Supplementary data

1. Introduction

Amyotrophic lateral sclerosis (ALS) is an adult-onset neurodegenerative disease defined by the impairment of the voluntary motor system and ubiquitin-positive intraneuronal aggregates in anterior horn cells characterized by progressive loss of motor neurons. Approximately 5% of ALS cases are familial (FALS), with the remaining being sporadic (SALS) (Chio et al, 2008). Although a specific genetic alteration responsible for the sporadic cases is still lacking, a number of modifier loci and associated genes have been recently identified, often with controversial results (Garber, 2008; Schymick et al., 2007b). Lack of replication and/or small effects of associated alleles are obvious problems in the understanding of a complex and genetically heterogeneous disorder such as sporadic ALS (Del Bo et al., 2008).

Increasing evidence points to a role for progranulin in the pathogenesis of neurodegenerative disorders. Null mutations in progranulin gene (*PGRN*) have been causally associated with frontotemporal lobar degeneration (FTLD) with ubiquitin-immunoreactive neuronal inclusions (FTLD-U) (Baker et al., 2008). Progranulin is a pleiotropic protein encoded by the single gene *PGRN*, located on human chromosome 17q21. It encodes a secreted precursor protein composed of a signal peptide and 7 tandem repeats of highly conserved motifs containing 12 cysteines that can be proteolytically cleaved to form a family of granulin peptides. Progranulin is a growth factor involved in the regulation of multiple processes including tumorigenesis, wound repair, development and inflammation. Progranulin is expressed in a number of ephitelia, as well as in specific neurons in the brain, including Purkinje cells, pyramidal cells of hippocampus and some cerebral cortical neurons (Eriksen and Mackenzie, 2008).

Clinicopathological data suggest that ALS and FTLD-U represent a disease continuum; the discovery of a common substrate for the ubiquinated inclusion bodies observed in both ALS and FTLD-U, the intranuclear DNA-binding protein TDP-43, further strengthens the association of these two disorders (Liscic et al., 2008). In addition, progranulin, stimulating the vascular endothelial growth factor expression, is of potential relevance for the pathogenesis of motor neuron diseases. Thus, a complex interaction between growth factors may play a role in ALS as well FTLD-U, as recently suggested. Furthermore, increased PGRN expression in areas of active degeneration in ALS has been reported through an immunohistochemical approach (Irwin et al., 2008). Recently, several rare missense nucleotide variants within *PGRN* gene have been identified in sporadic ALS patients, although the pathogenicity of these variants remains unclear (Schymick et al, 2007). Moreover, common polymorphisms and haplotypes were significantly associated with a reduction in age at onset and a shorter survival after onset of ALS, at least in Belgian and Dutch populations (Sleegers et al., 2008). However, replication of this positive association in additional independent cohorts of ALS patients with a different ethic origin is mandatory. Few missense mutations have been also reported in sporadic AD patients (Brouwers et al., 2008; Cortini et al., 2008) and in subjects affected by Parkinson's disease (Nuytemans et al., 2008). In contrast to FTLD-U patients, null mutations are very rare in other neurodegenerative disorders (Brouwers et

al., 2007). *PGRN* missense mutations were predicted to mildly affect PGRN function and so suggested to act as low penetrant risk factors for neurodegeneration.

All these above reported findings, as well as the overlap between ALS and FTLD-U prompted us to screen a series of Italian ALS patients for mutation in *PGRN* gene.

2. Methods

2.1. Patients and controls

The study was performed in accordance with the Declaration of Helsinki; an informed consent, approved by the Institutional Review Board of the "IRCCS Foundation Ospedale Maggiore Policlinico -Milan, Italy" was obtained by patients and controls.

The 251 Italian subjects affected by ALS (including 12 FALS patients) included in this cross-sectional study referred to three Italian ALS Centers: i) the Department of Neurology, Foundation IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan; ii) the Department of Neurosciences -University of Padua; iii) the Neurological Institute -University of Pisa. All patients fulfilled the El Escorial criteria for probable or definite ALS and were recruited from January 2001 to April 2008. In FALS cases (7 males and 5 females), mutations in SOD1, ANG, VAPB, TARDBP and FUS/TLS genes were excluded before their inclusion in the present study. Data generated from the FALS patients was not included in the association analysis. For association analysis, the control group consisted of 181 healthy volunteers matched for ethnic background and age to cases. Features of SALS subjects and healthy subjects are shown in Table S1. The coding region and intron/exon boundaries of PGRN gene were screened for mutations in all ALS patients recruited at the Department of Neurology -Ospedale Maggiore Policlinico-Milano and the Department of Neurosciences -University of Padua (n=146). Additionally, exon 3, exon 5 and exon 13 (which includes the 3'UTR region) were analyzed in 105 SALS patients recruited at the Neurological Institute -University of Pisa.

