This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Chrono-proteomics of human saliva: variations of the salivary proteome during human development.

Journal:	*Journal of Proteome Research*
Manuscript ID:	pr-2014-01270x.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Messana, Irene; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Cabras, Tiziana; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Iavarone, Federica; Facoltà di Medicina - Università Cattolica, Ist. di Biochimica e Biochimica Clinica Manconi, Barbara; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Huang, Liling; Facoltà di Medicina - Università Cattolica, Ist. di Biochimica e Biochimica Clinica Martelli, Claudia; Facoltà di Medicina - Università Cattolica, Ist. di Biochimica e Biochimica Clinica Olianas, Alessandra; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Sanna, Maria; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Pisano, Elisabetta; University of Cagliari, Dipartimento di Scienze Chirurgiche Sanna, Monica; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Arba, Morena; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente D'Alessandro, Alfredo; University of Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente Desiderio, Claudia; CNR, Istituto di Chimica del Riconoscimento Molecolare Vitali, Alberto; CNR, Istituto di Chimica del Riconoscimento Molecolare Pirolli, Davide; Facoltà di Medicina - Università Cattolica, Ist. di Biochimica e Biochimica Clinica Tirone, Chiara; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Pediatrica Lio, Alessandra; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Pediatrica Vento, Giovanni; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Pediatrica Romagnoli, Costantino; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Pediatrica Cordaro, Massimo; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Odontostomatologica Manni, Armando; Facoltà di Medicina - Università Cattolica, Ist. di Clinica

	Odontostomatologica Gallenzi, Patrizia; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Odontostomatologica Fiorita, Antonella; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Otorinolaringoiatrica Scarano, Emanuele; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Otorinolaringoiatrica Calò, Lea; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Otorinolaringoiatrica Passali, Giulio; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Otorinolaringoiatrica Picciotti, Pasqualina; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Otorinolaringoiatrica Paludetti, Gaetano; Facoltà di Medicina - Università Cattolica, Ist. di Clinica Otorinolaringoiatrica Fanos, Vassilios; University of Cagliari, Dipartimento di Scienze Chirurgiche Faa, Gavino; University of Cagliari, Dipartimento di Scienze Chirurgiche Castagnola, Massimo; Facoltà di Medicina - Università Cattolica, Ist. di Biochimica e Biochimica Clinica

SCHOLARONE™
Manuscripts

Chrono-proteomics of human saliva: variations of the salivary proteome during human development.

Irene Messana[a], Tiziana Cabras[a], Federica Iavarone[b], Barbara Manconi[a], Liling Huang[b], Claudia Martelli[b], Alessandra Olianas[a], Maria Teresa Sanna[a], Elisabetta Pisano[c], Monica Sanna[a], Morena Arba[a], Alfredo D'Alessandro[a], Claudia Desiderio[d], Alberto Vitali[d], Davide Pirolli[b], Chiara Tirone[e], Alessandra Lio[e], Giovanni Vento[e], Costantino Romagnoli[e], Massimo Cordaro[f], Armando Manni[f], Patrizia Gallenzi[f], Antonella Fiorita[g], Emanuele Scarano[g], Lea Calò[g], Giulio Cesare Passali[g], Pasqualina Maria Picciotti[g], Gaetano Paludetti[g], Vassilios Fanos[c], Gavino Faa[c], Massimo Castagnola[b,d*].

[a]Dip. di Scienze della Vita e dell'Ambiente, Università di Cagliari, Cagliari, Italy, [b]Ist. di Biochimica e Biochimica Clinica, Università Cattolica, Roma, Italy, [c]Dipartimento di Scienze Chirurgiche, Università di Cagliari, Cagliari Italy, [d]Istituto di Chimica del Riconoscimento Molecolare, CNR., Roma, Italy, [e]Ist. di Clinica Pediatrica, [f]Ist. di Clinica Odontostomatologica, [g]Ist. di Clinica Otorinolaringoiatrica, Università Cattolica, Roma, Italy.

* Corresponding author

Prof. Massimo Castagnola

Istituto di Biochimica e Biochimica Clinica

Facoltà di Medicina, Università Cattolica

Largo F. Vito, 00168, Roma, Italy

Tel. and/or Fax: ++39-06-3053598

E-mail: mcastagnola@rm.unicatt.it or massimo.castagnola@icrm.cnr.it

Abstract

An important contribution to the variability of any proteome is given by the time dimension that should be carefully considered to define physiological modifications. To this purpose whole saliva proteome was investigated in a wide age range. Whole saliva was collected from 17 preterm newborns with a postconceptional age at birth of 178−217 days. In these subjects sample collection was performed serially starting immediately after birth and within about 1 year follow-up gathering a total of 111 specimens. Furthermore, whole saliva was collected from 182 subjects aged between 0 and 17 years and from 23 adults aged between 27 and 57 years. The naturally occurring intact salivary proteome of the 316 samples was analyzed by low- and high-resolution HPLC-ESI-MS platforms. Proteins peculiar of the adults appeared in saliva with different time courses during human development. Acidic proline-rich proteins encoded by *PRH2* locus and glycosylated basic proline-rich proteins encoded by *PRB3* locus appeared since 180 days of post-conceptional age, followed at seven months (± two weeks) by histatin 1, statherin and P-B peptide. The other histatins and acidic proline-rich proteins encoded by *PRH1* locus appeared in whole saliva of babies from one to three weeks after the normal term of delivery, S-type cystatins appeared at one year (± three months) and basic proline-rich proteins at four years (± one year) of age. All the proteinases involved in the maturation of salivary proteins were more active in preterm than in at-term newborns, on the basis of the truncated forms detected. The activity of the Fam20C kinase, involved in the phosphorylation of various proteins, started around 180 days of post-conceptional age, slowly increased reaching values comparable to adults at about two years (± 6 months) of age. Instead, MAPK14 involved in the phosphorylation of S100A9 was fully active since birth also in preterm newborns.

Keywords: chrono-proteomics, saliva, human, proline-rich proteins, statherin, histatin, S-type cystatins, S100A9 protein, preterm newborns.

1. Introduction

Human saliva is a bodily fluid very attractive for diagnostic and prognostic purposes because it is characterized by an easy and non-invasive specimen collection.[1] For this reason in the last years many groups have studied by a large variety of –omic platforms human saliva under physiological and pathological conditions to evidence new, and possibly early, salivary biomarkers of disease with clinical utility as reported in recent outstanding reviews and articles.[2-5] Sampling of saliva can be accomplished without pain and stress for the donor, therefore representing a suitable bodily fluid that may be collected and investigated also in pediatric age.[6] In the last years our group, as well as other ones, investigated the salivary proteome of preterm newborns and subjects in the pediatric age highlighting striking differences with respect to adults.[7-11] By an integrated top-down/bottom-up platform we evidenced that uman preterm newborns saliva contains more than forty major salivary proteins either undetectable (according to the detection limit of our analytical method) or detectable in small amount in saliva of adults.[9] On the contrary, the proteins typically revealed in adult were undetectable or present in small amount in preterm newborn saliva. In the present study the human salivary proteome was investigated in a large group of subjects over a wide range of ages, starting from preterm newborns of about 180 days of post-conceptional age (PCA) up to 17 years-old subjects, to define the appearance and the level changes of the proteins typically found in adult saliva from the last months of fetal development to adulthood. Several post-translational-derived proteoforms have been also investigated as a function of age to obtain information on the time course activation of different enzymes, in particular kinases and convertases. The present study provides reference data to assess whether abnormal expression of salivary proteins and/or anomalous activities of specific enzymes involved in post-translational modifications occur during human development. The possibility to easily

highlight abnormalities of the salivary proteome, that may be clue of potential diseases, could have a significant impact on clinical practice, especially in pediatrics.

2. Experimental section

2.1 Reagents

All general chemicals and reagents of LC-MS grade were purchased from J.T.Baker (Deventer, the Netherlands), Merck (Darmstadt, Germany) and Sigma Aldrich (St. Louis, MO, USA).

2.2 Subjects enrolled, sample collection and treatment

The study protocol and written consent form were approved by the Ethical Committees of the University Hospital of Cagliari, of the Catholic University of Rome and by the Ethical Committee of the Neonatal Intensive Care Unit of the Institute of Clinical Pediatric of the Catholic University of Rome and, therefore, it has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. All rules were observed and written consent forms were signed or by the donor or by one parent of each child. For ethical reasons, saliva was only collected when sample collection caused no stress. 316 samples of resting whole saliva were gathered and analyzed from the following subjects:

a) 17 preterm newborns (8 males, 9 females) with a birth weight ranging between 500 and 1250 g and a PCA at birth of 178−217 days (26-31 weeks). Newborns with major congenital malformations or prenatal infections were excluded from the study. Sample collection was performed on the same preterm newborn during several weeks after birth at fixed time intervals (1 or 2 weeks). When possible,

it was also performed after discharge from the neonatal unit, during the periodical check visits within about 1 year follow-up. In this way 111 saliva specimens were analyzed covering a period from 178 to 545 days of PCA.

b) 182 donors (92 males, 90 females) with an age comprised between 0 to 17 years of age.

c) 23 adults (10 males, 13 females) with an age comprised between 27 and 57 years.

Collection time was usually established between 10.00 and 12.00 a.m. The donors were in healthy clinical conditions. Except for preterm newborns, donors did not eat or drink 2 h before the collection. Whole saliva was collected with a soft plastic aspirator as it flowed into the anterior floor of the mouth. After collection, each salivary sample was immediately diluted 1:1 (v/v) with 0.2% aqueous 2,2,2-trifluoroacetic acid (TFA) in ice bath. The solution was then centrifuged at 8000 g for 5 min (4 °C). Finally, the acidic supernatant was separated from the pellet and either immediately analyzed with the HPLC-ESI-MS apparatus or stored at -80 °C until the analysis.

2.3 RP-HPLC-ESI-MS analysis

2.3.1 RP-HPLC-high resolution MS/MS characterization of intact protein

High-resolution HPLC-ESI-MS/MS experiments were carried out by an Ultimate 3000 Nano/Micro HPLC apparatus (Dionex, Sunnyvale, CA, USA) equipped with a FLM-3000-Flow manager module and coupled to an LTQ Orbitrap XL apparatus (ThermoFisher, San Josè, CA, USA). A Dionex C8 column (3 μm particle diameter; column dimension 300 μm i.d.×15 cm) or a Zorbax 300 SB-C8 (3.5 μm particle diameter; column dimension 1 mm i.d.×15 cm) were used as chromatographic columns. The following eluents were used: (A) 0.1% (v/v) aqueous formic acid (FA) and (B) 0.01% (v/v) FA in

acetonitrile/water 80/20 v/v. The applied gradient was 0–4 min 5% B, 4–34 min from 5 to 50% B (linear), 34–64 min from 50 to 90% B (linear), at a flow rate of 80 μL/min. High-resolution positive MS/MS spectra were collected in full scan using the lock mass for internal mass calibration (polydimethyl cyclosiloxane, 445.1200 *m/z*) with the resolution of 60000 and 30000, respectively, and *m/z* range from 350 to 2000. In data-dependent acquisition mode the three most intense multiply-charged ions were selected and fragmented by using collision induced dissociation (CID, 35% normalized collision energy) and spectra were recorded. Alternatively, fragmentation was carried out using the same analysis conditions on selected multiply-charged ions corresponding to a specific protein mass. Tuning parameters were: capillary temperature 250 °C, source voltage 4.0 kV, capillary voltage 48 V, tube lens voltage 170 V. Detailed instrumental operative conditions are reported in Supplemental S1. The inject volume was set at 20 μL, corresponding to 10 μL of saliva.

