113	P22392-2	1.0000	Nucleoside diphosphate kinase B, isoform 3
114	P22314	1.0000	Ubiquitin-like modifier-activating enzyme 1 OS=Homo sapiens GN=UBA1 PE=1 SV=3
115	P21980-2	0.9910	Protein-glutamine gamma-glutamyltransferase 2, isoform 2
116	P21810	1.0000	Biglycan OS=Homo sapiens GN=BGN PE=1 SV=2
117	P21333-2	1.0000	Filamin-A, isoform 2
118	P20774	1.0000	Mimecan OS=Homo sapiens GN=OGN PE=1 SV=1
119	P20700	1.0000	Lamin-B1 OS=Homo sapiens GN=LMNB1 PE=1 SV=2
120	P19827	1.0000	Inter-alpha-trypsin inhibitor heavy chain H1 OS=Homo sapiens GN=ITIH1 PE=1 SV=3
121	P19823	1.0000	Inter-alpha-trypsin inhibitor heavy chain H2 OS=Homo sapiens GN=ITIH2 PE=1 SV=2
122	P18669	1.0000	Phosphoglycerate mutase 1 OS=Homo sapiens GN=PGAM1 PE=1 SV=2
123	P18206-2	1.0000	Vinculin, isoform 1
124	P17655	1.0000	Calpain-2 catalytic subunit OS=Homo sapiens GN=CAPN2 PE=1 SV=6
125	P13797	1.0000	Plastin-3 OS=Homo sapiens GN=PLS3 PE=1 SV=4
126	P13639	1.0000	Elongation factor 2 OS=Homo sapiens GN=EEF2 PE=1 SV=4
127	P13611-2	1.0000	Versican core protein, isoform V1
128	P13489	1.0000	Ribonuclease inhibitor OS=Homo sapiens GN=RNH1 PE=1 SV=2
129	P12814-2	1.0000	Alpha-actinin-1, isoform2
130	P12236	0.9994	ADP/ATP translocase 3 OS=Homo sapiens GN=SLC25A6 PE=1 SV=4
131	P12111-4	1.0000	Isoform 4 of Collagen alpha-3(VI) chain OS=Homo sapiens GN=COL6A3
132	P12110	1.0000	Collagen alpha-2(VI) chain OS=Homo sapiens GN=COL6A2 PE=1 SV=4
133	P12109	1.0000	Collagen alpha-1(VI) chain OS=Homo sapiens GN=COL6A1 PE=1 SV=3
134	P11021	1.0000	78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2
135	P10909-2	1.0000	Clusterin, isoform 2
136	P10809	1.0000	60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1 PE=1 SV=2
137	P10412	1.0000	Histone H1.4 OS=Homo sapiens GN=HIST1H1E PE=1 SV=2
138	P09651-2	1.0000	Heterogeneous nuclear ribonucleoprotein A1, isoform A1-A
139	P09525	1.0000	Annexin A4 OS=Homo sapiens GN=ANXA4 PE=1 SV=4
140	P09417	1.0000	Dihydropteridine reductase OS=Homo sapiens GN=QDPR PE=1 SV=2
141	P09382	1.0000	Galectin-1 OS=Homo sapiens GN=LGALS1 PE=1 SV=2
142	P09211	1.0000	Glutathione S-transferase P OS=Homo sapiens GN=GSTP1 PE=1 SV=2
143	P08865	1.0000	40S ribosomal protein SA OS=Homo sapiens GN=RPSA PE=1 SV=4
144	P08758	1.0000	Annexin A5 OS=Homo sapiens GN=ANXA5 PE=1 SV=2
145	P08670	1.0000	Vimentin OS=Homo sapiens GN=VIM PE=1 SV=4
146	P08123	1.0000	Collagen alpha-2(I) chain OS=Homo sapiens GN=COL1A2 PE=1 SV=7
147	P07996	1.0000	Thrombospondin-1 OS=Homo sapiens GN=THBS1 PE=1 SV=2
148	P07910-2	1.0000	Heterogeneous nuclear ribonucleoproteins C1/C2, isoform C1
149	P07900-2	1.0000	Heat shock protein HSP 90-alpha, isoform 2
150	P07741-2	1.0000	Adenine phosphoribosyltransferase, isoform 2