2.2. Genetic analysis

Total DNA was isolated from peripheral blood according to standard protocols. *PGRN* exons 1 through 13 (and at least 20 bp of flanking intronic sequence) were amplified by PCR; all fragments were purified and directly sequenced using BigDyeTerminatorTM protocol on an automated 3100 ABI Prism Genetic Analyzer (Applied Biosystem, Foster City, CA).

For haplotype analysis, the genomic region encompassing the two markers D17S946 and D17S809 was investigated and corresponding haplotypes were constructed manually. Briefly, haplotype analysis was carried out by PCR with fluorescently labeled primers on an automated 3100 ABI Prism Genetic Analyzer through the GeneMapper software-version 4.0 (Applied Biosystem). To compare the data of the markers, haplotyping of the PGRN gene region was carried out in the affected ALS-FTD subject (New York Brain Bank sample T-51), recently shown to carry the heterozygous pS120Y mutation (Schymick et al. 2007a). For haplotype analysis, DNA samples from relatives were not available. The frequency of shared allele for markers D17S1801, D17S951, D17S931 and D17S806, as well as for SNPs within *DBF4B* gene was calculated on 48 Italian healthy subjects.

2.3. Statistical analysis

Two-tailed Pearson's Chi-square and Fisher's exact tests were used to compare genotype or allele frequencies; corresponding odds ratios (ORs) and 95% confidence intervals

(CIs) were calculated. PGRN circulating were compared across patients and controls using the nonparametric Mann-Whitney Rank Sum Test. The same test was used for differences related to the survival after disease onset (time elapsed between first symptoms and ALS diagnosis) in different genotype carriers. Furthermore, survival analysis was estimated using the Kaplan-Meier method and compared by the log-rank test, using age at the time of disease onset as the end-point and the different genetic variants as predictors. The Student's *t*-test was used to compare the age at onset across the different genotypes.

Haploview v.4.1 software (http://www.broad.mit.edu/mpg/haploview) was used to test for Linkage Disequilibrium (LD), as previously reported (Barrett et al., 2005).

2.4. Serum progranulin determination

Serum samples for the progranulin dosage were available from 30 ALS patients (17 males) and 27 healthy subjects (16 males). For the measurement to be made, healthy subjects needed a normal platelet count, no memory impairment (MMSE>28) or psychobehavioural complaints, no clinical evidence of cancer, metastases or diabetes. Progranulin concentration was measured with a commercial human specific ELISA kit (AdipoGen, Korea), based on the quantitative sandwich enzyme immunoassay technique, according to the instruction of the manufacturer.

2.5. Isolation of PBMC cells and expression analysis

Total RNA was obtained from peripheral blood mononuclear cells (PBMC) of 11 ALS patients (5 males, mean age: 60.5 ± 12.7) and 9 control subjects (4 males, mean age: 57.4 ± 12.1).

Total RNA was reverse-transcribed using the Ready to go kit (Usb, USA). Subsequently, the relative expression of progranulin was measured by quantitative RT-PCR using TaqMan gene expression assays (ABI, Assay ID: Hs00173570_m1) with 18S (ABI, Assay ID: HS99999901_s1) as an endogenous control for sample normalization. In particular, the relative quantity of RNA was measured in triplicate by the ΔΔCT method relative to the data from a control subject.

2.6. In silico analysis

The effect of nonsynonymous missense mutation on protein function was estimated in silico using the Sorting Intolerant Fron Tolerant (SIFTv.2) program. SIFT score <0.05 is predicting affected protein function.