2.3.2 RP-HPLC-low resolution MS analysis of intact protein

For low resolution (1/6000) MS experiments the HPLC-ESI-IT-MS apparatus used was a Surveyor HPLC system (ThermoFisher Scientific) connected by a T splitter to a photo diode-array detector and either to an LCQ Advantage ion trap mass spectrometer (ThermoFisher Scientific) or to an LCQ Deca XP Plus ion trap mass spectrometer (ThermoFisher Scientific). The chromatographic column was a 150 × 2.1 mm Vydac (Hesperia, CA, USA) C8 column, with 5-μm particle diameter. The following solutions were utilized for the RP-HPLC separation: (eluent A) 0.056% aqueous TFA and (eluent B) 0.05% TFA in acetonitrile-water 80/20. The gradient applied for the analysis was linear going from 0 to 55% of B in 40 min, and from 55% to 100% of B in 10 min, at a flow rate of 0.30 mL/min. The T splitter permitted 0.20 mL/min to flow toward the diode array detector and 0.10 mL/min to flow toward

the ESI source. The UV diode array detector was set at 214 and 276 nm. The MS spray voltage was 4.0 kV and the capillary temperature was 220 °C. Detailed instrumental operative conditions are reported in Supplemental S2. The inject volume was set at 100 µL, corresponding to 50 µL of saliva.

2.4 MS data analysis and statistics

Deconvolution of averaged ESI-MS and ESI-MS/MS spectra was automatically performed by using MagTran 1.0 software[12] or the HPLC-MS apparatus management software (Xcalibur 2.0.7 SP1, Thermo Fisher Scientific). Pearson *r* coefficient was used to evaluate linear correlations. Differences with p value < 0.05 were considered statistically significant.

2.5 Intact protein/peptide characterization and relative quantification

The different salivary proteins and their proteoforms investigated have been previously characterized by means of an integrated top-down/bottom-up proteomic platform.[13] The data obtained from high resolution ESI-MS and MS/MS utilized for the protein/peptide characterization are summarized in Supplemental S3. Salivary peptide and protein quantitative analysis of all the 316 whole saliva specimens reported in Section 2.2 was based on the measurement of the low-resolution HPLC-ESI-MS eXtracted Ion Current (XIC) peak area (signal/noise ratio > 5). Relative quantification by XIC procedure was also carried out by high-resolution HPLC-ESI-MS in selected samples. The percentages calculated by high-resolution HPLC-ESI-MS XIC peak areas were in agreement with those calculated by low-resolution HPLC-ESI-MS peak areas (within the limit of the experimental error). The XIC search reveals the peak associated with the protein of interest by extracting along the total ion current (TIC) chromatographic profile the intensity of the ion current of specific multiply-charged ions (*m/z*)

generated by the ESI source. The ions used to quantify the proteins and peptides were carefully selected to exclude values in common with other co-eluting analytes. The area of the XIC peak was related to the saliva volume (1.0 mL) and it is proportional to protein/peptide concentration, therefore, under constant analytical conditions, it can be used for a quantitative analysis and comparative studies.[14-16] The estimated percentage error of the XIC procedure was minor than 8%.

The *m/z* values utilized for the determination of the (XIC peak area)/(mL of saliva) of the proteins of this study are reported in Supplemental S4. The *m/z* values of acidic and basic proline-rich proteins, histatins, statherin, P-B peptide and S-type cystatins were also reported in our previous study.[11] The *m/z* values utilized for the determination of the (XIC peak area)/(mL of saliva) of the different proteoforms of S100A9 were already described in reference.[17]

3. Results

3.1 Expression of human salivary proteins in the pediatric age

HPLC-ESI-MS investigation of the acidic soluble fraction of 293 human whole saliva samples collected from preterm newborns (N =111) and from subjects aged 0-17 years (N = 182) highlighted some relevant aspects of the human salivary proteome: a) it varies according with the age up to adulthood; b) it is characterized by a very high inter-individual variability; c) proteins typically found in adults appear in saliva with different time courses d) certain proteoforms were undetectable in saliva of preterm and at-term newborns. Therefore, the number of samples utilized for the computation of the percentages in the Figures varies for each protein or protein family.

Some of these features are evident in the Scheme of Fig. 1, which shows the time course expression of

the different salivary proteins and peptides investigated in this study and grouped on the basis of their locus. Indeed, the thickness of the lines in the scheme gives a rough indication of the percentages of protein level at the different ages. In Figure 1 and in all figures displaying temporal changes, a logarithmic scale has been used in the abscissa axis to better highlight the major changes occurring in preterm newborns in the first weeks after birth, which correspond to the last period of fetal development and those occurring immediately after the normal term of delivery and in the first years of life. In the scheme, the dashed lines highlights the age range in which proteins were detected in at least one baby evidencing the high variability observed in the first period of life.

3.1.1 Acidic proline-rich proteins

Acidic proline-rich proteins (aPRPs) are a family of salivary proteins encoded by *PRH1* and *PRH2* loci located on chromosome 12p13.2. The majority of individuals in the western population express five major isoforms of aPRPs: PIF-s, Db-s (where "s" stands for slow), and Pa (commonly detected as 2-mer) encoded by the *PRH1* locus, PRP-1 and PRP-2, encoded by the *PRH2* locus.[18] PRP-1, PIF-s (isobaric, average mass (Mav) 15514 Da) and PRP-2 (Mav 15515 Da) are indistinguishable by low-resolution HPLC-ESI-MS analysis, while Db-s (Mav 17633 Da) and Pa 2-mer (Mav 30921 Da) can be discriminated.[18] The proteoforms with Mav 15514-15515 Da were detected even though at small amount in all the samples from preterm newborns since 180-190 days of PCA, while Pa 2-mer and Db-s have been never detected in preterm newborns before 270-280 days of PCA, that correspond to the normal term of delivery. For this reason it was hypothesized that the *PRH2* gene expression starts during fetal life, slightly before 180 days of PCA, and that of *PRH1* gene starts later, after the normal term of delivery.

3.1.2 Basic proline-rich proteins and glycosylated basic proline-rich proteins

Basic proline-rich proteins (bPRPs) are the most complex family of salivary proteins secreted only by parotid gland. They are the expression product of four genes (*PRB1-PRB4*) located on chromosome 12p13.2 closely to the *PRH1-2* loci. *PRB1* and *PRB3* loci exhibit at least four common alleles (S, M, L, VL) in the western population, *PRB2* locus four (S, M. L(Con1+), L(Con1-)), and *PRB4* three (S, M, L).[13,19] As evidenced in Figure 2, bPRPs encoded by *PRB1*, *PRB2* and *PRB4* genes are only detectable as multiple peptide fragments generated by pre-secretory cleavages of the pro-proteins, and the connection between the most common haplotypes and salivary phenotypes is still pending complete definition.[13,14,19] Expression of some bPRPs occurs later with respect to all the other salivary proteins. Indeed, non-glycosylated bPRPs were undetectable, or detectable at small amount even in saliva of four ± one years old children. Their (XIC peak area)/(mL of saliva) increased with age reaching levels comparable to adults only after puberty. A peak attributed to a very heterogeneous mixture of glycoforms of bPRPs, whose structure is under characterization in our laboratories, was detected in preterm newborn saliva since 180 days of PCA. It is very likely that these glycoforms are the expression product of the *PRB3* gene, because diverse glycosylated bPRPs are coded by the *PRB4* gene together with P-D peptide (see Fig. 2) that has never been detected in saliva from newborns. Thus, the present data show that the expression of the *PRB3* gene occurs early during the last months of fetal life and precedes that of the other bPRPs genes (Fig. 1).

3.1.3 Histatins

The *HIS1(*or *HTN1)* and *HIS2* (or *HTN3*) genes, responsible for the expression of histatin 1 (Hst1) and

histatin 3 (Hst3), respectively, are closely located on chromosome 4q13.3.[20,21] From our data we assumed that *HIS1* protein expression started around seven months (210 days) ± two weeks of PCA. Furthermore, as already reported in a recent study, we confirmed that Hst1 showed two maxima of concentrations around 1 and 6 years of age, in correspondence with deciduous and permanent dentitions.[11] Concerning Hst3 and its derivatives histatin 6 (Hst6) and histatin 5 (Hst5) the present study suggested that *HIS2* protein expression commonly starts later than *HIS1* roughly after the normal term of delivery (270-280 days of PCA).

3.1.4 Statherin and P-B peptide

STATH and *PBII* genes located on chromosome 4q13.3 encode statherin, and P-B peptide (also called submaxillary gland androgen-regulated protein 3B), respectively.[20,21] As Hst1, statherin and P-B peptide were detected in whole saliva of preterm newborns at seven months (± two weeks) of PCA. The increase in concentration with age of the two peptides followed a similar trend. In some preterm newborns, both peptides were detected only few days after the normal term of delivery (270-280 days of PCA). The (XIC peak area)/(mL of saliva) of the two peptides showed a very significant linear correlation coefficient ($R = 0.567$, $N = 193$, $p < 0.00001$) in agreement with the hypothesis that P-B peptide is functionally more connected to statherin than to bPRPs.[13,22]

3.1.5 S-type cystatins

Cystatin S (S1, S2), SN and SA are called S-type cystatins and are the expression products of three loci closely located on chromosome 20p.11.21.[21] In previous studies, it has been shown that they are mainly

secreted by submandibular/sublingual glands.[21,22] S-type cystatins were not detected in any sample from preterm and at-term newborns and became detectable at one year (± three months) of age.

3.1.6 S100A9

S100A9, expression product of a locus located on chromosome 1q21 showed an opposite trend with respect to the other proteins investigated.[23] It belongs to the group of proteins found at high concentration in preterm newborn saliva,[9] that decreased as a function of PCA and became undetectable or detectable at low level after the normal term of delivery.

3.2 Post-translational modifications

Age-dependent variations of the relative concentration of post-translationally modified proteoforms of salivary proteins can disclose valuable information on the timing of activation of the enzymes involved.

3.2.1 Phosphorylation

Many phosphorylated proteins related to kinase activities have been detected in human whole saliva by bottom-up proteomic studies.[24-26] The phosphorylation sites of salivary proteins and peptides characterized in this study are described in the following sections.

3.2.1.1 Acidic proline-rich proteins

Ser_8 and Ser_{22} are the main phosphorylation sites of aPRPs. The di-phosphorylated proteins represent the major forms, but small amounts of tri-phosphorylated (on Ser_{17}), mono- and non-phosphorylated

proteoforms have been also detected in adult human saliva.[18] In the present study our previous observation of the decrease of non-phosphorylated and mono-phosphorylated aPRPs percentages as a function of PCA in preterm newborns[27] was confirmed on a greater number of subjects. The percentages of the different phosphorylated entire aPRPs proteoforms (PRP-1 types, including PRP-1, PRP-2 and PIF-s) measured in 263 samples are shown in Fig. 3 as a function of the logarithm of PCA. The panels a, b and c show the percentage of non-, mono- and di-phosphorylated aPRPs, respectively. The trend lines of Fig. 3 were obtained by fitting experimental data to exponential functions. The mean percentage value measured on adults (N = 23) is reported on the right side of all the figures at an arbitrary value of 9500 days of PCA. PRP-1 type aPRPs are all cleaved at Arg_{106} generating truncated PRP-3 type proteoforms (including PRP-3, PRP-4 and PIF-f, where "f" stands for fast). Fig. 4a,b,c show the percentages of the different phosphorylated proteoforms of truncated PRP-3 type aPRPs measured in 258 samples. Figure features are identical to those of Fig. 3. The present data, in agreement with previous observations,[27] highlight that the kinase involved in the phosphorylation of aPRPs is not fully active even at birth and reaches the activity comparable to that of adults at around two years of age.