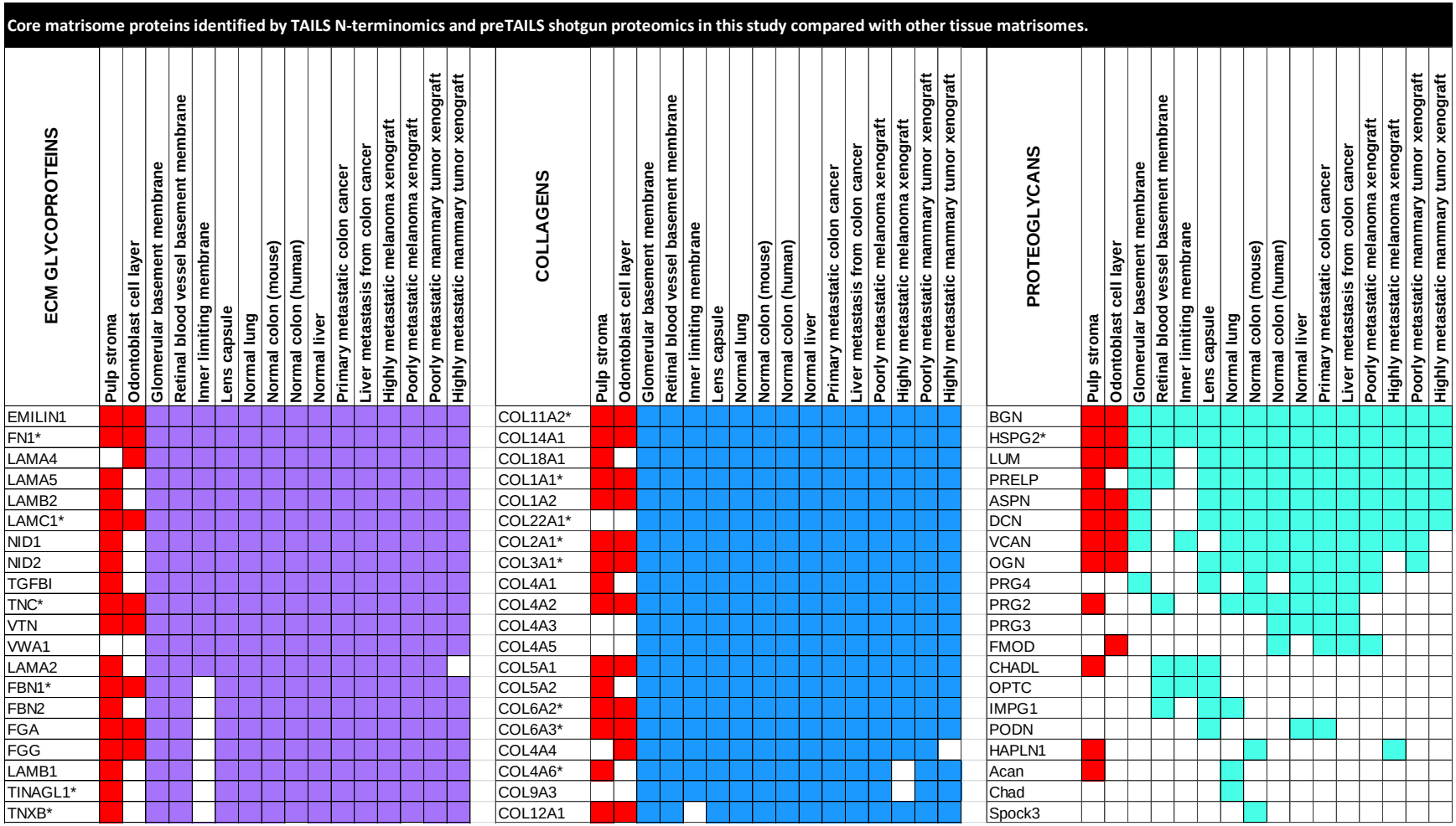
	UniProt ID	Probability	Protein Description
151	P07737	1.0000	Profilin-1 OS=Homo sapiens GN=PFN1 PE=1 SV=2
152	P07585	1.0000	Decorin OS=Homo sapiens GN=DCN PE=1 SV=1
153	P07437	1.0000	Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2
154	P07355	1.0000	Annexin A2 OS=Homo sapiens GN=ANXA2 PE=1 SV=2
155	P07339	1.0000	Cathepsin D OS=Homo sapiens GN=CTSD PE=1 SV=1
156	P07305	0.9918	Histone H1.0 OS=Homo sapiens GN=H1F0 PE=1 SV=3
157	P07237	1.0000	Protein disulfide-isomerase OS=Homo sapiens GN=P4HB PE=1 SV=3
158	P06727	1.0000	Apolipoprotein A-IV OS=Homo sapiens GN=APOA4 PE=1 SV=3
159	P06703	1.0000	Protein S100-A6 OS=Homo sapiens GN=S100A6 PE=1 SV=1
160	P06576	1.0000	ATP synthase subunit beta, mitochondrial OS=Homo sapiens GN=ATP5B PE=1 SV=3
161	P06396-3	1.0000	Gelsolin, isoform 3
162	P05155	0.9918	Plasma protease C1 inhibitor OS=Homo sapiens GN=SERPING1 PE=1 SV=2
163	P05141	0.9994	ADP/ATP translocase 2 OS=Homo sapiens GN=SLC25A5 PE=1 SV=7
164	P04908	1.0000	Histone H2A type 1-B/E
165	P04792	1.0000	Heat shock protein beta-1 OS=Homo sapiens GN=HSPB1 PE=1 SV=2
166	P04434	0.9996	Ig kappa chain V-III region VH (Fragment) OS=Homo sapiens PE=4 SV=1
167	P04433	1.0000	Ig kappa chain V-III region VG (Fragment) OS=Homo sapiens PE=1 SV=1
168	P04406	1.0000	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens GN=GAPDH PE=1 SV=3
169	P04196	1.0000	Histidine-rich glycoprotein OS=Homo sapiens GN=HRG PE=1 SV=1
170	P04083	1.0000	Annexin A1 OS=Homo sapiens GN=ANXA1 PE=1 SV=2
171	P04075-2	1.0000	Fructose-bisphosphate aldolase A, isoform 2
172	P04040	1.0000	Catalase OS=Homo sapiens GN=CAT PE=1 SV=3
173	P04004	1.0000	Vitronectin OS=Homo sapiens GN=VTN PE=1 SV=1
174	P04003	0.9918	C4b-binding protein alpha chain OS=Homo sapiens GN=C4BPA PE=1 SV=2
175	P02794	0.9918	Ferritin heavy chain OS=Homo sapiens GN=FTH1 PE=1 SV=2
176	P02790	1.0000	Hemopexin OS=Homo sapiens GN=HPX PE=1 SV=2
177	P02787	1.0000	Serotransferrin OS=Homo sapiens GN=TF PE=1 SV=3
178	P02766	1.0000	Transthyretin OS=Homo sapiens GN=TTR PE=1 SV=1
179	P02765	1.0000	Alpha-2-HS-glycoprotein OS=Homo sapiens GN=AHSG PE=1 SV=1
180	P02760	1.0000	Protein AMBP OS=Homo sapiens GN=AMBP PE=1 SV=1
181	P02751-10	1.0000	Fibronectin, isoform 10
182	P02750	0.9918	Leucine-rich alpha-2-glycoprotein OS=Homo sapiens GN=LRG1 PE=1 SV=2
183	P02749	1.0000	Beta-2-glycoprotein 1 OS=Homo sapiens GN=APOH PE=1 SV=3
184	P02743	1.0000	Serum amyloid P-component OS=Homo sapiens GN=APCS PE=1 SV=2
185	P02679-2	1.0000	Fibrinogen gamma chain, isoform Gamma-A
186	P02675	1.0000	Fibrinogen beta chain OS=Homo sapiens GN=FGB PE=1 SV=2
187	P02671-2	1.0000	Fibrinogen alpha chain, isoform 2
188	P02647	1.0000	Apolipoprotein A-I OS=Homo sapiens GN=APOA1 PE=1 SV=1
189	P02545	1.0000	Prelamin-A/C OS=Homo sapiens GN=LMNA PE=1 SV=1
190	P02511	0.9918	Alpha-crystallin B chain OS=Homo sapiens GN=CRYAB PE=1 SV=2
191	P02461-2	1.0000	Collagen alpha-1(III) chain, isoform 2
192	P02452	1.0000	Collagen alpha-1(I) chain OS=Homo sapiens GN=COL1A1 PE=1 SV=5
193	P01876	1.0000	Ig alpha-1 chain C region OS=Homo sapiens GN=IGHA1 PE=1 SV=2
194	P01871-2	1.0000	Ig mu chain C region, isoform 2
195	P01860	1.0000	Ig gamma-3 chain C region OS=Homo sapiens GN=IGHG3 PE=1 SV=2
196	P01834	1.0000	Ig kappa chain C region OS=Homo sapiens GN=IGKC PE=1 SV=1
197	P01772	1.0000	Ig heavy chain V-III region KOL OS=Homo sapiens PE=1 SV=1
198	P01766	1.0000	Ig heavy chain V-III region BRO OS=Homo sapiens PE=1 SV=1
199	P01717	1.0000	Ig lambda chain V-IV region Hil OS=Homo sapiens PE=1 SV=1
200	P01714	0.9918	Ig lambda chain V-III region SH OS=Homo sapiens PE=1 SV=1
201	P01706	0.9994	Ig lambda chain V-II region BOH
202	P01703	0.9918	Ig lambda chain V-I region NEWM OS=Homo sapiens PE=1 SV=1
203	P01042-2	1.0000	Isoform LMW of Kininogen-1 OS=Homo sapiens GN=KNG1
204	P01024	1.0000	Complement C3 OS=Homo sapiens GN=C3 PE=1 SV=2
205	P01019	1.0000	Angiotensinogen OS=Homo sapiens GN=AGT PE=1 SV=1
206	P01011	1.0000	Alpha-1-antichymotrypsin OS=Homo sapiens GN=SERPINA3 PE=1 SV=2
207	P01008	1.0000	Antithrombin-III OS=Homo sapiens GN=SERPINC1 PE=1 SV=1
208	P00915	1.0000	Carbonic anhydrase 1 OS=Homo sapiens GN=CA1 PE=1 SV=2
209	P00751	1.0000	Complement factor B OS=Homo sapiens GN=CFB PE=1 SV=2
210	P00747	1.0000	Plasminogen OS=Homo sapiens GN=PLG PE=1 SV=2
211	P00738	0.9910	Haptoglobin OS=Homo sapiens GN=HP PE=1 SV=1
212	P00734	1.0000	Prothrombin OS=Homo sapiens GN=F2 PE=1 SV=2
213	P00450	1.0000	Ceruloplasmin OS=Homo sapiens GN=CP PE=1 SV=1