2.7. Protein sequence conservation

Using ClustalW2, a program designated to create alignments of multiple target sequences, human PGRN was aligned with PGRN protein sequence of different species

2.8. Genotyping assays

For genotyping analysis, the two common single nucleotide polymorphisms rs9897526 and rs5848 were analyzed by using TaqMan methodology. Each Taqman 5'-nuclease assay employed 25ng of genomic DNA as template. Assay-on-demand products (ABI

assays: C_2548248_10 and C_7452046_20) were used for rs9897526 and rs5848 genotyping, respectively. Probes specific for these variants were labeled with 6-FAM and VIC as reporter dyes and MGB-NFQ (ABI) as quencher. All the assays were performed in 20 μl reactions in 96-well plates using an ABI PRISM ®7000 instrument (ABI).

3. **Additional results**

3.1. Case-control association analysis

Deviations from Hardy Weinberg Equilibrium were not observed for any of the two genetic variants (rs9897526 and rs5848) analyzed in the present study. The analysis of the age at disease onset according to the different genotypes showed no statistically significant results (figure S2). Genotype frequencies for both variants were similar when calculated on the basis of site of onset (bulbar vs spinal) and gender (data not shown). No difference was observed according to survival after disease onset by rs9897526 (GG vs GA/AA, p=0.63) and rs5848 (CC vs CT/TT, p=0.36).

3.2. p.S120Y gene variant

The p.S120Y variant was identified in a female sporadic patient with a predominant upper motor neuron phenotype. Her age at symptom onset was 68 years. Disease progression was slow, with a ALS-FRS of 31/ 40 after 3 years of disease onset. MMSE was 24/30, and Frontal Assessment Battery 13/18. Neuropsychological tests showed mild diffuse cognitive dysfunction, though she did not fulfill criteria for FTD. As shown in supplementary data- figure S3, S120Y is localised within the linker-GF domain of progranulin and the residue S120 is not highly conserved across species. Finally, substitution 120S>Y is predicted to be tolerated with a score of 0.12, as detected by SIFT program. The p.S120Y variant was absent in 181 healthy subjects and in all 105 SALS patients subsequently included in the study for association analysis.
Furthermore, there was no difference between the blood mRNA levels or the serum PGRN protein levels of the patient carrying p.S120Y variant and controls (relative quantity in patient carrying p.S120Y mutation = 0.94 versus relative quantity range in controls = 0.49-1.5; serum PGRN levels = 149ng/ml versus range in controls = 75.2-311.5ng/ml).

Haplotype analysis with intragenic SNPs and microsatellite markers flanking the *PGRN* region were carried out in the Italian SALS patient, as well as in sample T-51 carrying the p.S120Y change, previously reported (Schymick et al., 2007a). Although D17S1861 marker shows a different pattern in the two cases, haplotype analysis suggests the presence of a shared common founder associated to S120Y variant. The 2-bp variation observed at marker D17S1861 is probably the consequence of a mutation occurring in this microsatellite. The result of the allele sharing analysis is reported in table S4.

3.3. Haplotype analysis

Pairwise LD analysis computed with Haploview software indicated no LD between rs9897526 and rs5848 polymorphisms (D'=0.44), thus no haplotype block or markers combination was highlighted.

Supplementary tables

Table S1. Characteristics of Italian SALS patients and healthy subjects.

	SALS patients	**Healthy subjects**
Number of subjects	239	181
Gender (M/F)	153 / 86	102 / 79
Age at onset (years)	58.3 ± 12.9	
Age at blood collection (years)	60.8 ± 12.2	67.4 ± 8.8
Bulbar onset	21.4%	
Disease duration (months)	30.5 ± 28.4	
SOD1, ANG positive subjects	none	

Table S2. List of *PGRN* coding and intronic gene variations as observed in SALS patients genotyping

Genome	Predicted cDNA mutation	Predicted protein	Location	rs number	SALS
g.100121C>T	c.55C>T	p. Arg19Trp	EX2		1/134
g.100165C>T	c.99C>T	p. Asp33	EX2		1/134
g.100460G>A	c.264+7G>A (IVS3+7G>A)	-	IVS3		17/239
g.100474G>A	c.264+21G>A (IVS3+21G>A)	-	IVS3	rs9897526	47/239
g.101164T>C	c.384T>C	p. Asp128	EX5	rs25646	13/239
g.102072G>A	c.835+7G>A (IVS8+7G>A)	-	IVS8		2/134
g.103383C>T	c.1554C>T	p. Asp518	EX12		1/134
g.103778C>T	c.*78C>T (3'UTR+78C>T)	-	3'UTR	rs5848	110/239

Table S3. Rs9897526 and rs5848 genotype and allelic frequencies as detected in Italian SALS patients (n=239) and healthy subjects (n=181).

