Chrono-proteomics of human saliva: variations of the salivary proteome during human development investigated by a top-down platform.

Journal:	*Journal of Proteome Research*
Manuscript ID:	Draft
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

Chrono-proteomics of human saliva: variations of the salivary proteome during human development investigated by a top-down platform.

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

The salivary proteome has been investigated in 293 subjects over a wide range of ages. Proteins typically found in adult saliva became detectable at different times during development. Acidic proline-rich proteins expressed by *PRH2* locus and a glycosylated basic proline-rich protein expressed by *PRB3* locus appeared before seven months of post-conceptional age, followed at about seven months by histatin 1, statherin and P-B peptide. The other histatins and acidic proline-rich proteins appeared after the normal term of delivery, S type cystatins at about one year and basic proline-rich proteins around 4-5 years of age. All the proteinases involved in the maturation of salivary peptides were found more active during fetal growth, The activity of the Fam20C kinase, involved in phosphorylation of various proteins started its expression around 180 days of post-conceptional age, slowly increased and reached values comparable to adults at about two years of age. Instead, the MAPK14 involved in phosphorylation of S100A9 was found fully active till from the last months of fetal growth. This study establishes reference ranges of salivary proteins during the first years of life, during childhood through early adulthood. Deviations could provide indication of abnormal activities of enzymes involved in the development.

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

1. Introduction.

One of the most challenging topics of proteomics is to define the temporal changes of a specific proteome. This effort opens in proteomic studies a new dimension that can be termed chrono-proteomics. Our group has been engaged in the study of the salivary proteome with an integrated top-down/bottom-up platform for several years.[1] Human saliva is a bodily fluid characterized by an easy and non-invasive specimen collection.[2] Sampling can be performed without provoking any pain and stress to the donor and therefore it offers the opportunity to easily investigate its composition also in pediatric age. Taking into advantage this property, in the last years our group has investigated the proteome of preterm newborns and subjects in the pediatric age highlighting striking differences with respect to that of adults.[1,3,4] Saliva of human preterm newborns contains appreciable amounts of more than forty proteins that are either undetectable (according to the limit of detection of our analytical method) or detectable in minute quantity in saliva of adults.[1] On the contrary, the proteins typically revealed in adult saliva are undetectable or detectable at low amount in preterm newborn saliva. For this reason, the aim of this study was to investigate the variations of concentration of several adult salivary proteins in a large group of subjects over a wide range of ages, starting from preterm newborns with a post-conceptional age (PCA) of about 180 days up to 18 years-old subjects and compare the results with the levels measured in the adults. In addition, the determination of several post-translational derived proteoforms as a function of age has been carried out in order to highlight the activation of different enzymes, such as kinases and convertases, during human development. This information could be of great help in pediatric clinic in order to establish the reference ranges during

childhood through early adulthood. Deviations from the established reference ranges could provide an early indication of abnormal functions of enzymes involved in the development and thus be clue of potential diseases.[5]

2. Experimental section

2.1 Reagents and instruments

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). The HPLC-ESI-IT-MS apparatus was a Surveyor HPLC system (ThermoFisher Scientific, San Jose, CA, USA) 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. The MS apparatus was equipped with an electrospray ionization (ESI) source. The chromatographic column was a 150 × 2.1 mm Vydac (Hesperia, CA, USA) C8 column, with 5-µm particle diameter.

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 it has therefore been performed in accordance with the ethical standards laid down in the

1964 Declaration of Helsinki. All rules were respected 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 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 established 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 samples from 182 donors (82 males, 90 females) with an age comprised between 0 to 17 years of age.

c) 23 samples from 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 TFA in an ice bath. The solution was then centrifuged at 8000 g for 5 min (4 °C). Finally, the acidic supernatant was separated from the precipitate and either immediately analyzed with an HPLC-ESI-MS apparatus (100 μL of inject volume, corresponding to 50 μL of saliva) or stored at -80 °C until the analysis.

2.3 RP-HPLC-ESI-MS analysis

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 of saliva 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. Mass spectra were collected every 3 ms in the positive ionization mode. The MS spray voltage was 5.0 kV, and the capillary temperature was 260 °C.

2.4 Characterization and quantification

Salivary peptide and protein quantitative analysis was based on the measurement of the RP-HPLC-ESI-MS eXtracted Ion Current (XIC) peak area (signal/noise ratio > 5). The XIC layout reveals the peak associated with the protein of interest by searching along the total ion current (TIC) chromatographic profile the specific multiply-charged ions generated by the protein at the source. The ions used to quantify the proteins and peptides were carefully selected to exclude values in common with other co-eluting proteins. The area of the XIC peak was related to the saliva volume (1.0 mL) and it is proportional to protein/peptide concentration. Under constant analytical conditions, it can be used for a quantitative analysis and comparative studies.[6-8] The different salivary proteins and their proteoforms investigated have been previously characterized by means of an integrated top-down bottom-up proteomic platform.[1] The *m/z* values utilized for the determination of the (XIC peak

area)/(mL of saliva) of acidic and basic proline-rich proteins, histatins, statherin, P-B peptide and "S-type" cystatins were reported in our previous study.[3] 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.[9]

2.5 MS data analysis and statistics.

Deconvolution of averaged ESI-MS spectra was automatically performed by using MagTran 1.0 software.[10] Pearson *r* coefficient was used to evaluate linear correlations. Differences with p value < 0.05 were considered statistically significant.

3. Results

3.1 Expression of human salivary proteins in the pediatric age.

HPLC-ESI-MS investigation of the acidic soluble fraction of 293 whole human saliva samples collected from preterm newborns and from subjects aged 0-17 years highlighted a very high inter-individual variability. However, as shown in Figure 1, just by performing a XIC search in the TIC-MS profile it is possible to have a rough indication on the activation of the different loci responsible for the expression of the salivary peptides and proteins. It is relevant to outline that with the aim 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, a logarithmic scale has been used in the abscissa axis of Figure 1, as well as in all the figures displaying temporal changes.

3.1.1 Acidic proline-rich proteins

Acidic proline-rich proteins (aPRPs) are a family of salivary proteins expressed by *PRH1* and *PRH2* loci located on chromosome 12p13.2. In the majority of the western population aPRPs comprise five major isoforms: PIF-s, Db-s and Pa (commonly detected as 2-mer), expressed by the *PRH1* locus, PRP-1 and PRP-2, expressed by the *PRH2* locus.[10-12] 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. The proteoforms with Mav 15514-15515 Da were detected, although at very low amount, in all the samples of preterm newborns since 180-190 days of PCA, while Pa 2-mer and Db-s (where s stands for slow) were never detectable in preterm newborns before 270-280 days of PCA (corresponding to the normal term of delivery). For this reason it can be supposed that the expression of *PRH2* locus starts during fetal life, slightly before 180 days of PCA, and that of *PRH1* locus 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, detectable only in parotid secretion. They are the product of four loci (*PRB1-PRB4*) located on chromosome 12p13.2 closely to the *PRH1-2* loci. At least four common alleles (S, M, L, VL) are present at *PRB1* and *PRB3*

loci, four (S, M. L(Con1+), L(Con1-)) at *PRB2* locus and three (S, M. L) at *PRB4* locus in the western population.[3,13-17] All the bPRPs are only detectable as multiple peptide fragments deriving from the corresponding pro-proteins and, as shown in Figure 2, the connection between the most common haplotypes and salivary phenotypes is still pending for a complete definition. Broad peaks eluting in the bPRPs chromatographic cluster were attributed to glycosylated basic proline-rich proteins (gPRPs) on the basis of the crowded ESI spectra.