	UniProt ID	Probability	Protein Description
214	P00441	1.0000	Superoxide dismutase [Cu-Zn] OS=Homo sapiens GN=SOD1 PE=1 SV=2
215	P00367	0.9918	Glutamate dehydrogenase 1, mitochondrial OS=Homo sapiens GN=GLUD1 PE=1 SV=2
216	P00338-2	0.9918	L-lactate dehydrogenase A chain, isoform 2
217	O95336	1.0000	6-phosphogluconolactonase OS=Homo sapiens GN=PGLS PE=1 SV=2
218	O94875-10	0.9918	Sorbin and SH3 domain-containing protein 2, isoform 10
219	O94844	0.9862	Rho-related BTB domain-containing protein 1 OS=Homo sapiens GN=RHOBTB1 PE=1 SV=2
220	O60814	0.9999	Histone H2B type 1-K
221	O60493-2	0.9918	Sorting nexin-3, isoform 2
222	O43396	0.9918	Thioredoxin-like protein 1 OS=Homo sapiens GN=TXNL1 PE=1 SV=3
223	O15143	0.9918	Actin-related protein 2/3 complex subunit 1B OS=Homo sapiens GN=ARPC1B PE=1 SV=3
224	O14950	1.0000	Myosin regulatory light chain 12B
225	O14773	1.0000	Tripeptidyl-peptidase 1 OS=Homo sapiens GN=TPP1 PE=1 SV=2
226	O00445	0.9717	Synaptotagmin-5 OS=Homo sapiens GN=SYT5 PE=2 SV=2
227	O00151	1.0000	PDZ and LIM domain protein 1 OS=Homo sapiens GN=PDLIM1 PE=1 SV=4
228	B9A064	1.0000	Immunoglobulin lambda-like polypeptide 5