3.2.1.2 Histatin 1

Hst1 is mono-phosphorylated on Ser_2. The HPLC-ESI-MS profile of Fig. 5 shows that the non-phosphorylated proteoform eluted close to Hst1 under our experimental conditions.[13,22] The two broad and intense peaks co-eluting with Hst1 correspond to the highly charged aPRPs proteoforms (entire 23-24 min, truncated 24-25 min). The percentage of non-phosphorylated Hst1 is reported in Fig. 6 as a function of the logarithm of PCA. Percentages were calculated from the measured XIC peak area

only in 230 samples because Hst1 was not detected in saliva of most of the premature newborns with a low PCA. Fig. 6 shows that the activation timing of the kinase responsible for the phosphorylation of Hst1 is very similar to the one observed for aPRPs, though Hst1 becomes detectable approximately two weeks later than aPRPs.

3.2.1.3 Statherin

Statherin is di-phosphorylated on Ser_1 and Ser_2. In the HPLC-ESI-MS profile, the mono- and non-phosphorylated proteoforms were detected just before the peak of statherin.[13,22] The percentages of non- (a), mono- (b) and di-phosphorylated statherin (c) measured in 221 samples are shown in Fig.7 as a function of the logarithm of PCA. The trend of phosphorylation of statherin during human development was found very similar to that of aPRPs and Hst1.

3.2.1.4 Cystatin S, S1, S2

Cystatin S1 and S2 correspond to mono-phosphorylated on Ser_3 and di-phosphorylated on Ser_1 and Ser_3 proteoforms of cystatin S, respectively.[21] Fig. 8a,b,c shows the percentages of the three proteoforms as a function of the logarithm of PCA measured in 124 samples. It is evident that the percentage of the three cystatins (cystatin S 6±4 %, S1 68±9 %, S2 26±10 %) does not change in the age range investigated.

3.2.1.5 IB-1 and II-2 basic proline-rich proteins

IB-1 and II-2 are two bPRP peptides, 96 and 75 residues long, respectively, closely eluting in the HPLC-ESI-MS profile and sharing very similar sequences and both phosphorylated on $Ser_{8,}$.[13,19] These

bPRPs are undetectable in whole saliva before 4-6 years of age. When revealed, the phosphorylated proteoforms were the prevalent ones, being the non-phosphorylated sporadically detected, as in the adults (data not reported).

3.2.1.6 S100A9

At least four different proteoforms of S100A9 are commonly detectable in human saliva: S100A9 short, generated by the loss of the MTCKM N-terminal pentapeptide, S100A9 long, generated by the loss of the N-terminal Met residue, and the S-glutathionylated and S-cysteinylated derivatives (on Cys_2) of the long form, all N-terminally acetylated.[17] All the S100A9 proteoforms can be phosphorylated on the penultimate Thr residue. Fig. 9a shows the percentages of phosphorylation of the short proteoform and Fig. 9b the percentage of phosphorylation of all the long proteoforms measured in 187 samples. The percentages resulted very similar (mean value 32 ± 6) and did not change as a function of PCA, suggesting that the kinase involved in the phosphorylation of S100A9 is fully active since 190 days of PCA.

3.2.2.1 Endo- and exo-peptidases: acidic proline-rich proteins

Different endo- and exo- peptidases work during granule storage and secretion of salivary proteins. Among aPRPs, PRP-1, PRP-2, PIF-s and Db-s are cleaved at the Arg_{106} residue (Arg_{127} residue in Db-s) by the action of a convertase that recognizes the consensus sequence RXXR↓ generating four truncated proteoforms called PRP-3, PRP-4, PIF-f and Db-f and a common peptide of 44 amino acid residues called P-C peptide.[14,18,21] The Pa proteoform, due to the substitution Arg_{103}→Cys abolishing the RXXR↓ consensus sequence, is not cleaved. It is commonly detected in whole saliva as S-S 2-mer.[18]

As reported in Sections 3.1.1 and 3.2.1.1, PRP-1, PRP-2 and PIF-s (PRP-1 type) as well as their truncated derivatives PRP-3, PRP-4 and Pif-f (PRP-3 type, Mav 11162-11163 Da) were detected in whole saliva of preterm newborns since 180 days of PCA. The percentage of truncated PRP-3 type proteoforms computed with respect to the sum of PRP-1 and PRP-3 type proteoforms measured in 261 samples is shown in Fig.10a as a function of the logarithm of PCA. Because this percentage is indicative of the activity of the convertase responsible for aPRPs cleavage, data in Fig. 10a confirm that this enzyme is more active during fetal growth, as previously determined by our group.[27]

3.2.2.2 Endo- and exo-peptidases: histatins 3, 5 and 6

Hst3 is a small histidine-rich peptide of 32 amino acid residues partially cleaved before granule storage at Arg_{25} by a convertase generating Hst6 (Hst3 fr. 1/25).[22] The subsequent removal of Arg_{25} by a carboxy-peptidase generates Hst5 (Hst3 fr. 1/24). To evaluate the convertase activity, the percentage of Hst5+Hst6 was computed with respect to the sum of Hst3, Hst5 and Hst6 in 192 samples and it is shown in Fig. 10b as a function of the logarithm of PCA. Similarly, to evaluate the carboxy-peptidase activity, the percentage of Hst5 was computed with respect to the sum of Hst5 and Hst6 in 188 samples and it is shown in Fig.10c. From the Figures it is evident that the convertase and the carboxy-peptidase activities are high in pre-term newborns and decrease according to PCA reaching the activities characteristic of the adults at about one year of age.

3.2.2.3 Endo- and exo-peptidases: S100A9 short and long

The proteoforms of S100A9 detectable in whole saliva are the short one, and different derivatives of the long one.[17] The percentage of short S100A9 (phosphorylated and non-phosphorylated) with respect

to the sum of all the proteoforms of short and long S100A9 (cysteinylated, glutathionylated, phosphorylated and non-phosphorylated) measured in 186 samples (Fig. 10d) shows that the percentage of S100A9 short increases slightly during human development.

3.2.3 N-terminal acetylation

The various proteoforms of S100A9 detected in this study are all N-terminally acetylated. It was not possible to observe the non-acetylated S100A9 proteoforms even in preterm newborns with less than 190 days of PCA. This result suggests that the N-terminal acetyltransferase (NAT) responsible for the PTM is fully active at 6 months of fetal life.

3.2.4 Sulfation

Hst1 derivatives with molecular mass increments of 80, 160, 240 and 320 Da corresponding to mono-, di-, tri- and tetra-sulfo-tyrosines proteoforms are detectable in whole human saliva. These derivatives represent about 10% of total Hst1 being this post-translational modification (PTM) specific of submandibular/sublingual gland.[28] Sulfo-proteoforms of Hst1 appeared in whole saliva during the first year of life. They were detected in some newborns few weeks after birth, but they were absent in other subjects till one year of age. After one year of age, the sulfo-derivatives were always present in all the subjects under study. Therefore, the tyrosylprotein sulfotransferase (TPST) seems to be not active in submandibular/sublingual gland during fetal life and its activation seems to occur during the first year of life.

4 Discussion

4.1 General considerations

An important contribution to the variability of any proteome is given by the physiological modifications occurring according to time dimension. The exploration of proteomic temporal changes may be defined as "chrono-proteomics". The present investigation evidenced that individual variability of the salivary proteome is very high in the pediatric age, especially in preterm newborns, as shown in Figures 3,4 and 6-10. However, the large number of analyzed samples granted a statistical significance of the temporal trends observed. Two types of sampling were performed: longitudinal, on 17 preterm newborns up to 545 days of PCA for a total of 111 specimens and cross-sectional on subjects aged between 0-17 years for a total of 182 specimens. Dashed lines of Fig. 1 represent the age range in which specific proteins were detected in at least one baby. The upper limit of this range represents the age in which saliva of all the babies displayed the specific protein in the chromatographic profile, even if at small amount. Whole saliva of preterm newborns contains a lot of proteins involved in the fetal growth, some of them characterized by our group in a previous work.[9] Many of them pertain to the families of S100 proteins, β-thymosins, and cystatins. Several of these proteins have been also detected in a large proteomic survey of selected salivary glandular secretions from adults by other groups.[5] However, the different proteoforms were not characterized because of the bottom-up platforms applied. Concentration of these proteins dropped quickly after birth, and they almost disappeared at the normal term of delivery.[9] In the meantime, proteins typically found in adult saliva became detectable, even though with different time courses. aPRPs encoded by *PRH2* locus and glycosylated bPRPs encoded by *PRB3* locus appear first in saliva since 180 days of PCA, followed by Hst1, statherin and P-B peptide

that appear at about seven months of PCA. Hst3, Hst5, Hst6 and aPRPs encoded by *PRH1* locus appear after the normal term of delivery and S-type cystatins at about one year of age. Finally, bPRPs encoded by *PRB1, PRB2*, and *PRB4* loci appear not before 3-4 years of age. The hierarchical expression of the salivary proteins during human development should have a functional meaning. In this respect, it has been suggested that the important modifications observed in Hst1 salivary concentration during infancy could be linked to the wound healing and tissue formation properties of this peptide.[29,30] The late appearance of bPRPs, rather than linked to the ability to protect oral cavity against the harmful effects of tannins,[31] could be connected to changes in diet and taste perception[32] and/or perhaps to the maturation of secondary sexual traits. The majority of the salivary protein families are probably involved in the modulation and selection of the oral and gastrointestinal microbiota that experiences deep modifications during pediatric age.[33,34] Undoubtedly, the role of P-B peptide and its functional connection with statherin, the functional meaning of the partial cleavage of aPRPs, as well as the significance of the different proteome of the salivary glands are the most challenging questions concerning human saliva. The results of this study stimulate further questions on the functions of salivary proteins and in the meantime offer new elements to elucidate the role of some of them.