The expression of bPRPs occurs later with respect to other salivary proteins. Indeed, bPRPs peptides were undetectable or detected at very low amount even in saliva of 3-6 years old children. They increased with age reaching levels comparable to adults only after puberty, with the exception of a glycosylated bPRP, under characterization upon our laboratories, which was surprisingly detected in minute amounts in preterm saliva since 180 days of PCA. Due to the absence of PD peptide, which is expressed by *PRB4* locus (see Fig. 2), it can be assumed that this protein is a product of *PRB3* locus. This finding highlights that, differently from other bPRPs, the expression of *PRB3* locus starts during the last months of fetal life (Fig. 1).

3.1.3 Histatins

The *HIS1(*or *HTN1)* and *HIS2* (or *HTN3*) loci, responsible for the expression of histatin 1 (Hst1) and histatin 3 (Hst3), respectively, are closely located on chromosome 4q13.3.[16] From our data *HIS1* expression started around 200-210 days of PCA and as already reported in a recent study of our group Hst1 showed two maximum concentration levels around 1 and 6 years of age which could be linked to deciduous and permanent dentitions.[4] Concerning Hst3 and its proteolytic derivatives histatin 6 (Hst6)

and histatin 5 (Hst5) our observations highlight that *HIS2* expression commonly starts later than Hst1 and roughly immediately after the normal term of delivery (270-280 days of PCA).

3.1.4 Statherin and P-B peptide

STATH and *PBII* genes encode statherin and P-B peptide (also called submaxillary gland androgen-regulated protein 3B), respectively.[15] In several preterm newborns statherin and P-B peptide were detected since about 220 days of PCA and the increase in concentration with age showed a similar trend. In some preterm newborns, both peptides were detected in saliva 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 connected more to statherin than to bPRPs.[15]

3.1.5 "S-type" cystatins

Cystatin S (S1,S2), SN and SA are called "S type" cystatins and are expressed by three loci closely located on chromosome 20p.11.21.[17] In previous studies it has been shown that they are mainly secreted by submandibular/sublingual glands.[1,17] "S type" cystatins, undetectable in preterm newborn saliva, became detectable in whole saliva from approximately one year of age.

3.1.6 S100A9

S100A9 is expressed by a locus located on chromosome 1q21 and showed features opposite to the other proteins investigated in this study. It is representative of the proteins found at high concentration in preterm newborn saliva[1] decreasing during fetal development as a function of PCA and becoming detectable at low level or undetectable after the normal term of delivery.

3.2 Post-translational modifications.

Variations of the relative concentration of several proteoforms of salivary proteins during human development can disclose precious information on the timing of activation of the different enzymes involved in the post-translational modifications (PTMs) occurring during saliva secretion.

3.2.1 Phosphorylation

Many phosphorylated proteins and related kinase activities have been detected in whole saliva by bottom-up proteomic studies.[18-20] The different phosphorylation sites detected on salivary proteins and peptides by means of the top-down HPLC-MS platform utilized 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, which are commonly di-phosphorylated. However, very small amounts of a three-phosphorylated proteoform (on Ser_{17}) as well as

mono-phosphorylated and non-phosphorylated proteoforms were also detected.[11] Our group already reported[21] that the percentages of non-phosphorylated and mono-phosphorylated aPRPs decrease as a function of PCA in preterm newborns. In this study the investigation was extended to a greater number of subjects. Figure 3 shows the percentages of the different phosphorylated proteoforms of entire aPRPs (PRP-1 types, including PRP-1, PRP-2 and PIF-s) measured in 263 samples. The panels a, b and c show the percentage of non-phosphorylated, mono-phosphorylated and di-phosphorylated aPRPs, respectively.

The trend lines on Fig.3a and 3c represent a fit obtained by exponential functions, while the trend line of Fig. 3b was obtained by the differences of the fits of Fig.s 3a and 3c. On the right part of the figures at an arbitrary value of 9500 days of PCA (corresponding to about 26 years of age) the mean value measured on the adults is reported. PRP-1 type aPRPs are all cleaved at the Arg_{106} residue 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. Data confirmed previous observations[21] and showed that the kinase involved in the phosphorylation of aPRPS is not fully active even at birth, reaching the activity measured in the adults around two years of age.

3.2.1.2 Histatin 1

Hst1 is mono-phosphorylated on Ser_2. As shown in Fig. 5, the non-phosphorylated proteoform was always detectable in the HPLC-ESI-MS close to Hst1.[14,22] Fig. 6 reports the percentage of the non-phosphorylated Hst1 proteoform measured as a function of the logarithm of PCA. Percentages

were determined only in 230 samples because Hst1 was not detectable in most of the premature infants with lower PCA. Fig. 6 shows that, even Hst1 becomes detectable approximately two weeks later than aPRPs, the activation timing of the kinase responsible for the phosphorylation of Hst1 is very similar to that observed for aPRPs.

3.2.1.3 Statherin

Statherin is di-phosphorylated on Ser_1 and Ser_2. In the HPLC-ESI-MS profile, the mono-phosphorylated and the non-phosphorylated proteoforms of statherin are detectable just before the peak of statherin.[15,23] Fig.7 shows the percentages of non- (a), mono- (b) and di-phosphorylated statherin (c) measured in 217 samples as a function of the logarithm of PCA. Fig. 7 shows that the trend of phosphorylation of statherin during human development is very similar to that observed for aPRPs and Hst1. The phosphorylation of statherin reaches the percentages observed in adult saliva around two years of age (1000 days of PCA).

3.2.1.4 Cystatin S, S1, S2

Cystatin S is a non-phosphorylated protein, while cystatin S1 is phosphorylated on Ser_3 and cystatin S2 is phosphorylated on Ser_1 and Ser_3 residues.[17] Fig.s 8a,b,c show the percentages of the three proteoforms as a function of the logarithm of PCA measured in 124 samples. It is evident that the percentages of phosphorylation of these cystatins do not change during the development. The mean percentage values were 6 ± 4, 68 ± 9 and 26 ± 10 for cystatin S, S1 and S2, respectively.

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, sharing very similar sequences and both phosphorylated on Ser_8, closely eluting in the HPLC-ESI-MS profile.[13] These bPRPs are undetectable in whole saliva before 4-6 years of age. The non-phosphorylated proteoforms were detected sporadically, with a frequency and a percentage comparable to those observable in adult saliva (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 S100A9 long, all N-terminally acetylated.[10] 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. 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 before 190 days of PCA.