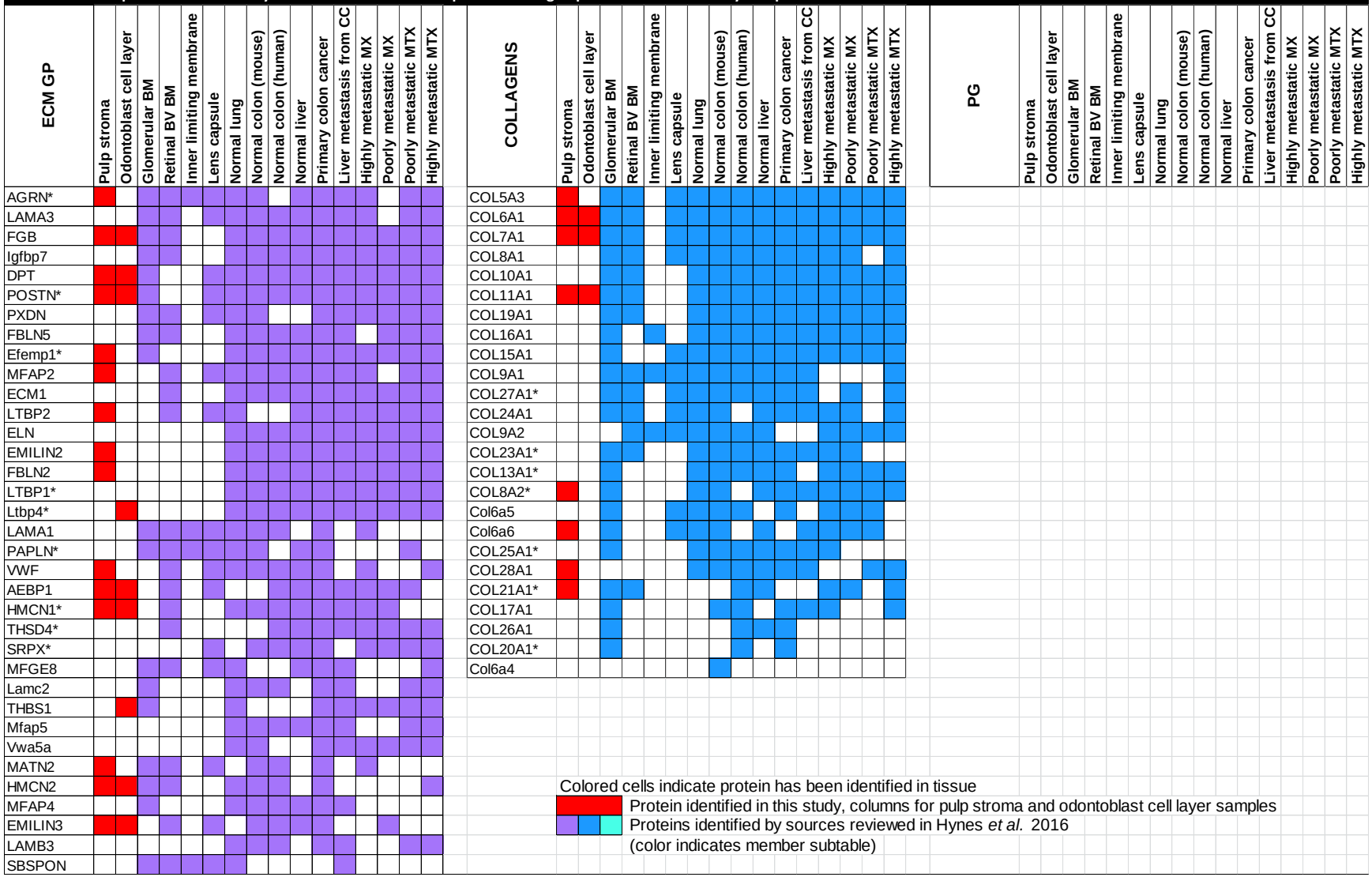
Tables 3-5: All proteins identified in the odontoblast cell layer sample are grouped according to the age of the donor tissue, (**3**) ≤ 20 years for young pulp, (**4**) > 20 years for mature pulp, (**5**) Proteins found in both young and mature pulp tissues. Proteins are listed with their Uniprot ID, protein probability from TPP analysis of PSM data, and full protein name.

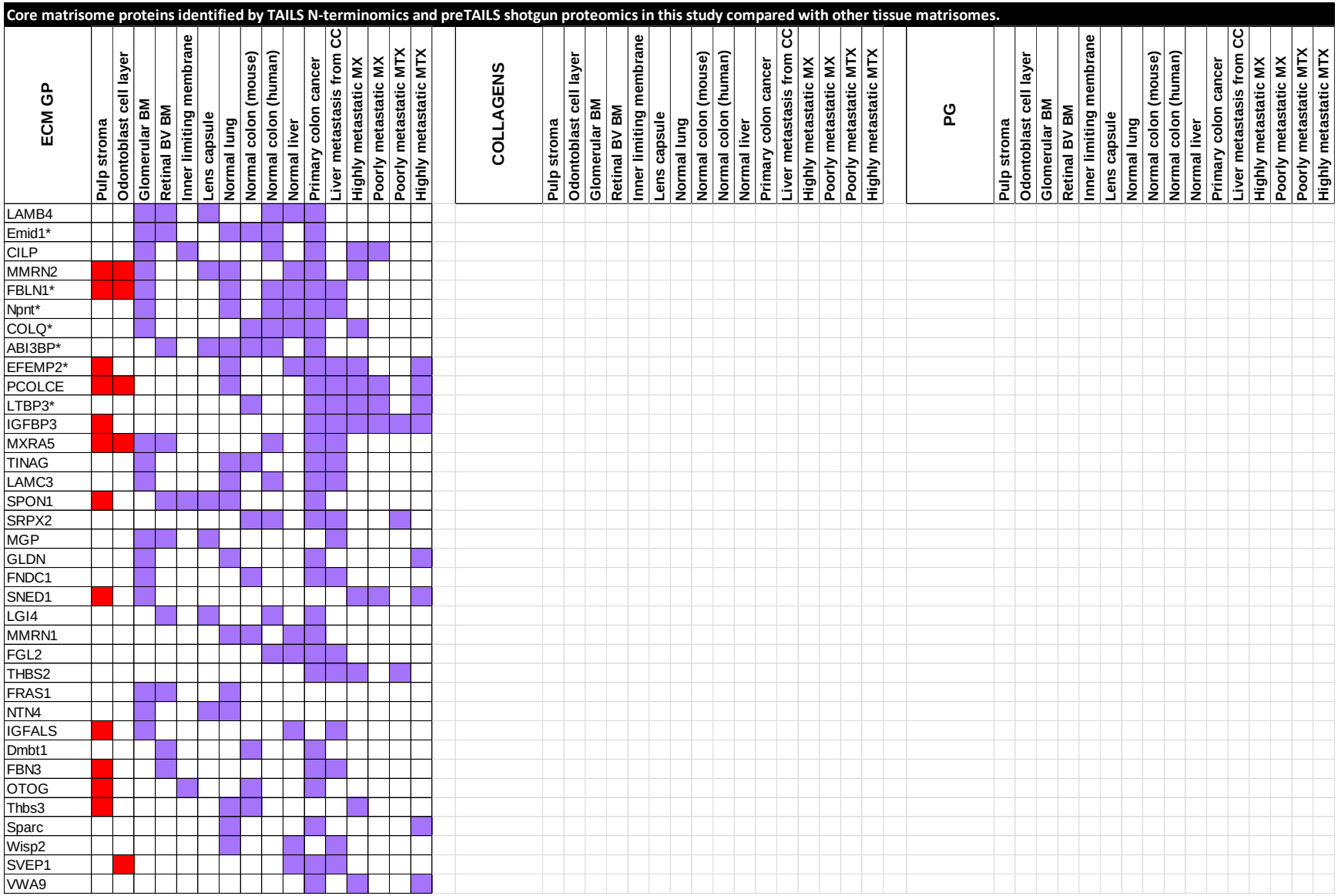
Appendix E Matrisome comparison

Table 6. Core matrisome proteins



Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.




Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.
ECM GP
Pulp stroma
Odontoblast cell layer
Glomerular BM
Retinal BV BM
Inner limiting membrane
Lens capsule
Normal lung
Normal colon (mouse)
Normal colon (human)
Normal liver
Primary colon cancer
Liver metastasis from CC
Highly metastatic MX
Poorly metastatic MX
Poorly metastatic MTX
Highly metastatic MTX
LAMB4
Emid1*
CILP
MMRN2
FBLN1*
Npnt*
COLQ*
ABI3BP*
EFEMP2*
PCOLCE
LTBP3*
IGFBP3
MXRA5
TINAG
LAMC3
SPON1
SRPX2
MGP
GLDN
FNDC1
SNED1
LGI4
MMRN1
FGL2
THBS2
FRAS1
NTN4
IGFALS
Dmbt1
FBN3
OTOG
Thbs3
Sparc
Wisp2
SVEP1
VWA9
COLLAGENS
PG

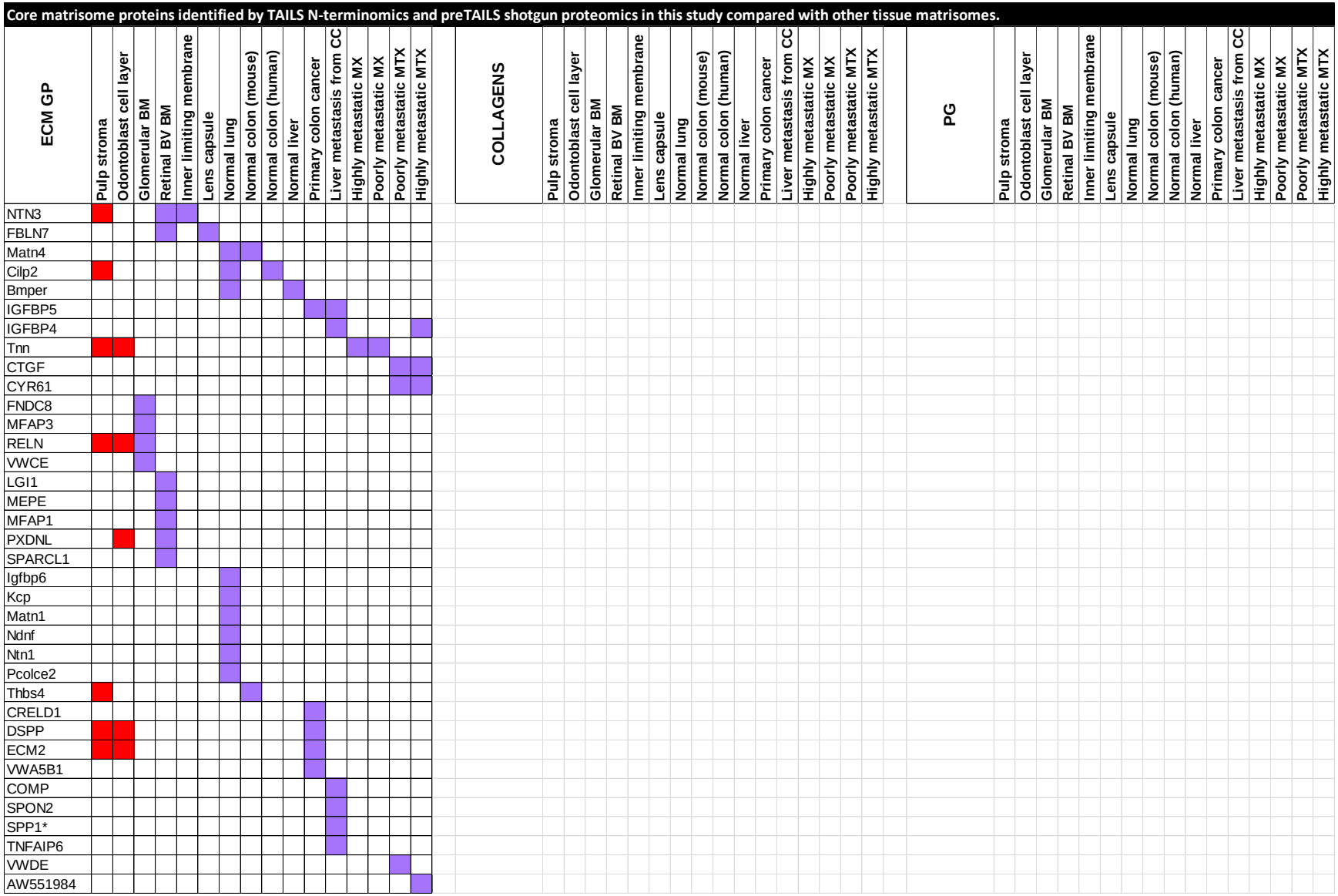
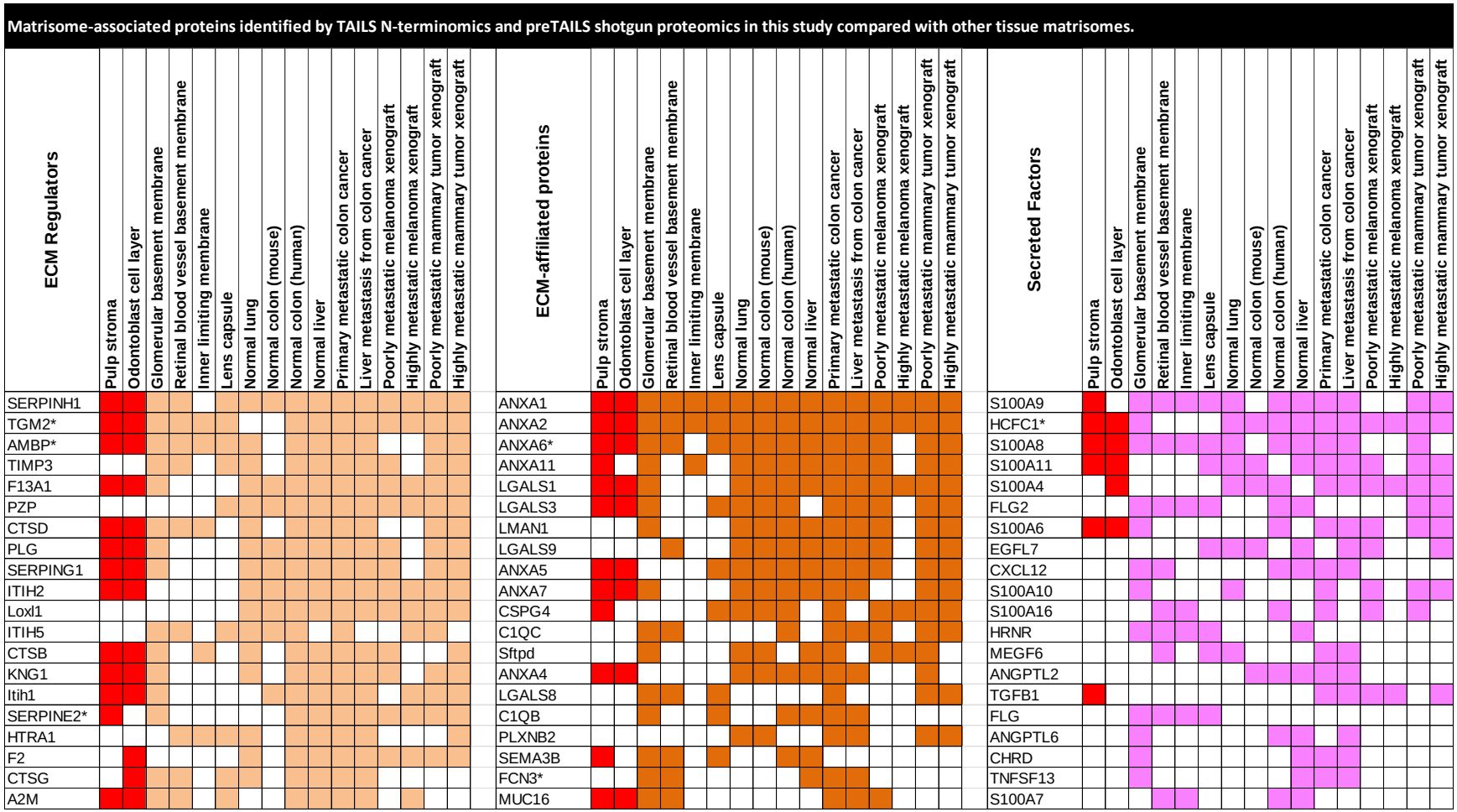
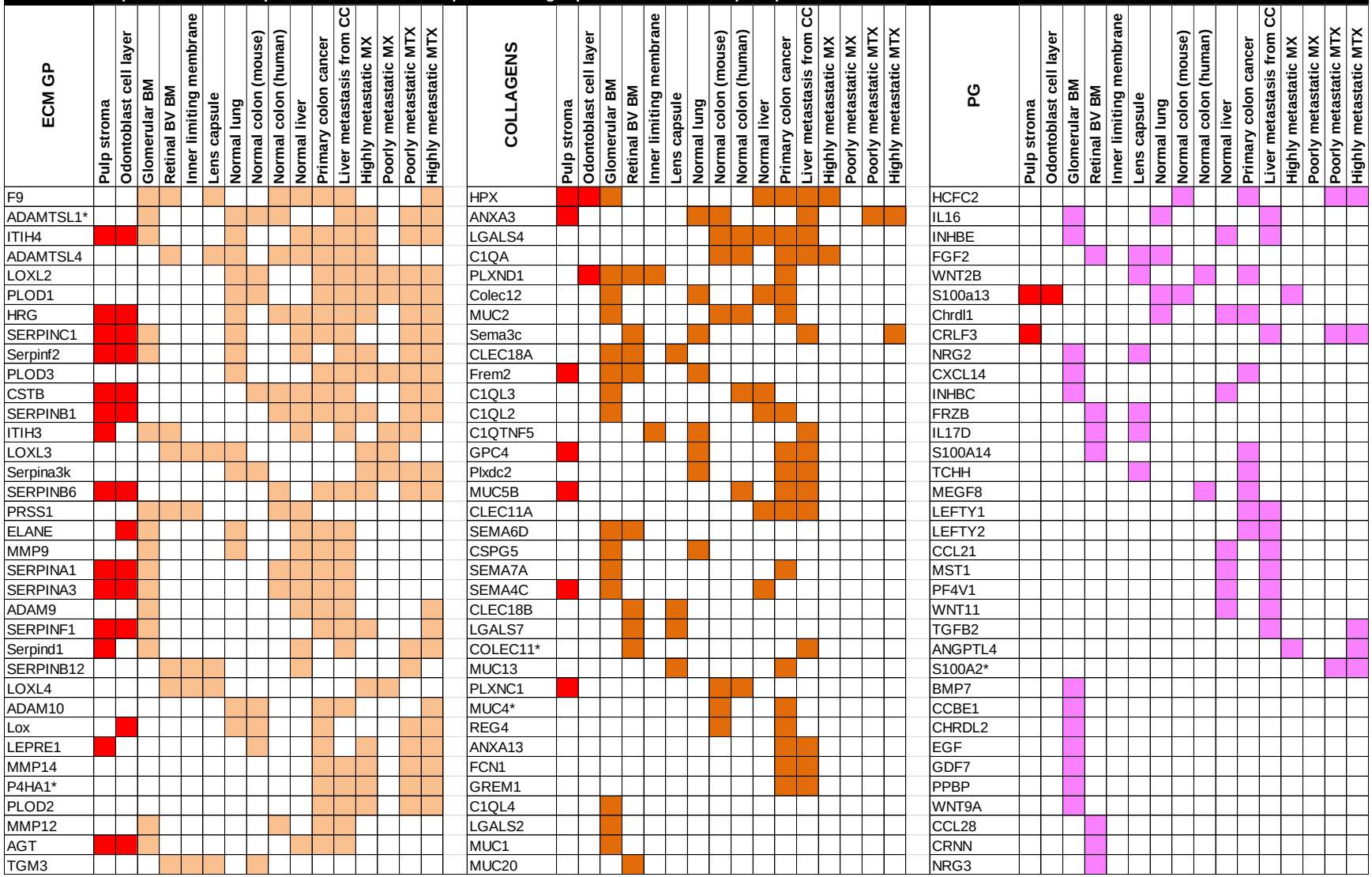

Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.
ECM GP
Pulp stroma
Odontoblast cell layer
Glomerular BM
Retinal BV BM
Inner limiting membrane
Lens capsule
Normal lung
Normal colon (mouse)
Normal colon (human)
Normal liver
Primary colon cancer
Liver metastasis from CC
Highly metastatic MX
Poorly metastatic MX
Poorly metastatic MTX
Highly metastatic MTX
NTN3
FBLN7
Matn4
Cilp2
Bmper
IGFBP5
IGFBP4
Tnn
CTGF
CYR61
FNDC8
MFAP3
RELN
VWCE
LGI1
MEPE
MFAP1
PXDNL
SPARCL1
Igfbp6
Kcp
Matn1
Ndnf
Ntn1
Pcolce2
Thbs4
CRELD1
DSPP
ECM2
VWA5B1
COMP
SPON2
SPP1*
TNFAIP6
VWDE
AW551984
COLLAGENS
PG

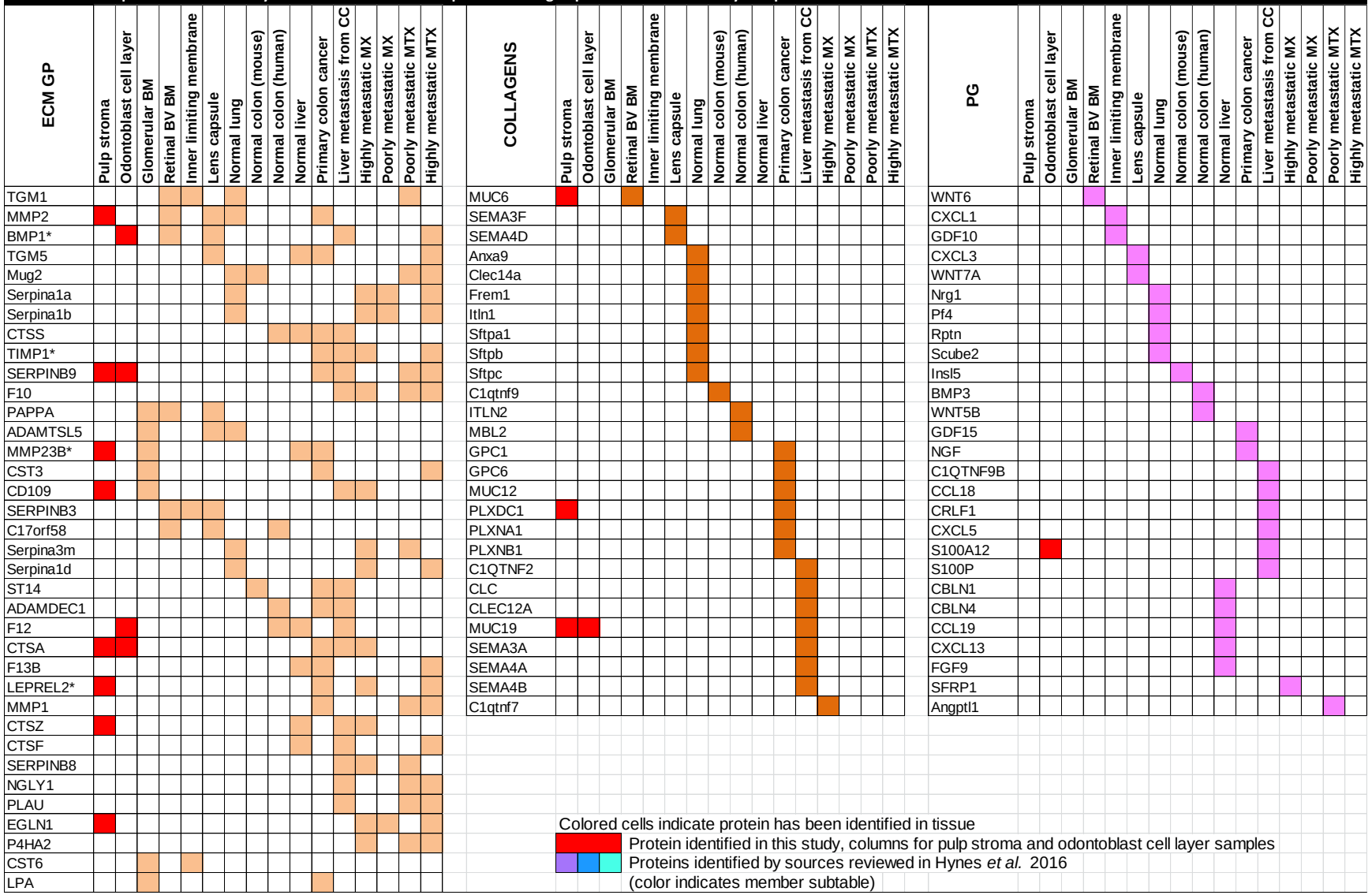
Table 7. Matrisome-associated proteins

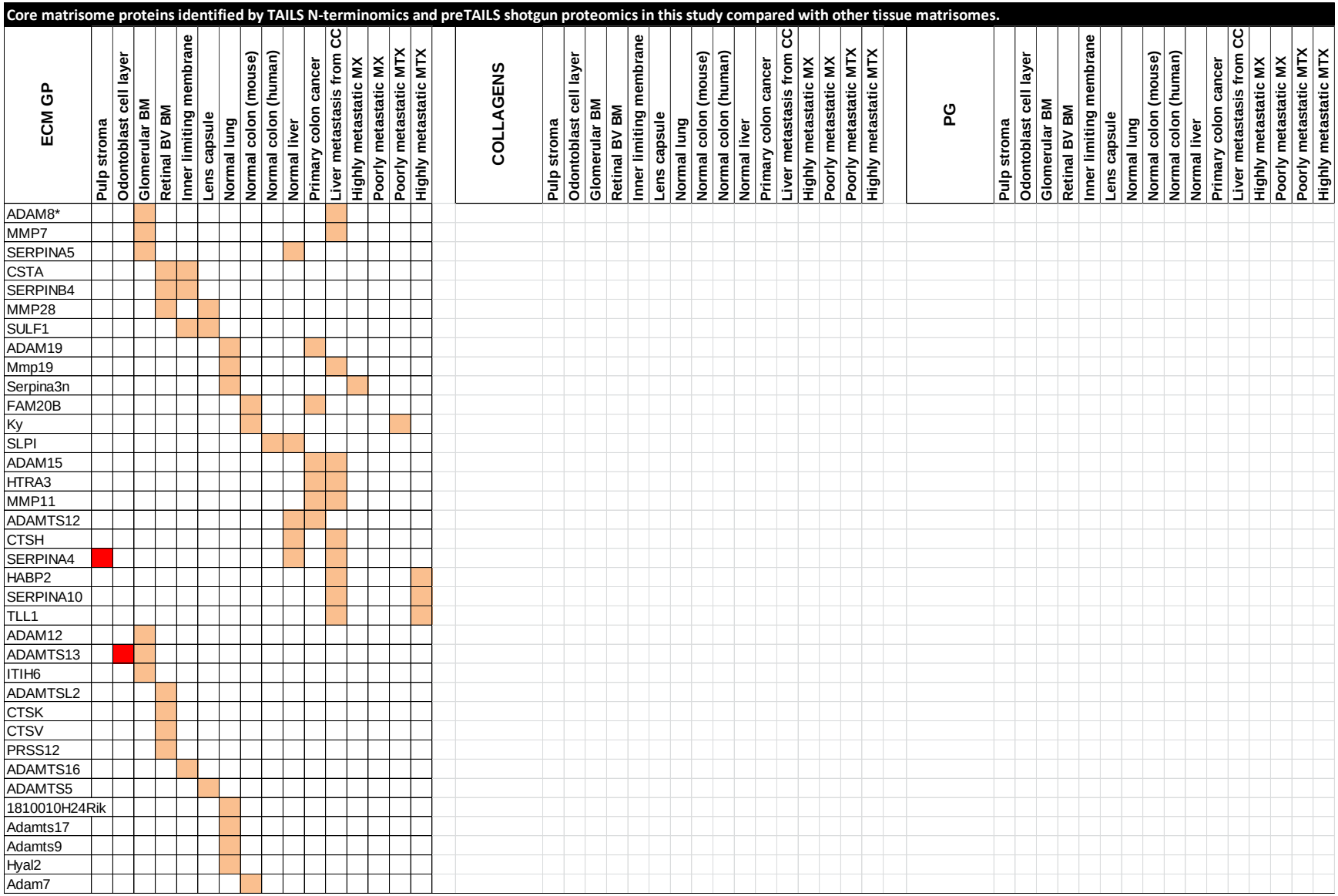


Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.



Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.




Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.
ECM GP
Pulp stroma
Odontoblast cell layer
Glomerular BM
Retinal BV BM
Inner limiting membrane
Lens capsule
Normal lung
Normal colon (mouse)
Normal colon (human)
Normal liver
Primary colon cancer
Liver metastasis from CC
Highly metastatic MX
Poorly metastatic MX
Poorly metastatic MTX
Highly metastatic MTX
ADAM8*
MMP7
SERPINA5
CSTA
SERPINB4
MMP28
SULF1
ADAM19
Mmp19
Serpina3n
FAM20B
Ky
SLPI
ADAM15
HTRA3
MMP11
ADAMTS12
CTSH
SERPINA4
HABP2
SERPINA10
TLL1
ADAM12
ADAMTS13
ITIH6
ADAMTSL2
CTSK
CTSV
PRSS12
ADAMTS16
ADAMTS5
1810010H24Rik
Adamts17
Adamts9
Hyal2
Adam7
COLLAGENS
PG

Core matrisome proteins identified by TAILS N-terminomics and preTAILS shotgun proteomics in this study compared with other tissue matrisomes.