4.2 Phosphorylation

The similar trend of the level of phosphorylation (see Fig.s 3,4,6) and the common and highly specific SxE/pS consensus sequence supported the hypothesis that aPRPs, Hst1 and statherin are all substrates of the same kinase.[14] The kinase, working in the Golgi apparatus of granule secreting cells, was firstly defined as Golgi casein kinase and for long time escaped any attempt of purification and structural characterization.[35] Recently, it was recognized as Fam20C, which belongs with Fam20A and B to a

family of kinases with sequence similarity, all implicated in the homeostasis of mineralized tissues.[36] The results of this study suggested that the expression of this enzyme starts in the salivary glands slightly before 200 days of PCA. Its activity slowly increases and reaches values comparable to the adults at around two years of age. Fam20C is a pleiotropic enzyme responsible for the phosphorylation of a lot of different substrates in many organs and tissues, the central nervous system included.[37] Mutations in Fam20C gene in humans cause the Raine syndrome, an extremely rare autosomal recessive osteosclerotic bone dysplasia.[38] The Raine syndrome is usually fatal. In the few cases described worldwide, the patients (less than 40) did not survive longer than few weeks after birth. Few non-lethal cases have been reported, with great neurological and physiological impairments.[39] In a recent study by our group significant hypo-phosphorylation of aPRPs, statherin and Hst1 in a subgroup of patients with autism spectrum disorders has been observed.[40] The crucial role of Fam20C suggests that it has to be strictly synchronized with other enzymes and relevant proteins during development and that a small delay in the activation during fetal growth could generate severe neurological impairments.[41] Also cystatin S and the bPRPs IB-1 and II-2 peptides show the consensus sequence recognized by Fam20C. The level of phosphorylation of these proteins did not show a trend to increase with age, because these proteins appear in human saliva when the enzyme is fully active. S100A9, instead, is phosphorylated by MAPK14, a kinase pertaining to the p38 mitogen-activated protein kinase pathway.[42] Results of Fig. 9 clearly show that this kinase is fully active during the last months of fetal growth and its expression largely anticipates that one of Fam20C during fetal development.

4.3 Endo- and exo-peptidases

To date, few investigations have been performed in order to characterize proteinases involved in the cleavage of salivary proteins.[21] The principal cleavage consensus sequences are potentially recognizable by enzymes pertaining to the furin-like convertase families.[14,22] Although the cleavage consensus sequences are very similar, it is not known if a common convertase acts on the different families of salivary proteins or rather different specific enzymes are involved in the cleavage of each protein. bPRPs are completely cleaved before granule storage and the pro-protein is not detectable at all in whole saliva. Instead, aPRPs and Hst3 are partially cleaved and a high percentage of the intact with respect to the truncated proteoforms can be detected in saliva. Two studies addressed the partial characterization of the proteinases acting on aPRPs and bPRPs.[43,44] A study of Basak and colleagues[45] established that Hst3 acts *in vitro* as substrate for rat PC1 (proprotein convertase 1), but no one study was carried out in order to characterize the proteinase *in vivo* and to establish if the convertase involved in the cleavage of PRPs could be the same. Results from the present study evidenced that the activity of the convertase/s is higher during the last months of fetal growth (preterm newborn saliva) than after birth. We have previously observed this phenomenon in a limited number of subjects in the cleavage of aPRPs.[27] This study confirms this observation in a large number of subjects and highlights a similar trend in the cleavage of Hst3 (Fig. 10b). An increased enzymatic activity during the last months of fetal growth was also detectable for the carboxy-peptidase involved in the conversion of Hst6 to Hst5 (Fig. 10c). The removal of C-terminal Arg residues after the convertase cleavage is an event widespread in many secretory processes.[46] In salivary glands this activity could be ascribed to the metalloproteinase carboxypeptidase Z,[47] which is active toward substrates with basic C-terminal amino acids, such as in the case of the formation of Hst5 from Hst6.[22] On the whole, our results suggested that all the

proteinases involved in the maturation of salivary proteins are more active during fetal growth, as expected in a period of very rapid cell proliferation and tissue growth.

The relative proportion between short and long S100A9 is probably due to the recognition of similar consensus sequences by the enzyme methionine aminopeptidase (MAP), responsible for the removal of the initiator methionine after protein synthesis.[48] A particular N-terminal structural arrangement could induce MAP to cleave the fifth instead of the first methionine residue of the protein chain. Figure 10d suggests that the MAP recognition slightly changes during development, because the percentage of the S100A9 short proteoforms seems to increase a little according to age. Further experiments will be necessary to better clarify these results and to establish if the short and long S100A9 proteoforms have different functional roles.

4.4 N-terminal acetylation

N-terminal acetylation is a common protein modification present on more than 80% of cytosolic human proteins. It occurs mainly co-translationally by the action of N-terminal acetyltransferases (NATs).[49] On the basis of the S100A9 N-terminal residues, a threonine in the long proteoforms and a serine in the short proteoforms, the enzyme should be NATa.[49] From the present data and from our previous studies,[9] NATa as well as any other NATs working on salivary proteins resulted fully active before 7 months of fetal life.

4.5 Sulfation

Tyrosine sulfation is a widespread PTM implicated in various cellular and extra-cellular processes.[50] Sulfation of tyrosines occurs by the transfer of sulfate from the universal sulfate donor

3'-phosphoadenosine 5'-phosphosulfate to the hydroxyl group of tyrosine residues. Until now two different tyrosylprotein sulfotransferases (TPST1 and TPST2) have been identified, and all mammalian cell types and cell lines studied to date express both.[51] Hst1 has been the first sulfo-peptide detected in human saliva.[28] Tyrosines 27, 30, 34, and 36 are involved in the sulfation of Hst1 and the PTM is probably hierarchical. Tyr_{27} satisfies several features required for enzyme recognition, such as the presence of two acidic amino acid residues at positions +2 and -4, only one basic residue at -5 position, and several turn-inducing amino acid residues (Pro, Ser, Gly, Asn) in the ±7 proximity.[52] Therefore, O-sulfation of Tyr_{27} in Hst1 may be the first event that activates the sequential sulfation of the other three Tyr residues. The results of the present study indicate that in submandibular/sublingual glands TPST activity starts during the first year of life. It could be of great interest to establish in the future if other fetal organs and tissues show the delayed activation of TPSTs in addition to salivary glands.

5. Acknowledgments

The authors acknowledge the financial support of the Nando Peretti Foundation, the Cagliari University, the Catholic University of Rome, the MIUR, the Italian National Research Council (CNR), and the Regione Autonoma Sardegna, Fin. 2013, Lg. Reg. N. 7, project "Studio Proteomico della Saliva Umana in età pediatrica per la diagnosi precoce delle malattie metaboliche congenite". Morena Arba and Monica Sanna gratefully acknowledge Sardinia Regional Government for the financial support of their PhD scholarships (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.).

The authors declare no competing financial interest

6. References

(1) Tabak, L. A. A revolution in biomedical assessment: the development of salivary diagnostics. *J. Dent. Educ.* **2001**, *65*, 1335–1339.

(2) Wang, Q.; Yu, Q.; Lin, Q.; Duan, Y. Emerging salivary biomarkers by mass spectrometry. *Clin. Chim. Acta* **2015**,*438*,214-221.

(3) Schafer, C. A; Schafer, J. J, ; Yakob, M.; Lima, P.; Camargo, P.; Wong, D. T. Saliva diagnostics: utilizing oral fluids to determine health status. *Monogr. Oral Sci.* **2014**,*24*,88-98.

(4) Amado, F. M.; Ferreira, R. P.; Vitorino, R. One decade of salivary proteomics: current approaches and outstanding challenges. *Clin. Biochem.* **2013**,*46*,506-517.

(5) Denny, P.; Hagen, F. K.; Hardt, M.; Liao, L.; Yan, W.; Arellanno, M.; Bassilian, S.; Bedi, G. S.; Boontheung, P.; Cociorva, D.; Delahunty, C. M.; Denny, T.; Dunsmore, J.; Faull, K. F.; Gilligan, J.; Gonzalez-Begne, M.; Halgand, F.; Hall, S. C.; Han, X.; Henson, B.; Hewel, J.; Hu, S.; Jeffrey, S.; Jiang, J.; Loo, J. A.; Ogorzalek Loo, R. R.; Malamud, D.; Melvin, J. E.; Miroshnychenko, O.; Navazesh, M.; Niles, R.; Park, S. K.; Prakobphol, A.; Ramachandran, P.; Richert, M.; Robinson, S.; Sondej, M.; Souda, P.; Sullivan, M. A.; Takashima, J.; Than, S.; Wang, J.; Whitelegge, J. P.; Witkowska, H. E.; Wolinsky, L.; Xie, Y.; Xu, T.; Yu, W.; Ytterberg, J.; Wong, D. T.; Yates, J. R. 3rd; Fisher, S. J. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. *J Proteome Res.* **2008,** *7*, 1994-2006.

(6) Romano-Keeler, J.; Wynn, J. L.; Maron, J. L. Great expectorations: the potential of salivary 'omic'

approaches in neonatal intensive care. *J. Perinatol.* **2014** ,*34*,169-173.

(7) Morzel, M.; Palicki, O.; Chabanet, C.; Lucchi, G.; Ducoroy, P.; Chambon, C.; Nicklaus, S. Saliva electrophoretic protein profiles in infants: changes with age and impact of teeth eruption and diet transition. *Arch. Oral Biol.* **2011**,*56*,634-642.

(8) Morzel, M.; Jeannin, A.; Lucchi, G.; Truntzer, C.; Pecqueur, D.; Nicklaus, S.; Chambon, C.; Ducoroy, P. Human infant saliva peptidome is modified with age and diet transition. *J Proteomics* **2012**,*75*,3665-3673.

(9) Castagnola, M.; Inzitari, R.; Fanali, C.; Iavarone, F.; Vitali, A.; Desiderio, C.; Vento, G.; Tirone, C.; Romagnoli, C.; Cabras, T.; Manconi, B.; Sanna, M. T.; Boi, R.; Pisano, E.; Olianas, A.; Pellegrini, M.; Nemolato, S.; Heizmann, C. W.; Faa, G.; Messana, I. The surprising composition of the salivary proteome of preterm human newborn. *Mol. Cell. Proteomics* **2011**, *10*, M110.003467.

(10) Cabras, T.; Pisano, E.; Boi, R.; Olianas, A.; Manconi, B.; Inzitari, R.; Fanali, C.; Giardina, B.; Castagnola, M.; Messana, I. Age-dependent modifications of the human salivary secretory protein complex. *J. Proteome Res.* **2009**, *8*, 4126-4134.

(11) Manconi, B.; Cabras, T.; Pisano, E.; Sanna, M. T.; Olianas, A.; Fanos, V.; Faa, G.; Nemolato, S.; Iavarone, F.; Castagnola, M.; Messana, I. Modifications of the acidic soluble salivary proteome in human children from birth to the age of 48 months investigated by a top-down HPLC-ESI-MS platform. *J. Proteomics* **2013**, *91*, 536-543.

(12) Zhang, Z.; Marshall, G. A. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 225–233.

(13) Castagnola, M.; Cabras, T.; Iavarone, F.; Vincenzoni, F.; Vitali, A.; Pisano, E.; Nemolato, S.;

Scarano, E.; Fiorita, A.; Vento, G.; Tirone, C.; Romagnoli, C.; Cordaro, M.; Paludetti, G.; Faa, G.; Messana, I. Top-down platform for deciphering the human salivary proteome. *J. Matern. Fetal Neonatal Med.* **2012**, *25* (Suppl 5), 27-43.

(14) Messana, I.; Inzitari, R.; Fanali, C.; Cabras, T.; Castagnola, M. Facts and artifacts in proteomics of body fluids. What proteomics of saliva is telling us? *J. Sep. Sci.* **2008**, *31*, 1948-1963.

(15) Levin, Y.; Schwarz, E.; Wang, L.; Leweke, F. M.; Bahn, S. Labelfree LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. *J. Sep. Sci.* **2007**, *30*, 2198−2203.