3.2.2 Endo- and exo-peptidases

Different endo- and exo- peptidases are active during granule storage and secretion of salivary proteins.

3.2.2.1 Acidic proline-rich proteins

Among aPRPs, PRP-1, PRP-2, PIF-s and Db-s are cleaved by the action of a convertase that recognizes the consensus sequence RXXR↓ at the 106 Arg residue (127 Arg residue in Db-s) generating four truncated proteoforms called PRP-3, PRP-4, PIF-f and Db-f and a peptide of 44 amino acid residues common to all these proteoforms, called P-C peptide.[11,12,21] The Pa proteoform, due to the substitution Arg→Cys at the 103 residue that abolishes the RXXR↓ consensus sequence, is not cleaved. It is commonly detected in whole saliva as S-S 2-mer.[11] 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. Fig.10a shows the percentage of truncated PRP-3 type proteoforms computed with respect to the total of PRP-1 and PRP-3 type proteoforms measured in 261 samples as a function of the logarithm of PCA. Because this percentage is indicative of the activity of the convertase responsible for aPRPs cleavage, data of Fig. 10a confirm that this enzyme is more active during fetal growth, as previously supposed by our group.[21]

3.2.2.2 Histatins 3, 5 and 6

Hst3, expressed by the *HIS2* gene, as described in Section 3.2.3, is a small histidine-rich peptide of 32

amino acid residues partially cleaved before granule storage at the Arg_{25} residue by a convertase generating Hst6 (Hst3 fr. 1/25).[22] The subsequent removal of the Arg_{25} by a carboxy-peptidase generates Hst5 (Hst3 fr. 1/24). To evaluate the convertase activity, the percentage of Hst5+Hst6 were computed with respect of total Hst3+Hst5+Hst6. The data determined in 192 samples are reported in Fig. 10b as a function of the logarithm of PCA. To evaluate the carboxy-peptidase activity, the percentage of Hst 5 was computed with respect of total Hst5+Hst6. The data determined in 188 samples are reported in Fig.10c as a function of the logarithm of PCA. From the Figures it is evident that the convertase and the carboxy-peptidase activities are higher in pre-term newborns and decrease according to increased PCA reaching the activities characteristic of the adults roughly after the normal term of delivery.

3.2.2.3 S100A9 short and long

As reported in Section 3.3.1.6 the proteoforms of S100A9 detectable in whole saliva are the short one, and different derivatives of the long one.[10] The plot of the percentage of short S100A9 (phosphorylated and non-phosphorylated) with respect to total S100A9 measured in 186 samples and reported in Fig. 10d shows that the short form slightly increase 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 evidence the non acetylated S100A9 proteoforms even in preterm newborns with less than

190 days of PCA, showing that the N-terminal acetyltransferase (NAT) responsible for the PTM is fully active at 6 months of fetal life.

3.2.4 Sulfation

In whole human saliva it is possible to detect Hst1 derivatives showing mass increases of 80, 160, 240 and 320 Da corresponding to proteoforms having from 1 to 4 sulfo-tyrosines. These derivatives represent a small percentage of Hst1 (around 10% of total Hst1) and this PTM is specific of submandibular/sublingual gland.[24] Sulfo-derivatives of Hst1 appear in whole saliva during the first year of life. They were detected in some newborns few weeks after birth, but were absent in other subjects till to one year of age. After one year the sulfo-derivatives were always present in all the subjects under study. Therefore, the sulfotransferase responsible for this PTM is not active during fetal life and its activation occurs during the first year of life, according to the characteristic of each subject.

4 Discussion

4.1 General considerations

Top-down proteomic platform and label free quantification applied to the study of intact salivary proteoforms allowed to obtain a lot of information on the changes occurring during human development. Whole saliva of preterm newborns contains a lot of proteins characteristic of the fetal growth in part characterized by our group in a previous work, many of them pertaining to the families

of S100 proteins, thymosins and stefins.[1] The concentration of these proteins drops quickly and they almost disappear at the normal term of delivery. In the meantime proteins typically found in adult saliva became detectable, but at different times. aPRPs expressed by *PRH2* locus and a glycosylated bPRP expressed by *PRB3* locus appear first in whole saliva during the last months of fetal life before 180 days of PCA. They are followed by Hst1, statherin and P-B peptide at about seven months of PCA. Hst3, Hst5, Hst6 and aPRPs encoded by *PRH1* locus appear after the normal term of delivery. "S type" cystatins appear at about one year of age. bPRPs expressed by *PRB1, PRB2*, and *PRB4* loci have been commonly detected not before 4-5 years of age. This hierarchical expression of salivary proteins during human development should have a functional meaning. For instance, it has been suggested that the increase of concentration of Hst1 in some periods of infancy could be linked to the wound healing properties of this peptide.[25,26] The late appearance of bPRPs, more than linked to the ability to protect oral cavity against the harmful effects of tannins,[27] could be connected to variations of taste perception[28,29] 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,[30] apart from the great inter- and intra-individual variability in the adult.[31] 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 among 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 trends in the increase of phosphorylation (see Fig.s 3,4,6) and the common and characteristic SxE/pS consensus sequence supported the hypothesis that aPRPs, Hst1 and statherin are all substrates of a unique kinase, working before secretion. On the basis of its highly specific consensus sequence[32] and localization in the Golgi apparatus of granule secreting cells, it was firstly defined as Golgi casein kinase and for long time escaped any attempt of purification and structural characterization. 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.[33]

The results of this study indicated 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 roughly at two years of age. Fam20C is a pleiotropic enzyme responsible for the phosphorylation of a lot of different substrates in many organs and tissues, including central nervous system [34]. Mutations in Fam20C gene in humans cause Raine syndrome an extremely rare autosomal recessive osteosclerotic bone dysplasia.[35] 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.[36] In a recent study of our group significant hypo-phosphorylation of aPRPs, statherin and Hst1 in a subgroup of patients with autism spectrum disorders have been observed.[37] The crucial role of Fam20C indicates that it has to be strictly synchronized with other enzymes and proteins relevant during development, and a small delay of activation during fetal growth could generate severe neurological impairments.[38]

On the basis of the consensus sequence, also cystatin S and the bPRPs IB-1 and II-2 peptides should be substrates of Fam20C. However, these proteins appear in human saliva later when the enzyme is fully

active.