		SALS	Controls	p*	OR [95%CI]	p**
rs9897526	**GG**	80.3 %	79.0 %	0.74	Ref	
	GA	18.0 %	18.2 %		0.97 [0.67-1.66]	>0.99
	AA	1.7 %	2.8 %		0.59 [0.11-2.82]	0.50
	G	89.3 %	88.1 %	0.58	Ref	
	A	10.7 %	11.9 %		0.88 [0.56-1.39]	0.58
rs5848	**CC**	53.4 %	53.6 %	>0.99	Ref	
	CT	36.0 %	35.9 %		1.00 [0.68-1.63]	>0.99
	TT	10.6 %	10.5 %		1.01 [0.46-2.19]	>0.99
	C	71.4 %	71.5 %	0.91	Ref	
	T	28.6 %	28.5 %		1.00 [0.71-1.42]	>0.99

*p values were calculated using χ^2 exact test with two degrees of freedom (for genotypes) or with one degree of freedom (for allele comparisons). **p values referred to OR [95%CI].

Table S4. Allele sharing analysis in p.S120Y carriers.

Marker / SNP	Physical position (Mb)	T-51 case (Schymick et al)		Italian patient (this study)	Shared allele frequency
D17S946	34,1	179	189	187	
D17S1299	36,24	265	269	265	
D17S1801	37,66	161	163	161	52% *
D17S951	39,17	176	174	176	21% *
***PGRN* gene**	39,77-39,78				
rs9897526		G	A	G	88%
(39782466)		A	C	A	
p.S120Y		G	A	G	76% **
rs850713		C	T	C	71%
(39783258)					
rs5848 (39785770)					
DBF4B gene	40,14-40,18				
rs58812052	40,15	G	G	G	100% *
rs6503399	40,15	C	C	C	46% *
rs2906781	4015	T	T	T	96% *
D17S1861	40,16	176	192	174	
rs34337196	40,18	G	G	G	100 *
D17S931	42,35	168	166	168	39% *
D17S806	43,16	197	197	197	25% *
D17S958	43,6	179	175	179	
D17S1877	46,83	162	162	162	
D17S809	47,27	140	140	140	

* Frequency of shared allele based on 48 Italian control individuals. ** Frequency of shared allele based on 96 Italian control individuals.

Figure legends

Figure S1

Quantitative analyses of circulating progranulin as detected in a subgroup of ALS patients and controls. Panel A: mRNA levels. Panel B: serum levels.

Figure S2

Analysis of the age at disease onset according to the different genotypes: IVS3+21G>A variant (rs9897526): panels A and C; 3'UTR78C>T variant (rs5848): panels B and D. Panels A and B: age of onset (mean ± standard deviation); panels C and D: survival analysis estimated using the Kaplan-Meier method and compared by the log-rank test.

Figure S3

A. Representative section of the electropherogram from genomic DNA *PGRN* exon 5 fragment encompassing the identified heterozygous c.359C>A mutation, as observed in SALS patient.

B. Phylogenetic conservation of the amino acid residue S120 within the amino acid sequence from residue aa111 to aa127 of the human *PGRN* gene as observed in multiple, evolutionarily diverse species.

C. Schematic presentation of the full-length human progranulin protein sequence showing all coding variations as detected in ALS patients: p.S120Y mutation is in red, rare variants/polymorphisms are in green. The 7.5 grn-domains are shown in bold, linker domains and N-terminal region are shown in italic. Other missense mutations previously reported in ALS, ALS-FTD, AD and PD patients and absent in healthy individuals, are indicated with blue circles.