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

(17) Castagnola, M.; Cabras, T.; Iavarone, F.; Fanali, C.; Messana, I. Detection of Ca(2+)-binding S100 proteins in human saliva by HPLC-ESI-MS. *Methods Mol. Biol.* **2013**, *963*, 357-371.

(18) Inzitari, R.; Cabras, T.; Onnis, G.; Olmi, C.; Mastinu, A.; Sanna, M. T.; Pellegrini, M.; Castagnola, M.; Messana, I. Different isoforms and post-translational modifications of human salivary acidic proline-rich proteins. *Proteomics* **2005**, *5*, 805-815.

(19) Messana, I.; Cabras, T.; Inzitari, R.; Lupi, A.; Zuppi, C.; Olmi, C.; Fadda, M. B.; Cordaro, M.; Giardina, B.; Castagnola., M. Characterization of the human salivary basic proline-rich protein complex by a proteomic approach. *J. Proteome Res.* **2004**, *3*, 792-800.

(20) Helmerhorst, E. J. and Oppenheim, F. G. Saliva: a dynamic proteome. *J. Dent. Res.* **2007**, *86*, 680-693.

(21) Oppenheim, F. G.; Salih, E.; Siqueira, W. L.; Zhang, W.; Helmerhorst, E. J. Salivary proteome and its genetic polymorphisms. *Ann. N.Y. Acad. Sci.* **2007**, *1098*, 22-50.

(22) Messana, I.; Cabras, T.; Pisano, E.; Sanna, M. T.; Olianas, A.; Manconi, B.; Pellegrini, M.;

Paludetti, G.; Scarano, E.; Fiorita, A.; Agostino, S.; Contucci, A. M.; Calò, L.; Picciotti, P. M.; Manni, A.; Bennick, A.; Vitali, A.; Fanali, C.; Inzitari, R.; Castagnola, M. Trafficking and post-secretory events responsible for the formation of secreted human salivary peptides: a proteomics approach. *Mol. Cell. Proteomics* **2008**, *7*, 911-926.

(23) Odink, K.; Cerletti, N.; Bruggen, J.; Clerc, R. G.; Tarcsay, L.; Zwadlo, G.; Gerhards, G.; Schlegel, R.; Sorg, C. Two calcium-binding proteins in infiltrate macrophages of rheumatoid arthritis. *Nature* **1987**,*330*,80-82.

(24) Vitorino, R.; Alves, R.; Barros, A.; Caseiro, A.; Ferreira, R.; Lobo, M. C.; Bastos, A.; Duarte, J.; Carvalho, D.; Santos, L. L; and Amado, F. L. Finding new posttranslational modifications in salivary proline-rich proteins. *Proteomics* **2010**, *10*, 3732-3742.

(25) Stone, M. D.; Chen, X.; Mc Gowan, T.; Bandhakavi, S.; Cheng, B.; Rhodus, N. L.; Griffin, T.J. Large-scale phosphoproteomics analysis of whole saliva reveals a distinct phosphorylation pattern. *J. Proteome Res.* **2011**, *10*, 1728-1736.

(26) Salih E.; Siqueira, W. L.; Helmerhorst, E.J.; Oppenheim, F. G. Large-scale phosphoproteome of human whole saliva using disulfide-thiol interchange covalent chromatography and mass spectrometry. *Anal. Biochem.* **2010**, *407*, 19-33.

(27) Inzitari, R.; Vento, G.; Capoluongo, E.; Boccacci, S.; Fanali, C.; Cabras, T.; Romagnoli, C.; Giardina, B.; Messana, I.; Castagnola, M. Proteomic analysis of salivary acidic proline-rich proteins in human preterm and at-term newborns. *J. Proteome Res.* **2007**, *6*, 1371-1377.

(28) Cabras, T.; Fanali, C.; Monteiro, J. A.; Amado, F.; Inzitari, R.; Desiderio, C.; Giardina, B.; Castagnola, M.; Messana, I. Tyrosine polysulfation of human salivary histatin 1. A post-translational modification specific of the submandibular gland. *J. Proteome Res.* **2007**, *6*, 2472-2480.

(29) Oudhoff, M. J.; Bolscher, J. G.; Nazmi, K.; Kalay, H.; van 't Hof, W.; Amerongen, A. V.; Veerman, E. C. Histatins are the major wound-closure stimulating factors in human saliva as identified in a cell culture assay. *FASEB J.* **2008**, *22*, 3805-3812.

(30) Brand, H. S.; Ligtenberg, A. J.; and Veerman, E. C. Saliva and wound healing. *Monogr Oral Sci.* **2014**, *24*, 52-60.

(31) Bennick, A. A. Interaction of plant polyphenols with salivary proteins. *Crit. Rev. Oral Biol. Med.* **2002,** *13*, 184-196.

(32) Cabras, T.; Melis, M.; Castagnola, M.; Padiglia, A.; Tepper, B. J.; Messana, I.; Tomassini Barbarossa, I. Responsiveness to 6-n-propylthiouracil (PROP) is associated with salivary levels of two specific basic proline-rich proteins in humans. *PLoS One* **2012**, *7*, e30962.

(33) Guarino, A.; Wudy, A.; Basile, F.; Ruberto, E.; Buccigrossi, V. Composition and role of intestinal microbiota in children. *J. Matern. Fetal Neonatal Med.* **2012,** *25* (Suppl 1), 63-66.

(34) Karlsson, F. H.; Nookaew, I.; Petranovic, D.; Nielsen, J. Prospects for systems biology and modeling of the gut microbiome. *Trends Biotechnol.* **2011,** *29*, 251-258.

(35) Brunati, A. M.; Marin, O.; Bisinella, A.; Salviati, A.; Pinna, L. A. Novel consensus sequence for the Golgi apparatus casein kinase, revealed using proline-rich protein-1 (PRP1)-derived peptide substrates. *Biochem. J.* **2000,** *351*, 765-768.

(36) Tagliabracci, V. S.; Pinna, L. A.; and Dixon, J. E. Secreted protein kinases. *Trends Biochem. Sci.* **2013**, *38*, 121-130.

(37) Tibaldi, E.; Arrigoni, G.; Brunati, A. M.; James, P.; Pinna, L. A. Analysis of a sub-proteome which co-purifies with and is phosphorylated by the Golgi casein kinase. *Cell Mol. Life Sci.* **2006,** *63*, 378-389.

(38) Raine, J.; Winter, R. M.; Davey, A.; Tucker, S. M. Unknown syndrome: microcephaly, hypoplastic nose, exophthalmos, gum hyperplasia, cleft palate, low set ears, and osteosclerosis. *J. Med. Genet.* **1989**, *26*, 786-788.

(39) Fradin, M.; Stoetzel, C.; Muller, J.; Koob, M.; Christmann, D.; Debry, C.; Kohler, M.; Isnard, M.; Astruc, D.; Desprez, P.; Zorres, C.; Flori, E.; Dollfus, H.; Doray, B. Osteosclerotic bone dysplasia in siblings with a Fam20C mutation. *Clin. Genet.* **2011**, *80*, 177-183.

(40) Castagnola, M.; Messana, I.; Inzitari, R.; Fanali, C.; Cabras, T.; Morelli, A.; Pecoraro, A. M.; Neri, G.; Torrioli, M. G.; Gurrieri, F. Hypo-phosphorylation of salivary peptidome as a clue to the molecular pathogenesis of autism spectrum disorders. *J. Proteome Res.* **2008**, *7*, 5327-5332.

(41) Simpson, M. A.; Scheuerle, A.; Hurst, J.; Patton, M. A.; Stewart, H.; Crosby, A. H. Mutations in FAM20C also identified in non-lethal osteosclerotic bone dysplasia. *Clin. Genet.* **2009**, *75*, 271-276.

(42) Lominadze, G.; Rane, M. J.; Merchant, M.; Cai, J.; Ward, R. A.; McLeish, K. R. Myeloid-related protein-14 is a p38 MAPK substrate in human neutrophils. *J. Immunol.* **2005**, *174*, 7257-7267.

(43) Azen, E. A.; Amberger, E.; Fisher, S.; Prakobphol, A.; Niece, R. L. PRB1, PRB2, and PRB4 coded polymorphisms among human salivary concavalin-A binding, II-1 and Po proline-rich proteins. *Am. J. Hum. Genet.* **1996**, *58*, 143-153.

(44) Cai, K.; and Bennick, A. Processing of acidic proline-rich proprotein by human salivary gland convertase. *Arch. Oral. Biol.* **2004**, *49*, 871-879.

(45) Basak, A.; Ernst, B.; Brewer, D.; Seidah, N. G.; Munzer, J. S.; Lazure, C.; Lajoie, G. A. Histidine-rich human salivary peptides are inhibitors of proprotein convertases furin and PC7 but act as substrates for PC1. *J. Pept. Res* **1997**, *49*, 596-603.

(46) Steiner, D. F. The proprotein convertases. *Curr. Opin. Chem. Biol.* **1998**, *2*, 31-39.

(47) Song, L.; and Fricker, L. D. Cloning and expression of human carboxypeptidase Z, a novel metallocarboxypeptidase. *J. Biol. Chem.* **1997**, *272*, 10543–10550.

(48) Lowther, W. T.; and Matthews, B. W. Structure and function of the methionine aminopeptidases. *Biochim Biophys Acta* **2000**, *1477*, 157-167.

(49) Starheim, K. K.; Gevaert, K.; Arnesen, T. Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem. Sci.* **2012**, *37*, 152–161.

(50) Sasaki, N. Current status and future prospects for research on tyrosine sulfation. *Curr. Pharm. Biotechnol.* **2012**, *13*, 2632-2641.

(51) Stone, M. J.; Chuang, S.; Hou, X.; Shoham, M.; Zhu, J. Z. Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *N. Biotechnol.* **2009**, *25*, 299-317.

(52) Monigatti, F.; Hekking, B.; Steen, H. Protein sulfation analysis--A primer. *Biochim. Biophys. Acta* **2006**, *1764*, 1904-1913.

Legends to Figures

Figure 1

Scheme describing the approximate time course of the different salivary proteins and peptides investigated in this study. The sum of XIC peak areas of all the members belonging to the same family of proteins/peptides, grouped on the basis of their locus, is reported as function of age. A logarithmic scale has been used in the abscissa axis to better highlight the major changes occurring in preterm newborns during the first months of life, which correspond to the last months of fetal development, and those occurring immediately after the normal term of delivery and in the first years of life. Dashed lines represent the age range in which the proteins were detected in at least one baby. The upper limit of this range corresponds to the age in which saliva of all the babies displayed the protein in the chromatographic profile, even if at small amount.

Figure 2

Known bPRPs loci and their phenotypes, according to the most recent mass spectrometry characterizations.[13,19] The nomenclature of the different bPRPs was described by Castagnola et al.[13]

Figure 3

Percentages of non-phosphorylated (panel a), mono-phosphorylated (panel b) and di-phosphorylated (panel c) PRP-1 types aPRPs (comprehending PRP-1, PRP-2 and PIF-s) measured in 263 samples as a function of age (logarithmic scale). The trend lines were obtained by fitting experimental data to exponential functions. On the right side of the figure the mean value measured in the adults is reported

at an arbitrary value of 9500 days of PCA.

Figure 4

Percentages of non-phosphorylated (panel a), mono-phosphorylated (panel b), and di-phosphorylated (panel c) PRP-3 types aPRPs (comprehending PRP-3, PRP-4 and PIF-f) measured in 258 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.