S100A9, instead, is phosphorylated by the MAPK14, a kinase pertaining to the p38 mitogen-activated protein kinase pathway.[39] 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.[16] The principal cleavage consensus sequences observed are potentially recognizable by enzymes pertaining to the furin-like convertase families.[6,16] Although the cleavage consensus sequences of cleavage are very similar, it is not known if a common unique convertase acts on the different families of salivary proteins or rather different enzymes are specifically devoted to the cleavage of each. bPRP are completely cleaved before granule storage and the pro-protein is not detectable, while aPRPs and Hst3 are partially cleaved and the intact proteoforms are always detectable in high percentage in human saliva. Two studies addressed the partial characterization of the aPRPs and bPRPs proteinases.[40,41] A study of Basak and colleagues[42] 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 obtained 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 for the cleavage of

aPRPs.[21] 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.[43] In salivary glands this activity could be ascribed to the metalloprotease carboxypeptidase Z[44], 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 regeneration.

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 excision of the initiator methionine after protein synthesis.[45] A particular N-terminal structural arrangement could induce MAP to cleave the fifth instead of the first methionine residue of the protein chain. Figure 12 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).[46] On the basis on the S100A9 N-terminal residues, a threonine in the long proteoforms and a serine in the short proteoforms, the enzyme should be NATa.[46] From the present data, and from our previous studies[2] 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 post-translational modification (PTM) implicated in various cellular and extra-cellular processes.[47] 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.[48] Hst1 has been the first sulfo-peptide detected in human saliva.[25] The residues involved in the sulfation of Hst1 are tyrosines 27, 30, 34, and 36 and the PTM is probably hierarchical. The sequence surrounding Tyr_{27} shows indeed several features required for the 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.[49] Therefore, it can be proposed that the O-sulfation of Tyr_{27} in Hst1 activates sequentially the sulfation of other three Tyr residues. The results of this study indicate that TPSTs activity in salivary glands starts during the first year of life. It could be of great

interest to establish in the future if this delay in the activation of TPSTs concerns other fetal organs and tissues, or rather is restricted only to the 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".

The authors declare no competing financial interest

6. References

(1) 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.

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

(3) 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.

(4) 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.

(5) Castagnola, M.; Cabras, T.; Vitali, A.; Sanna, M. T.; Messana, I. Biotechnological implications of the salivary proteome. *Trends Biotechnol.* **2011**, *29*, 409-418.

(6) 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.

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

(8) 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.

(9) 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.

(10) Zhang, Z. and 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.

(11) 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.

(12) Iavarone, F.; D'Alessandro, A.; Tian, N.; Cabras, T.; Messana, I.; Helmerhorst, E. J.; Oppenheim, F. G; Castagnola, M. High-resolution high-performance liquid chromatography with electrospray

ionization mass spectrometry and tandem mass spectrometry characterization of a new isoform of human salivary acidic proline-rich proteins named Roma-Boston Ser_{22} (Phos) → Phe variant. *J. Sep. Sci.* **2014**, *37*, 1896-1902.

(13) 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.

(14) Cabras, T.; Boi, R.; Pisano, E.; Iavarone, F.; Fanali, C.; Nemolato, S.; Faa, G.; Castagnola, M.; Messana, I. HPLC-ESI-MS and MS/MS structural characterization of multifucosylated N-glycoforms of the basic proline-rich protein IB-8a CON1(+) in human saliva. *J. Sep. Sci.* **2012**, *35*, 1079-1086.

(15) 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.

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

(17) 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.

(18) 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.

(19) 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.

(20) 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.

(21) 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.

(22) Castagnola, M.; Inzitari, R.; Rossetti, D. V.; Olmi, C.; Cabras, T.; Piras, V.; Nicolussi, P.; Sanna, M. T.; Pellegrini, M.; Giardina, B.; Messana, I. A cascade of 24 histatins (histatin 3 fragments) in human saliva: suggestions for a pre-secretory sequential cleavage pathway. *J. Biol. Chem.* **2004**, *279*, 41436-41443.

(23) 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 postsecretory events responsible for the formation of secreted human salivary peptides: a proteomics approach. *Mol. Cell. Proteomics* **2008**, *7*, 911-926.

(24) 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.

(25) 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.

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

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

(28) 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.

(29) Melis, M.; Aragoni, M. C.; Arca, M.; Cabras, T.; Caltagirone, C.; Castagnola, M.; Crnjar, R.; Messana, I.; Tepper, B. J.; Tomassini Barbarossa, I. Marked increase in PROP taste responsiveness following oral supplementation with selected salivary proteins or their related free amino acids. *PLoS One* **2013**, *8*, e59810.

(30) 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.

(31) 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.

(32) 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.

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

(34) 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.

(35) 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.

(36) 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.

(37) 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.

(38) 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.

(39) 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.

(40) 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.

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

(42) 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.

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

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

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

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

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

(48) 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.

(49) 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 expression timing of the different salivary proteins and peptides investigated in this study. 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.

Figure 2

Known bPRPs loci and their phenotypes, according to the most recent mass spectrometry characterizations [3,13-17].

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 the logarithm of PCA (or age). The trend lines on panel a and c represent a fit obtained by exponential functions, while the trend line of panel b was obtained by the differences of the fits of panels a and c. On the right of the figures at an arbitrary value of 9500 days of PCA the mean value measured in the adults is reported.

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 the logarithm of PCA (or age). Figure features are identical to those of Fig. 3.

Figure 5

Example of a XIC procedure for the relative quantitation of co-eluting proteins/peptides. The upper panel shows the HPLC-MS total ion current (TIC) profile in the elution range 22.29-26.71 min. The two broad and intense peaks are mainly due to the highly charged aPRPs proteoforms. In the middle and lower panels, XIC peaks of the almost co-eluting histatin 1 and its non-phosphorylated derivatives, are shown, respectively. They were obtained by selecting and only revealing the signal of the ions with the *m/z* values reported on the panels specific of the two peptides [3]. XIC peak areas (MA: measured area) are proportional to peptide concentration [6,7] and were utilized to calculate the percentages of the various proteoforms. RT: retention time. NL: normalization level.

Figure 6

Percentages of non-phosphorylated histatin 1 measured in 230 samples as a function of the logarithm of PCA (or age). 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 218 samples as a function of the logarithm of PCA (or age). Figure

features are identical to those of Fig. 3.