Figure S1

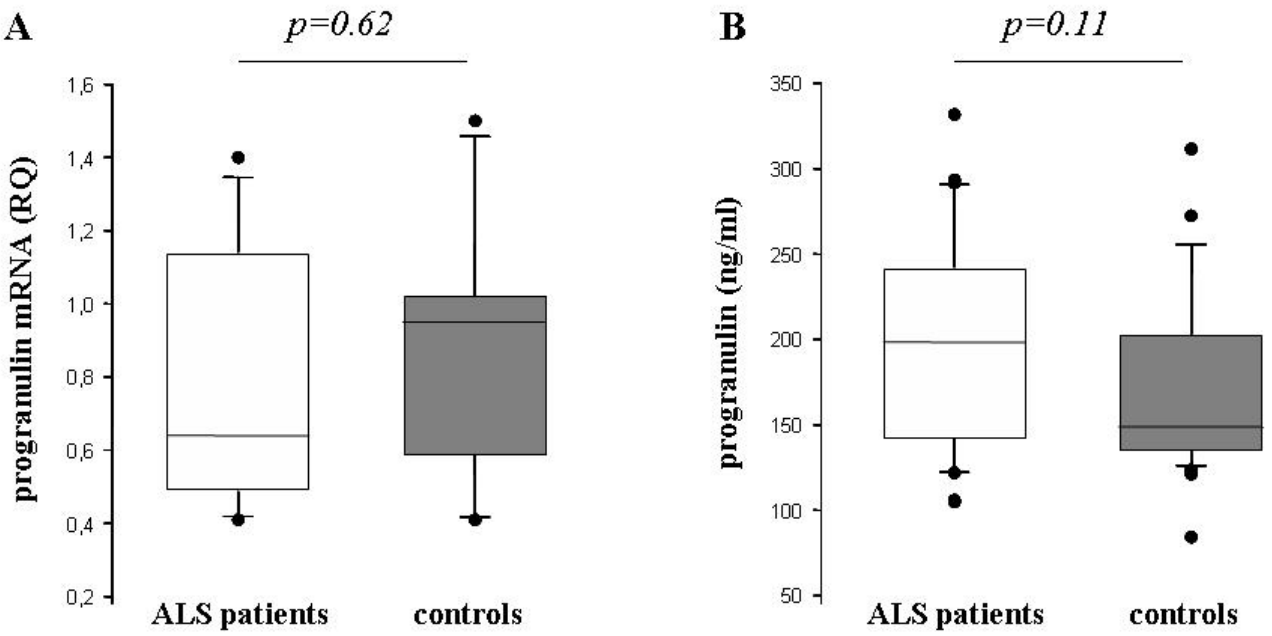

Figure S1

Figure S2