Figure 5

XIC procedure for the relative quantification of co-eluting proteins/peptides. The upper panel shows the HPLC-ESI-MS total ion current profile in the elution range 22.3-26.7 min. The two broad and intense peaks are mainly due to the highly charged aPRPs proteoforms (entire at 23-24 min, truncated at 24-25 min). The XIC procedure allowed to selectively reveal the current associated to the specific ion pertaining to Hst1 (middle panel) and to non-phosphosphorylated Hst-1 (lower panel). XIC peak areas (MA: measured area) are proportional to peptide concentration and were utilized to calculate the percentages of the two proteoforms.[14-16] RT: retention time. NL: normalization level.

Figure 6

Percentages of non-phosphorylated histatin 1 measured in 230 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.

Figure 7

Percentages of non-phosphorylated (panel a), mono-phosphorylated (panel b) and di-phosphorylated

(panel c) statherin measured in 221 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.

Figure 8

Percentages of cystatin S (non-phosphorylated; panel a), cystatin S1 (mono-phosphorylated; panel b) and cystatin S2 (di-phosphorylated, panel c) measured in 124 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.

Figure 9

Percentages of non-phosphorylated S100A9 short (panel a) and of all the non-phosphorylated S100A9 long proteoforms (panel b) measured in 187 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.

Figure 10

Percentages of truncated PRP-3 type proteoforms with respect to the sum of entire (PRP-1 type) and truncated (PRP-3 type) proteoforms measured in 261 samples (panel a). Percentage of Hst5+Hst6 with respect to Hst3+Hst5+Hst6 measured in 192 samples (panel b). Percentage of Hst 5 with respect to Hst5+Hst6 measured in 188 samples (panel c). Percentage of short S100A9 (phosphorylated and non-phosphorylated) with respect to total S100A9 measured in 186 samples (panel d). All percentages are as a function of age (logarithmic scale). Other features are identical to those of Fig. 3.

Supplemental Files

Supplemental S1

Tune Method and Instrument Method parameters used for the high-resolution HPLC-MS/MS analyses.

Supplemental S2

Tune Method and Instrument Method parameters used for the low-resolution HPLC-MS analyses.

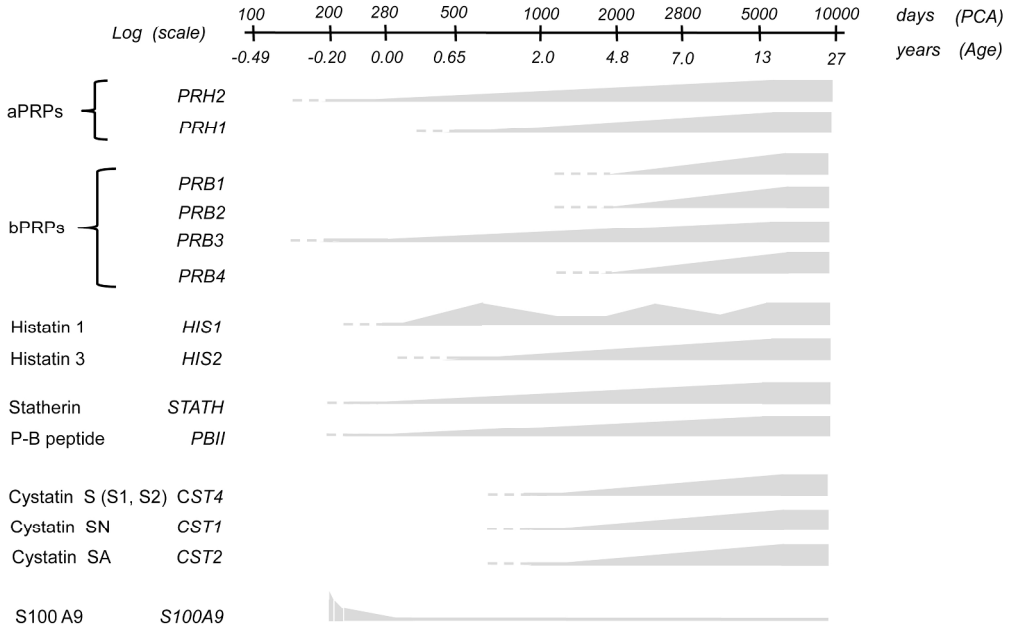
Supplemental S3

MS and MS/MS spectra assignment and characterization of salivary proteins. The "n fragment matching" is calculated as the ratio between the number of the experimental fragments y_n and b_n detected and the total principal y_n and b_n theoretical fragments. The percentage of matching considering also all the internal fragments and the possible neutral loss (fragment matching (internal fragments)) is also reported .

Supplemental S4

Molecular weight (theor), Elution time and *m/z* values utilized for the manual XIC label-free analyses.

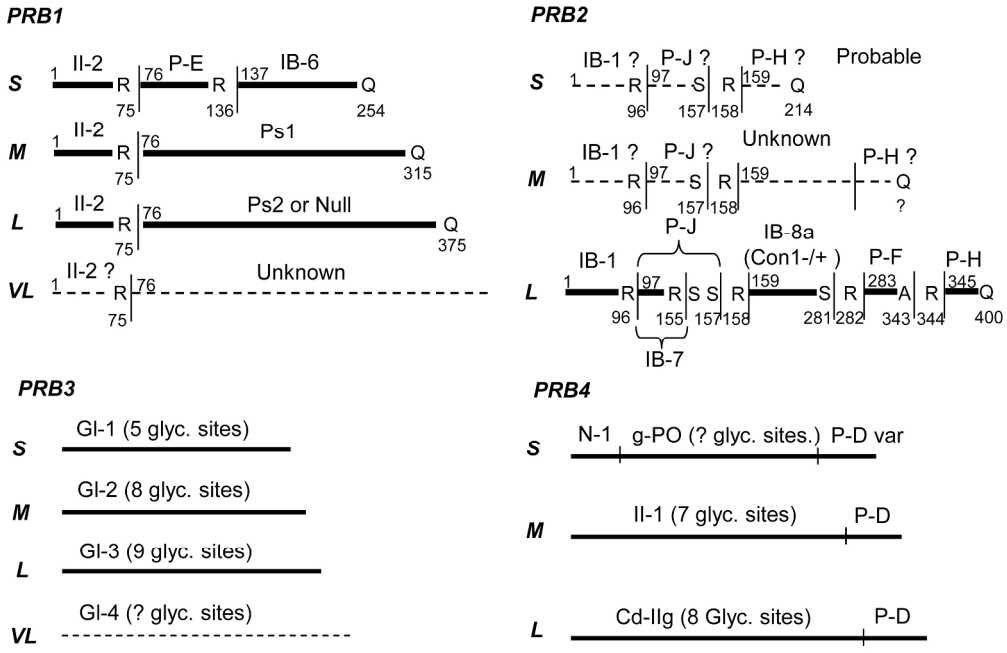
Figure 1



Scheme describing the approximate time course of the different salivary proteins and peptides investigated in this study. The sum of XIC peak areas of all the members belonging to the same family of proteins/peptides, grouped on the basis of their locus, is reported as function of age. A logarithmic scale has been used in the abscissa axis to better highlight the major changes occurring in preterm newborns during the first months of life, which correspond to the last months of fetal development, and those occurring immediately after the normal term of delivery and in the first years of life. Dashed lines represent the age range in which the proteins were detected in at least one baby. The upper limit of this range corresponds to the age in which saliva of all the babies displayed the protein in the chromatographic profile, even if at small amount.

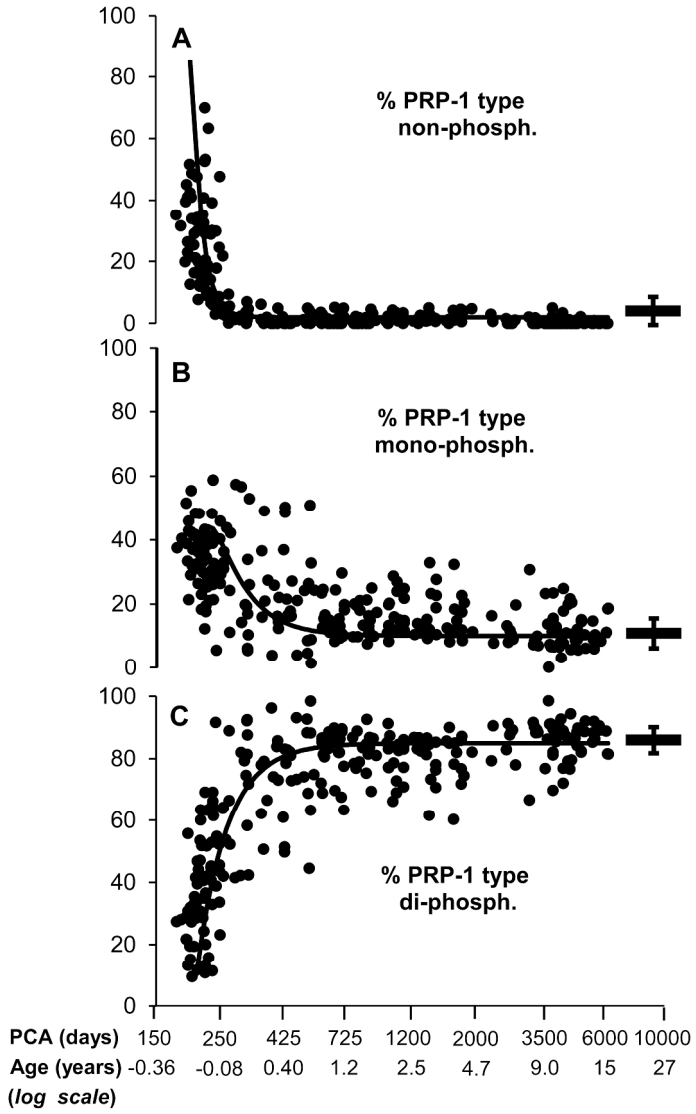
124x87mm (600 x 600 DPI)

Figure 2



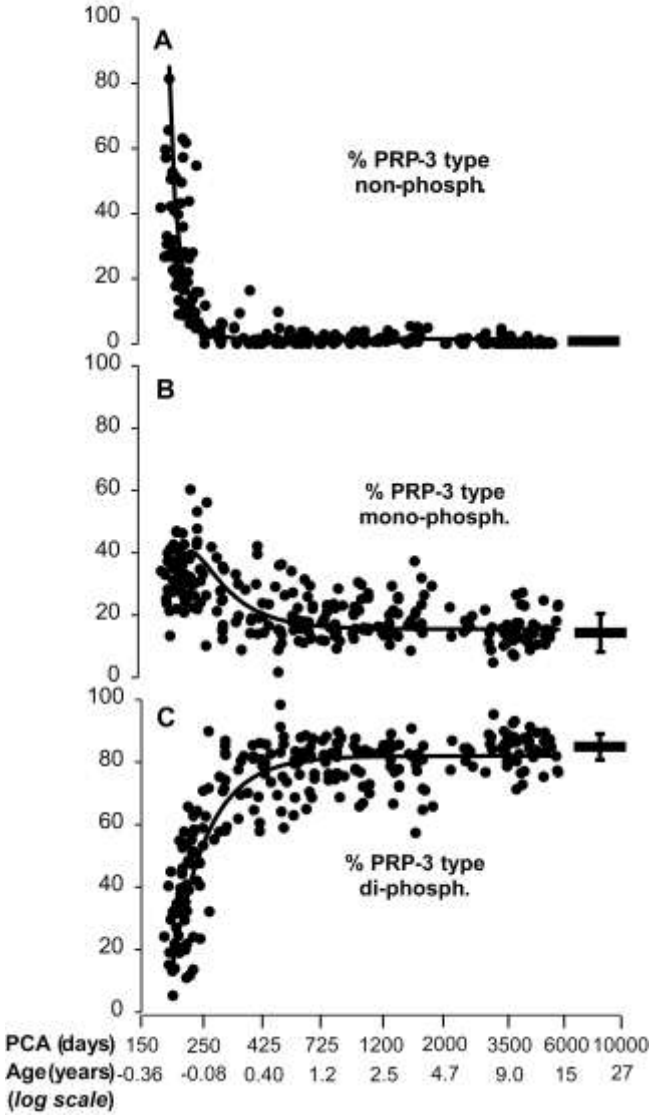
Known bPRPs loci and their phenotypes, according to the most recent mass spectrometry characterizations.13,19 The nomenclature of the different bPRPs was described by Castagnola et al.13
130x95mm (600 x 600 DPI)