Figure 8

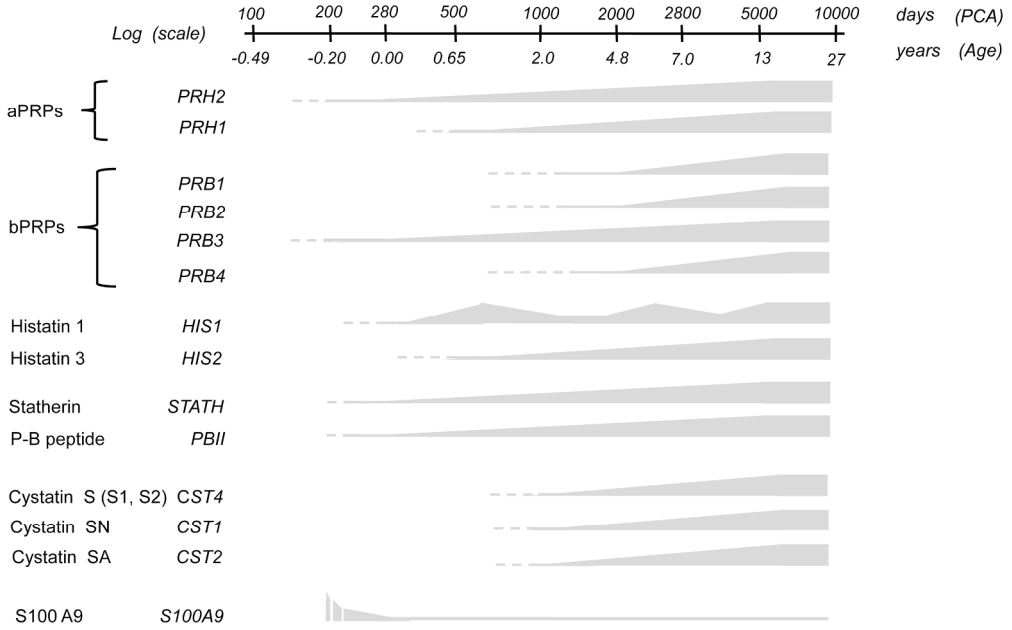
Percentages of cystatin S (non-phosphorylated; panel a), cystatin S1 (mono-phosphorylated; panel b) and cystatin S2 (di-phosphorylated, panel c) measured in 124 samples as a function of the logarithm of PCA (or age). Figure features are identical to those of Fig. 3.

Figure 9

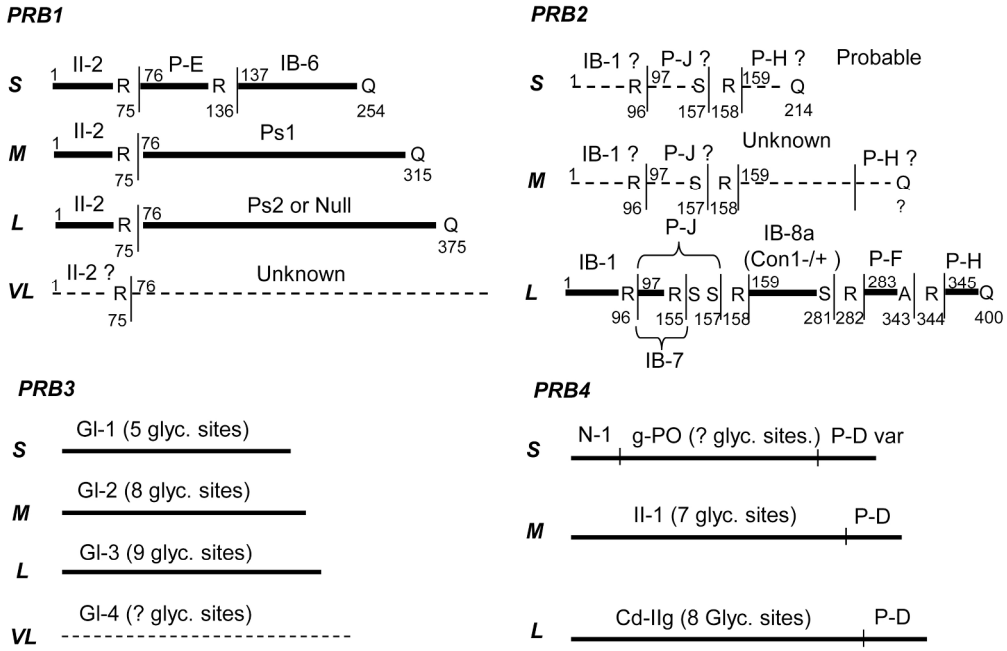
Percentages of non-phosphorylated S100A9 short (panel a) and of all the non-phosphorylated S100A9 long proteoforms (panel b) measured in 187 samples as a function of the logarithm of PCA (or age). Figure features are identical to those of Fig. 3.

Figure 10

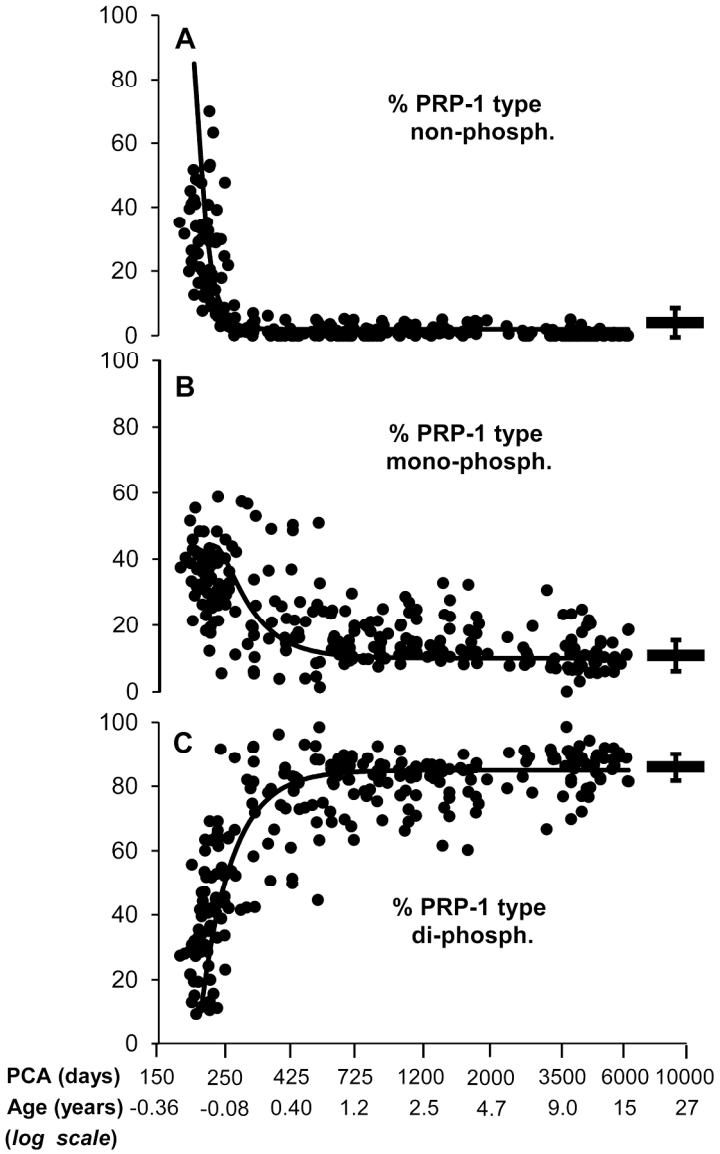
Panel a: percentages of truncated PRP-3 type proteoforms with respect to the sum of entire (PRP-1 type) and truncated (PRP-3 type) proteoforms measured in 261 samples. Panel b shows the percentage of Hst5+Hst6 with respect to Hst3+Hst5+Hst6 measured in 192 samples. Panel c shows the percentage of Hst 5 with respect to Hst5+Hst6 measured in 188 samples. Panel d shows the percentage of short S100A9 (phosphorylated and non-phosphorylated) with respect to total S100A9 measured in 186 samples as a function of the logarithm of PCA. Other features are identical to those of Fig. 3.