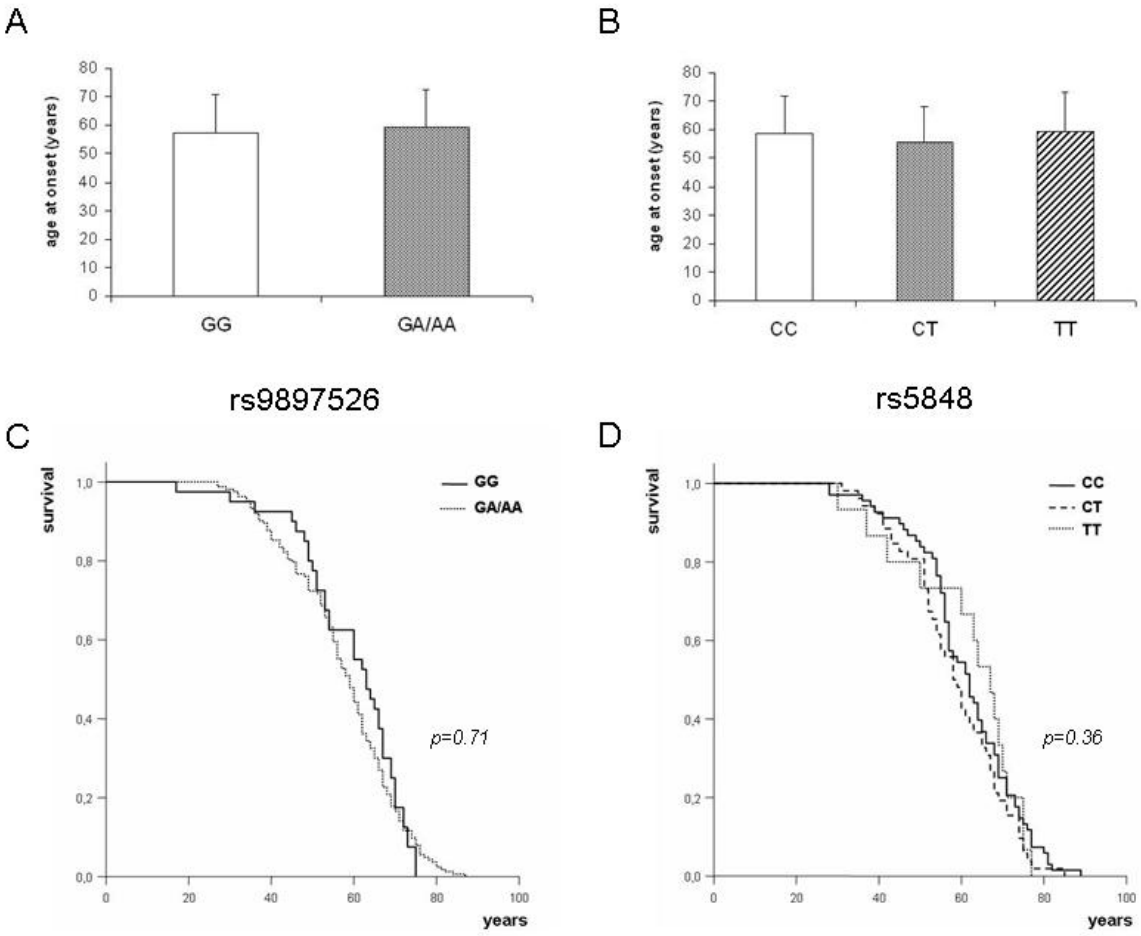

Figure S2

Figure S3

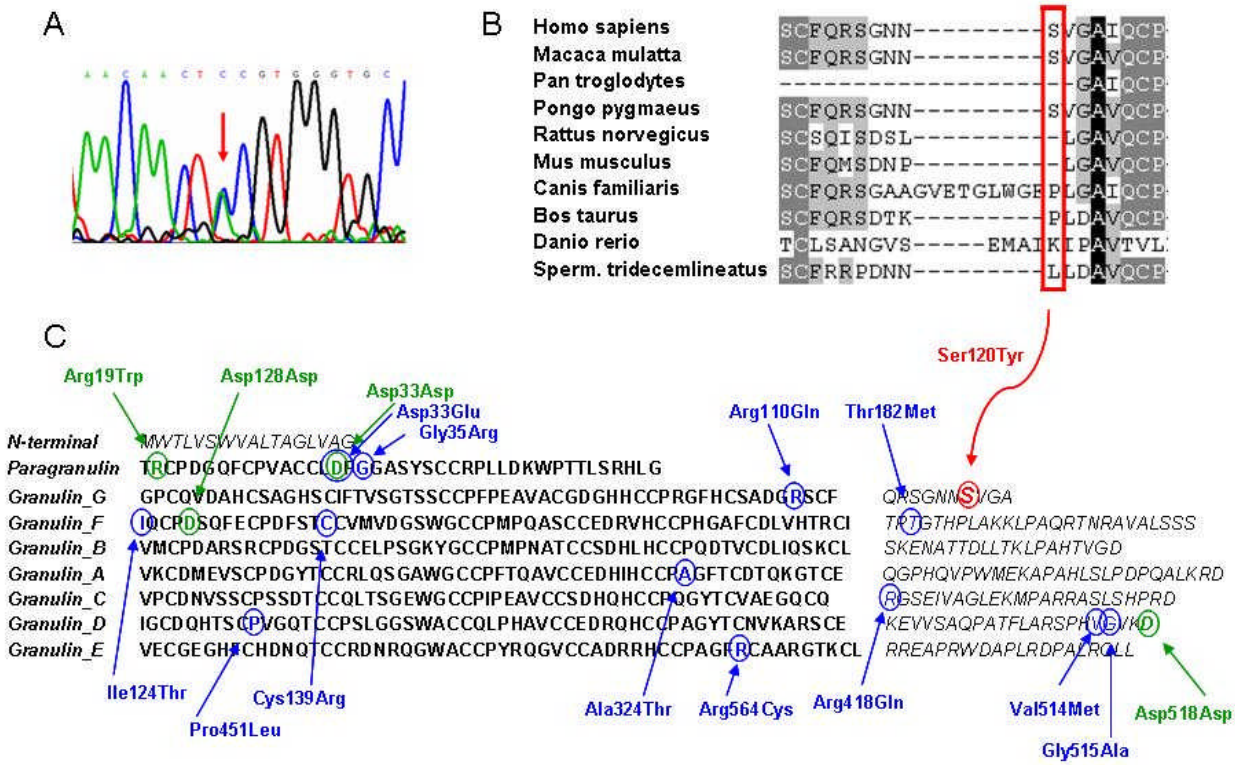

Figure S3