ECM GP	Pulp stroma	Odontoblast cell layer	Glomerular BM	Retinal BV BM	Inner limiting membrane	Lens capsule	Normal lung	Normal colon (mouse)	Normal colon (human)	Normal liver	Primary colon cancer	Liver metastasis from CC	Highly metastatic MX	Poorly metastatic MX	Poorly metastatic MTX	Highly metastatic MTX
Leprel1																
Mep1a																
Serpinb1a																
FAM20C																
HPSE2																
ADAMTS10																
HPSE																
MMP10																
MMP3																
SERPINB7																
SERPINB10																
SERPINB5																
TLL2																
CTSL																
F7																
Adam23																
SERPINE1																
Try10																
PLAT																
ADAM17																
AI182371																
CTSC																
P4HTM																

COLLAGENS	Pulp stroma	Odontoblast cell layer	Glomerular BM	Retinal BV BM	Inner limiting membrane	Lens capsule	Normal lung	Normal colon (mouse)	Normal colon (human)	Normal liver	Primary colon cancer	Liver metastasis from CC	Highly metastatic MX	Poorly metastatic MX	Poorly metastatic MTX	Highly metastatic MTX

PG	Pulp stroma	Odontoblast cell layer	Glomerular BM	Retinal BV BM	Inner limiting membrane	Lens capsule	Normal lung	Normal colon (mouse)	Normal colon (human)	Normal liver	Primary colon cancer	Liver metastasis from CC	Highly metastatic MX	Poorly metastatic MX	Poorly metastatic MTX	Highly metastatic MTX

Tables 6 & 7. ECM core matrisome and matrisome-associated proteins identified in dental pulp and the odontoblast cell layer and other tissues according to Hynes *et al.*, 2016. Colors are used only to clarify subgroupings, *e.g.* one per column, with red indicating a protein identified in our pulp and/or odontoblast cell layer samples.