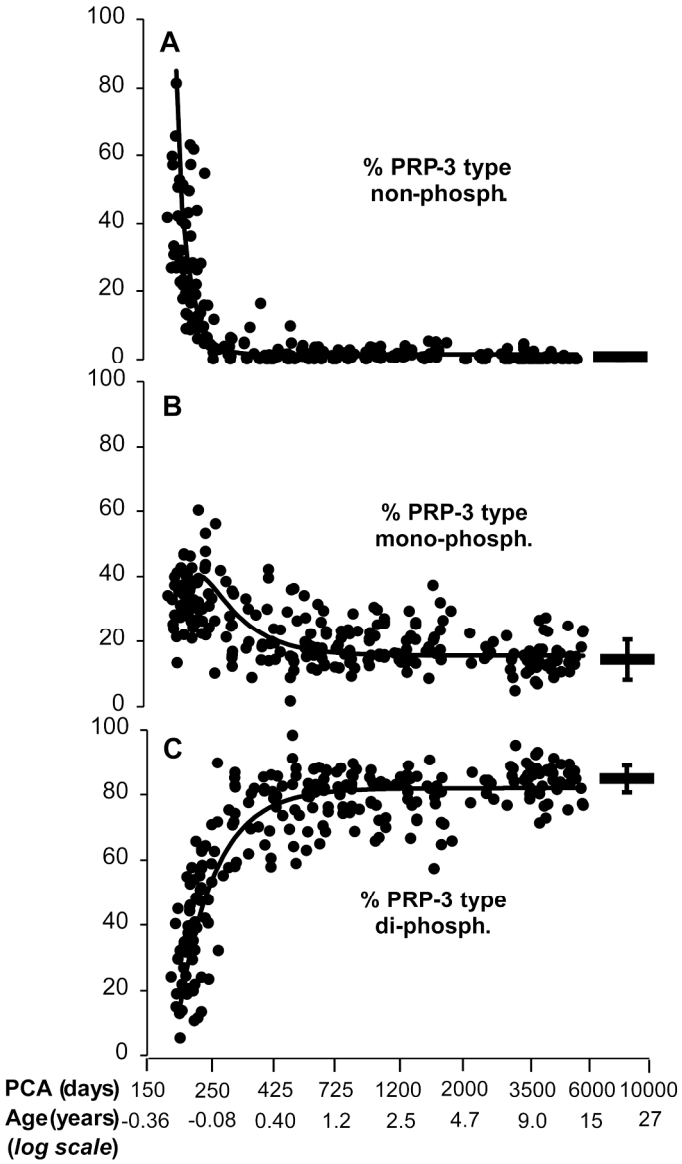
110x68mm (600 x 600 DPI)



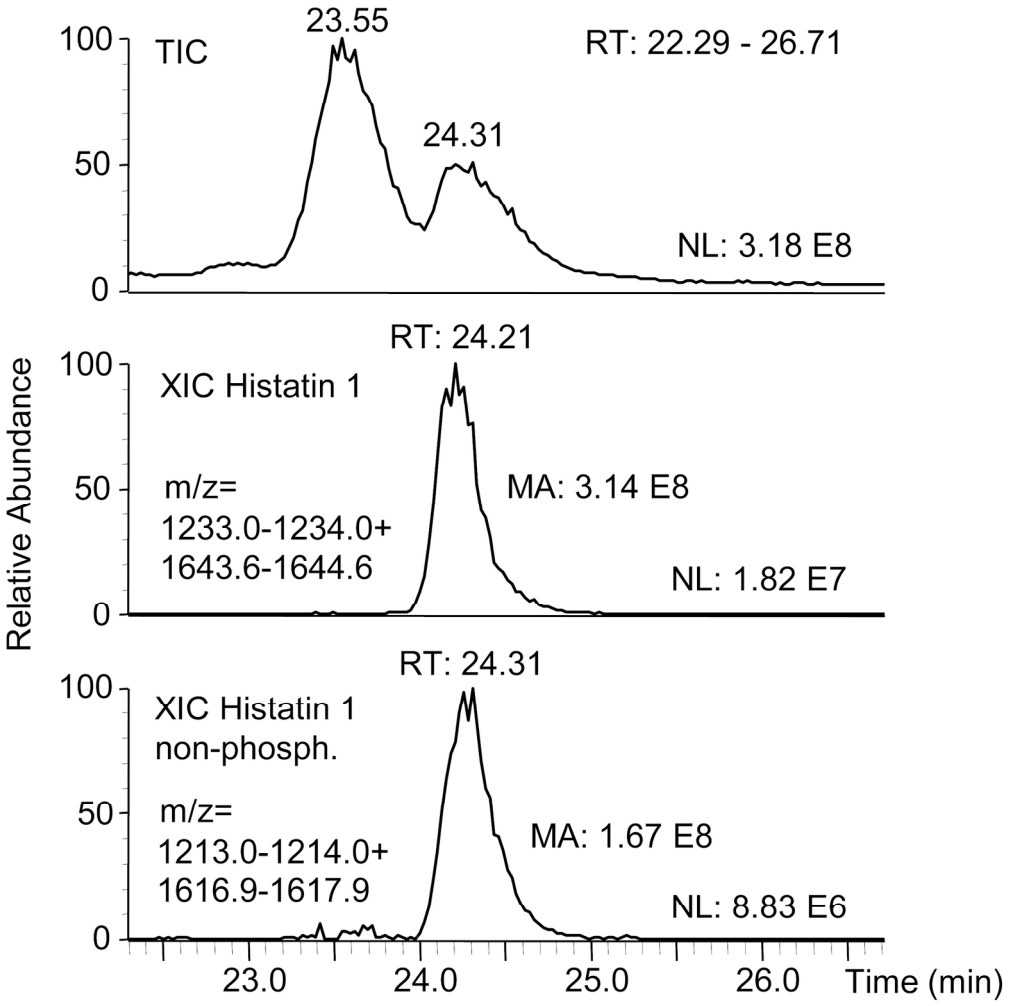
115x75mm (600 x 600 DPI)