Figure 3



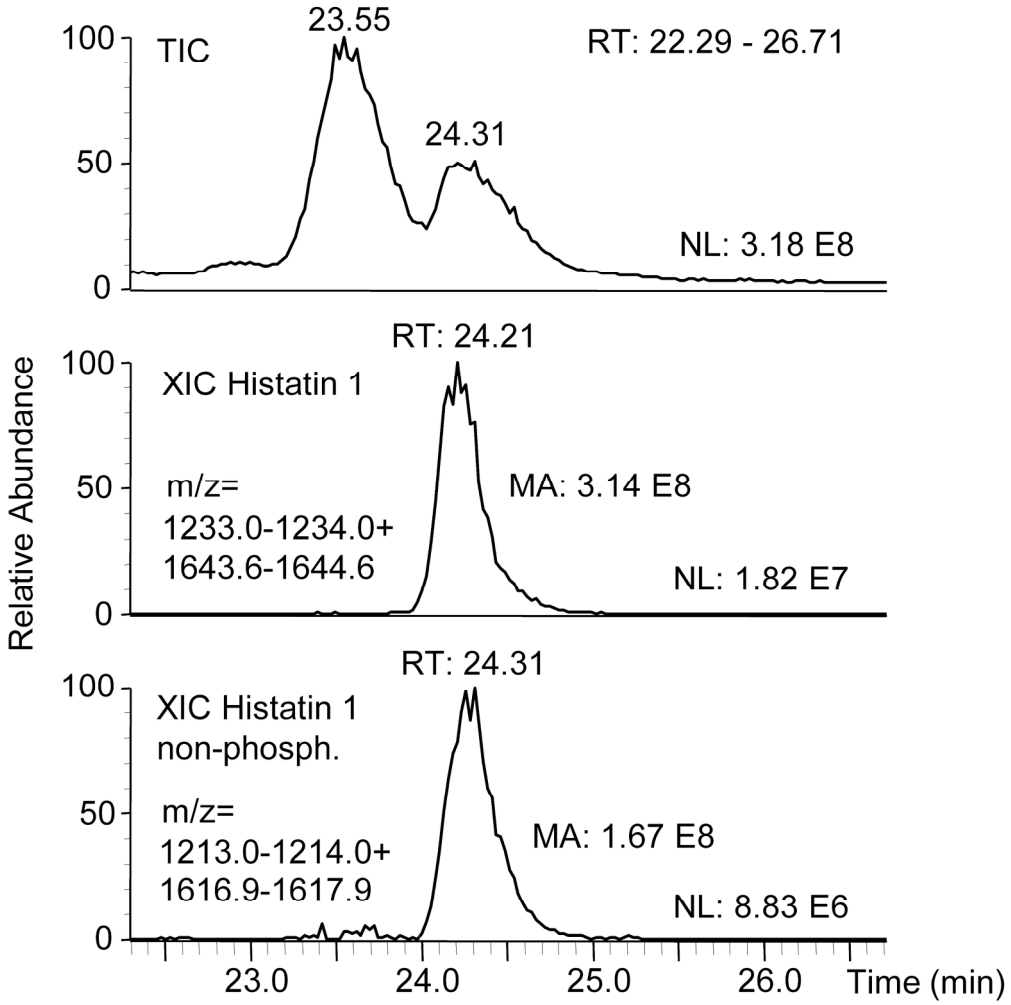
Percentages of non-phosphorylated (panel a), mono-phosphorylated (panel b) and di-phosphorylated (panel c) PRP-1 types aPRPs (comprehending PRP-1, PRP-2 and PIF-s) measured in 263 samples as a function of age (logarithmic scale). The trend lines were obtained by fitting experimental data to exponential functions. On the right side of the figure the mean value measured in the adults is reported at an arbitrary value of 9500 days of PCA.
133x234mm (600 x 600 DPI)

Figure 4



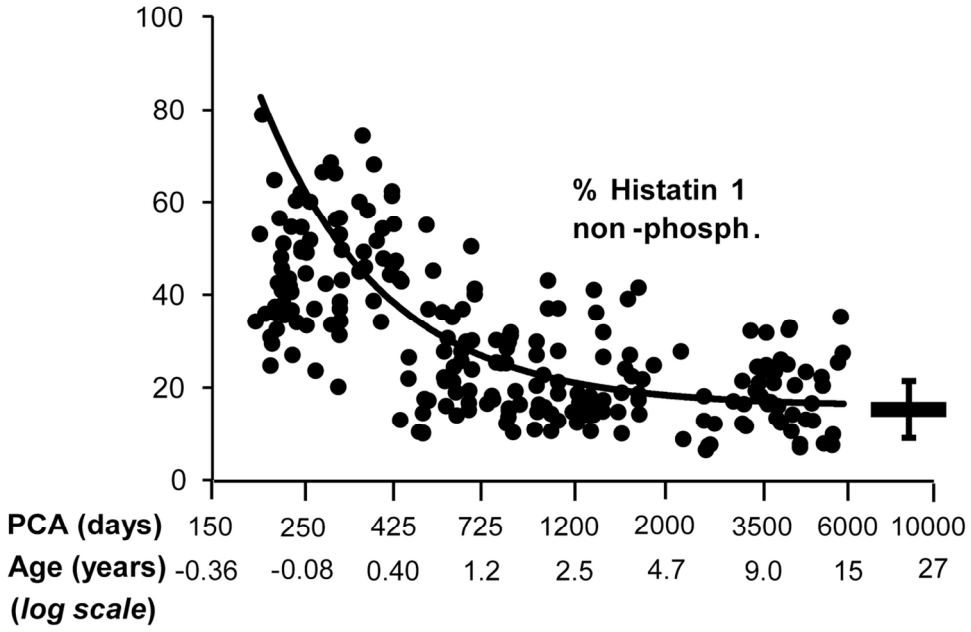
Percentages of non-phosphorylated (panel a), mono-phosphorylated (panel b), and di-phosphorylated (panel c) PRP-3 types aPRPs (comprehending PRP-3, PRP-4 and PIF-f) measured in 258 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.
135x252mm (600 x 600 DPI)

Figure 5



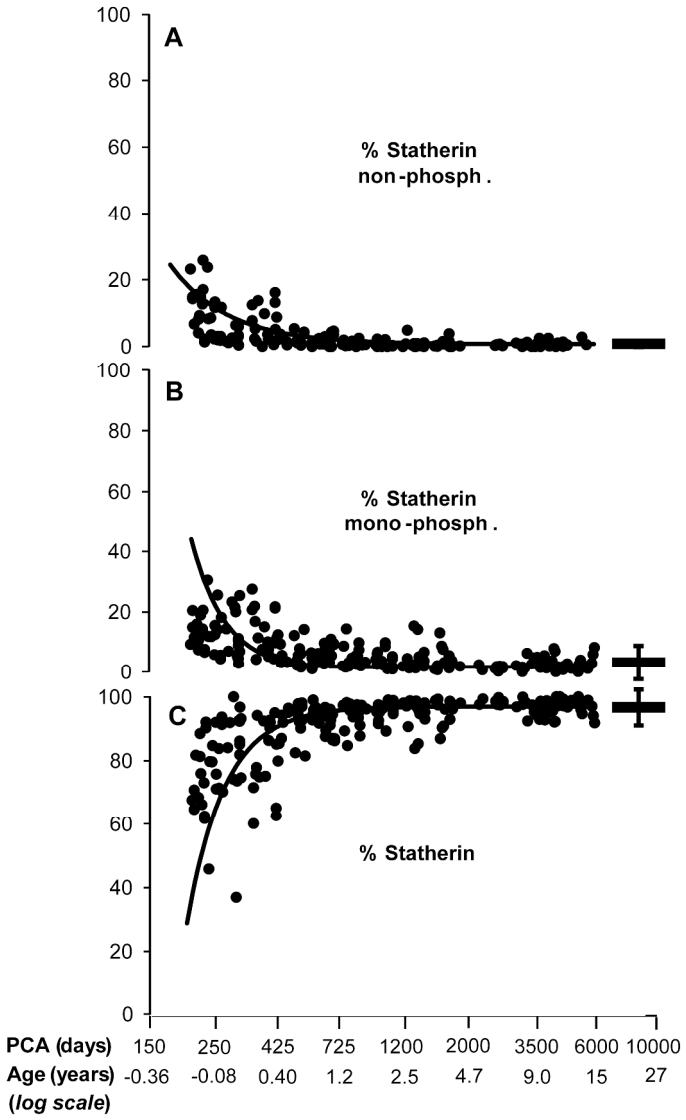
XIC procedure for the relative quantification of co-eluting proteins/peptides. The upper panel shows the HPLC-ESI-MS total ion current profile in the elution range 22.3-26.7 min. The two broad and intense peaks are mainly due to the highly charged aPRPs proteoforms (entire at 23-24 min, truncated at 24-25 min). The XIC procedure allowed to selectively reveal the current associated to the specific ion pertaining to Hst1 (middle panel) and to non-phosphosphorylated Hst-1 (lower panel). XIC peak areas (MA: measured area) are proportional to peptide concentration and were utilized to calculate the percentages of the two proteoforms.14-16 RT: retention time. NL: normalization level.
98x113mm (600 x 600 DPI)