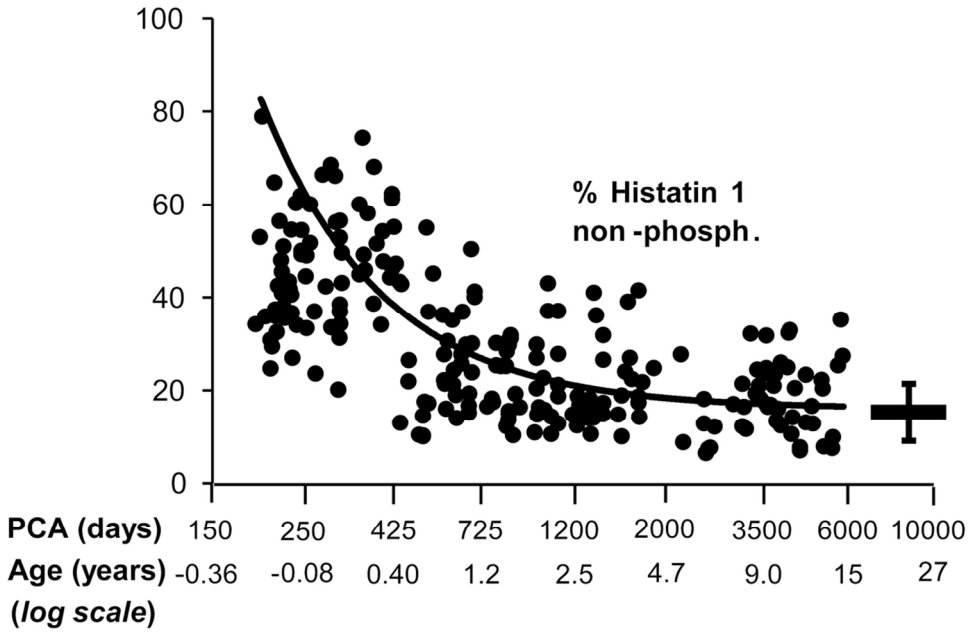
128x197mm (600 x 600 DPI)



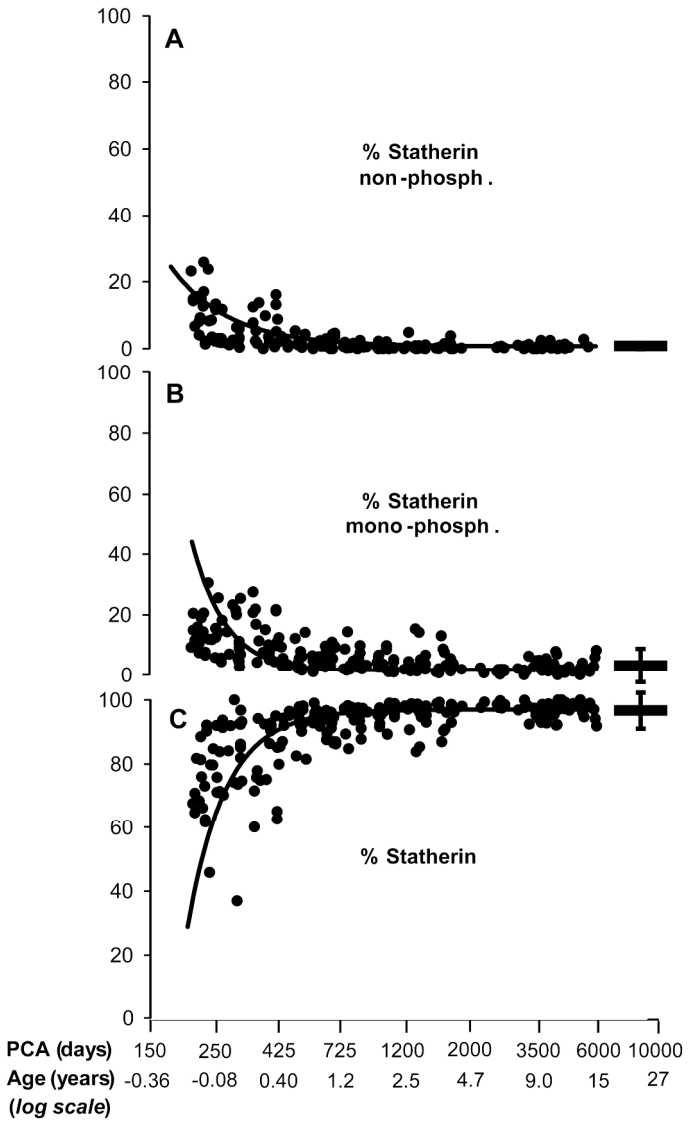
129x199mm (600 x 600 DPI)



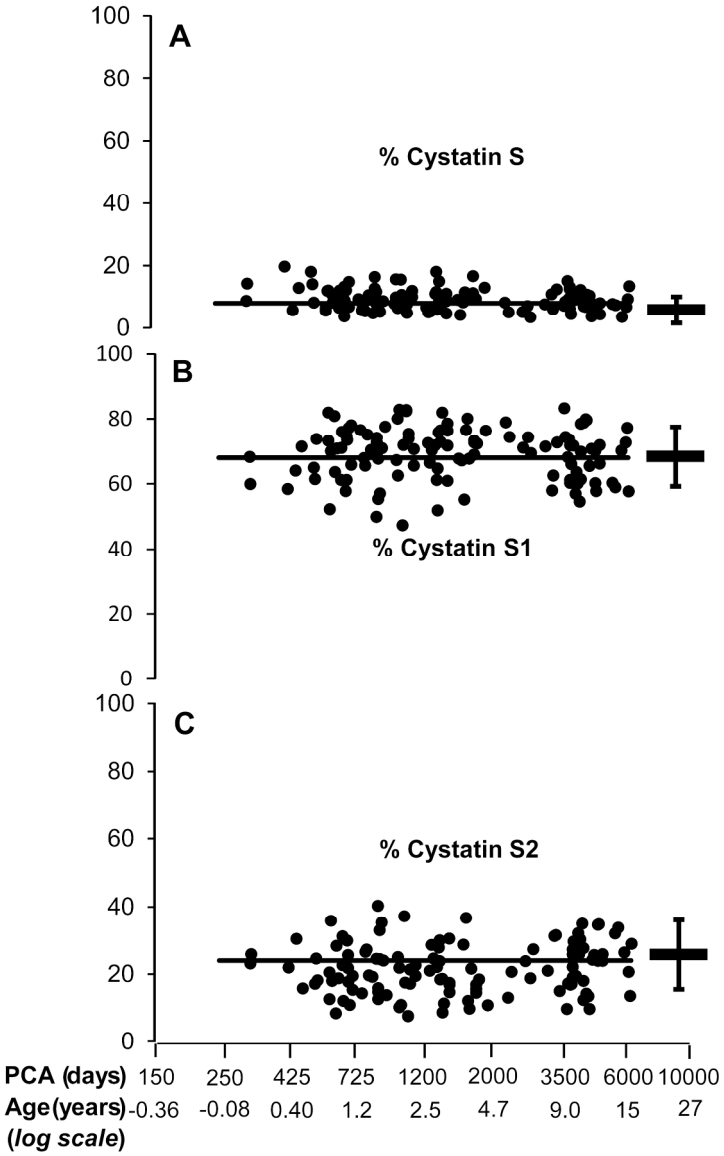
85x85mm (600 x 600 DPI)



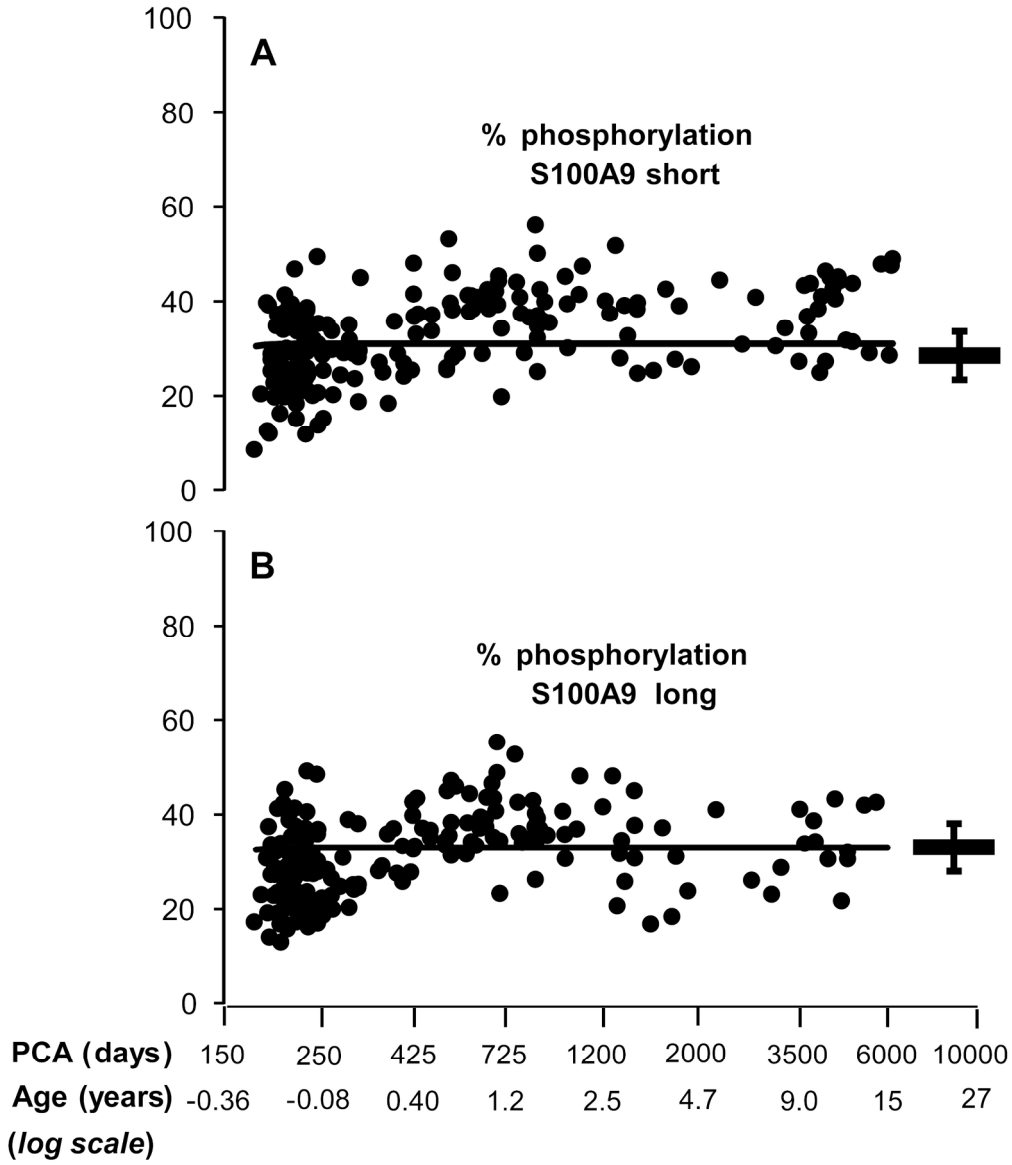
53x34mm (600 x 600 DPI)



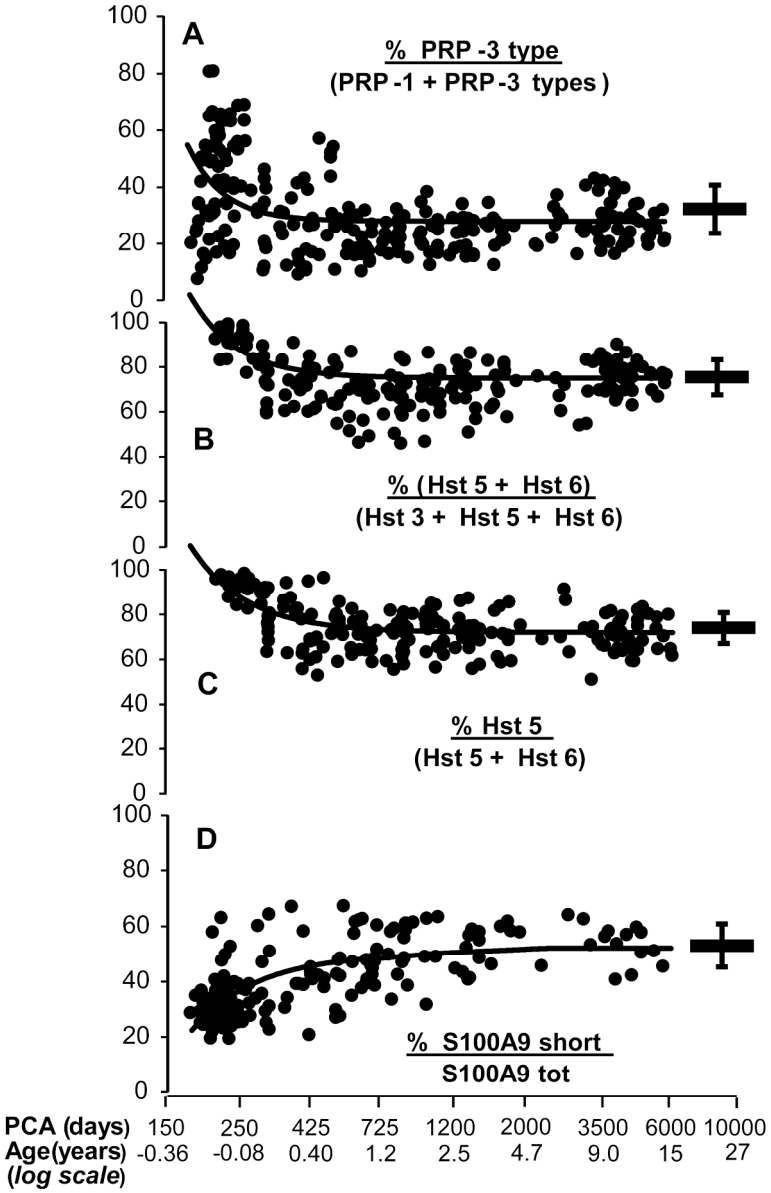
142x240mm (600 x 600 DPI)



122x178mm (600 x 600 DPI)



96x112mm (600 x 600 DPI)



119x171mm (600 x 600 DPI)