Figure 6



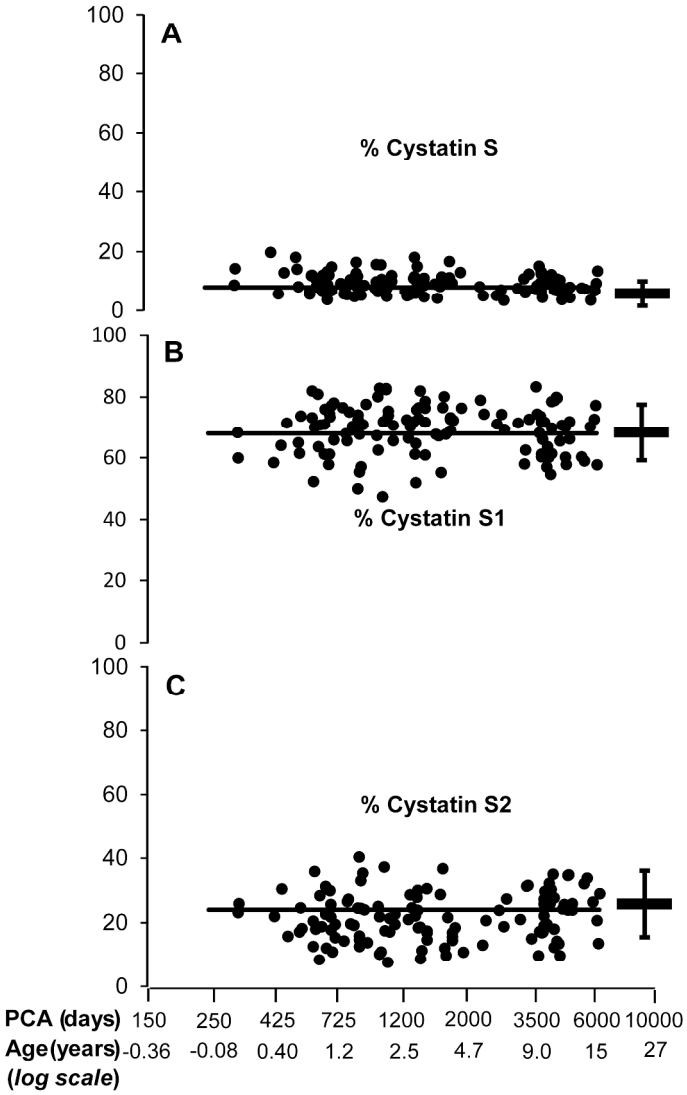
Percentages of non-phosphorylated histatin 1 measured in 230 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.
61x44mm (600 x 600 DPI)

Figure 7



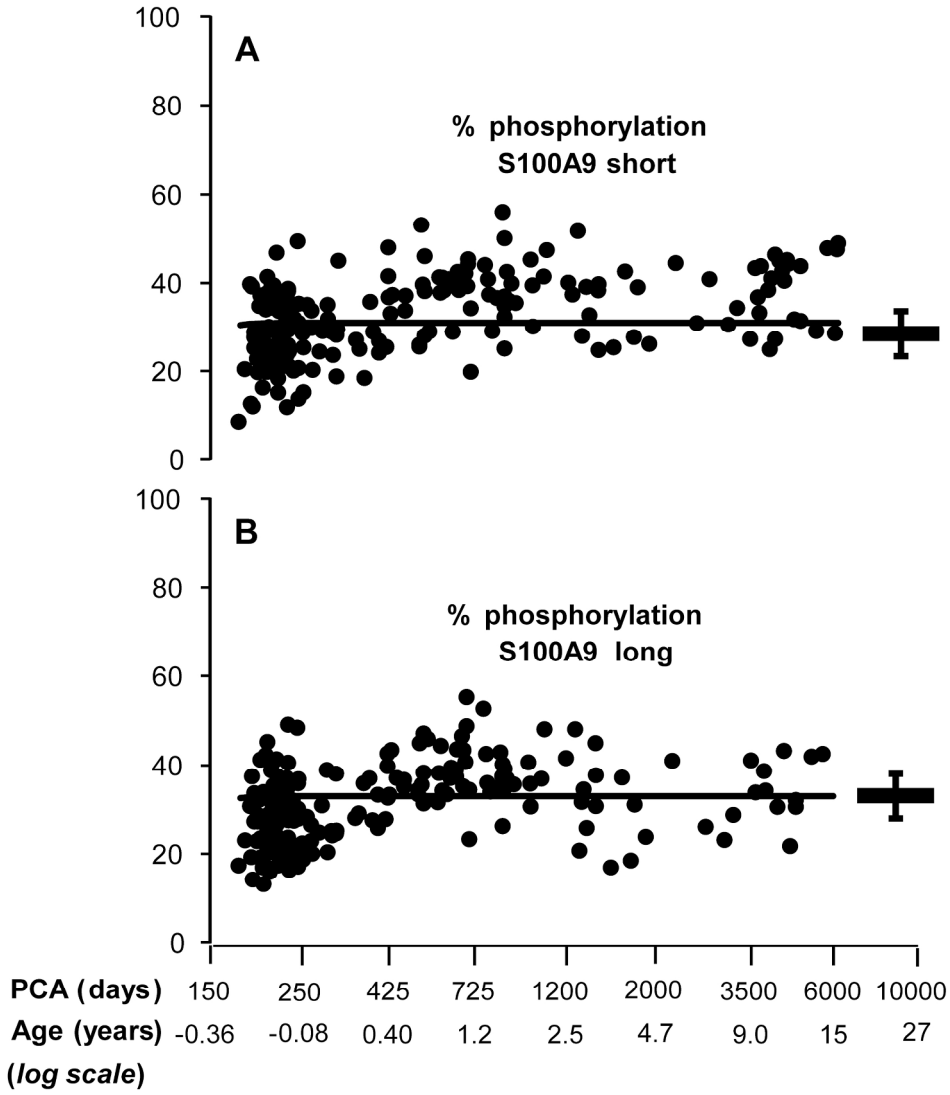
Percentages of non-phosphorylated (panel a), mono-phosphorylated (panel b) and di-phosphorylated (panel c) statherin measured in 221 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.
142x241mm (600 x 600 DPI)

Figure 8



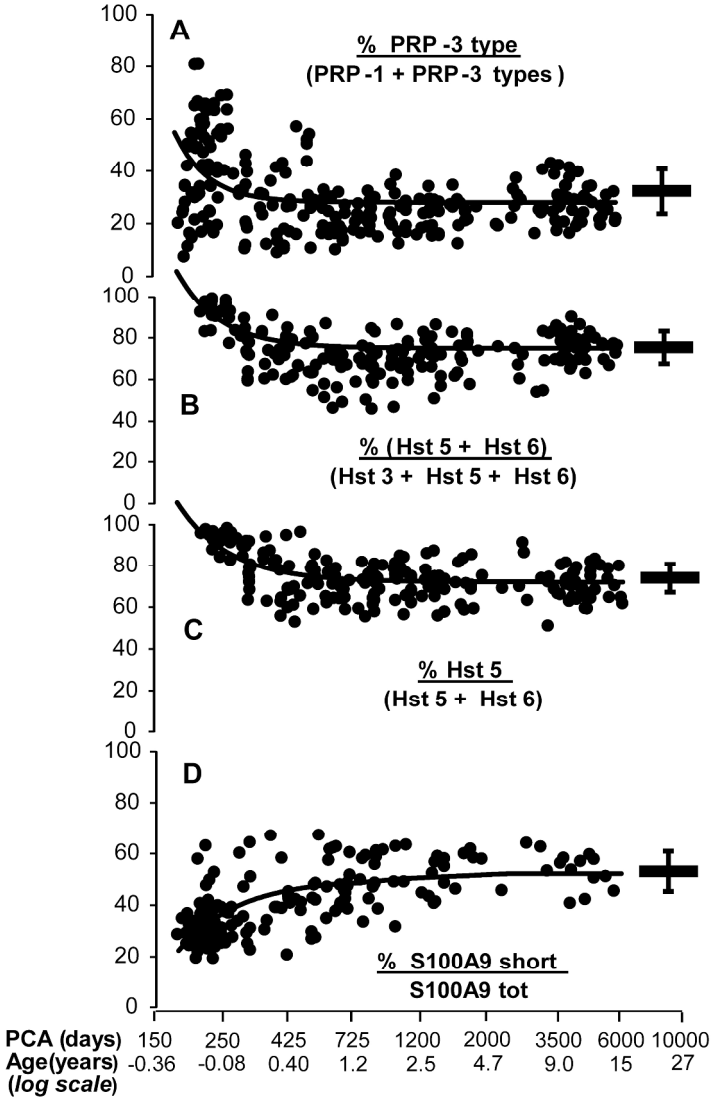
Percentages of cystatin S (non-phosphorylated; panel a), cystatin S1 (mono-phosphorylated; panel b) and cystatin S2 (di-phosphorylated, panel c) measured in 124 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.
129x198mm (600 x 600 DPI)

Figure 9

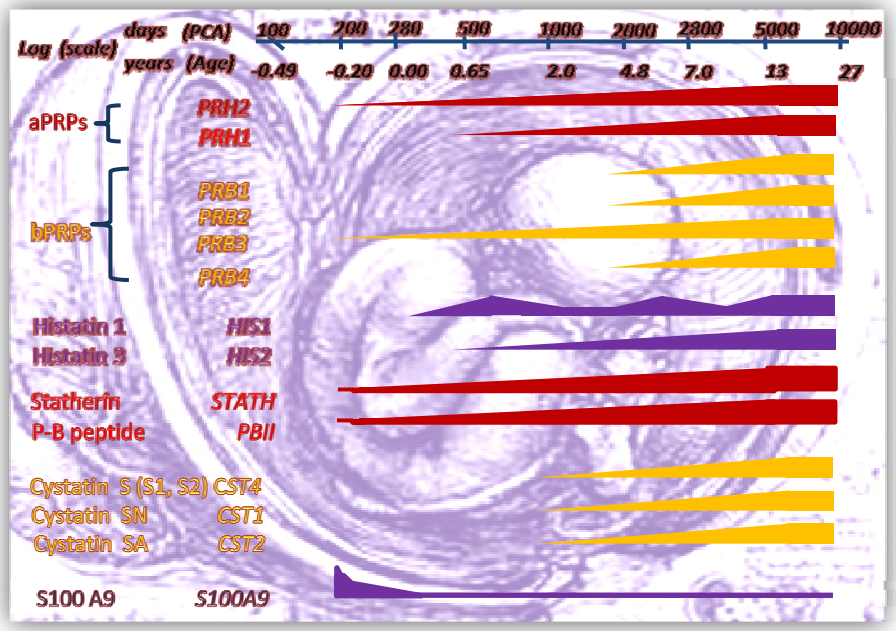


Percentages of non-phosphorylated S100A9 short (panel a) and of all the non-phosphorylated S100A9 long proteoforms (panel b) measured in 187 samples as a function of age (logarithmic scale). Figure features are identical to those of Fig. 3.
106x135mm (600 x 600 DPI)

Figure 10



Percentages of truncated PRP-3 type proteoforms with respect to the sum of entire (PRP-1 type) and truncated (PRP-3 type) proteoforms measured in 261 samples (panel a). Percentage of Hst5+Hst6 with respect to Hst3+Hst5+Hst6 measured in 192 samples (panel b). Percentage of Hst 5 with respect to Hst5+Hst6 measured in 188 samples (panel c). Percentage of short S100A9 (phosphorylated and non-phosphorylated) with respect to total S100A9 measured in 186 samples (panel d). All percentages are as a function of age (logarithmic scale). Other features are identical to those of Fig. 3.
129x201mm (600 x 600 DPI)


Log (scale)
days (PCA)
years (Age)
100
200
280
500
1000
2000
2800
5000
10000
-0.49
-0.20
0.00
0.65
2.0
4.8
7.0
13
27
aPRPs
PRH2
PRH1
bPRPs
PRB1
PRB2
PRB3
PRB4
Histatin 1
Histatin 3
HIS1
HIS2
Statherin
P-B peptide
STATH
PBII
Cystatin S (S1, S2)
Cystatin SN
Cystatin SA
CST4
CST1
CST2
S100 A9
S100A9