



Università degli Studi di Cagliari

DOTTORATO DI RICERCA

BIOLOGIA E BIOCHIMICA DELL'UOMO E DELL'AMBIENTE

Ciclo XXVIII

**Sequenziamento e Analisi Molecolare di
Varianti Alleliche del Gene PRB1**

Settore scientifico disciplinare di afferenza: BIO/10

Presentata da:

Dottor Roberto Orrù

Coordinatore Dottorato:

Prof. Emanuele Sanna

Tutor/Relatore:

Prof.ssa Alessandra Padiglia

Esame finale anno accademico 2014 – 2015

INDICE

ABSTRACT	2
INTRODUZIONE	5
Le Proteine Salivari Umane ricche in Prolina	5
PRB4	11
PRB3	12
PRB2	14
PRB1	15
MATERIALI E METODI	21
Selezione dei soggetti	21
Estrazione DNA	22
Primer Walking	23
Gel elettroforesi	30
Purificazione frammenti PCR	31
Sequenziamento	32
Analisi in silico	33
RISULTATI E DISCUSSIONE	34
Amplificazione con il primo set di primers e purificazione dei cloni PCR	34
Sequenziamento dei cloni amplificati col primo set di primers	38
Amplificazione con il secondo set di primers e purificazione dei cloni PCR	43
Sequenziamento dei cloni amplificati col secondo set di primers	50
Amplificazione con il terzo set di primers e purificazione dei cloni PCR	59
Sequenziamento dei cloni amplificati col terzo set di primers	62
Analisi in silico	73
CONCLUSIONI	85
BIBLIOGRAFIA	88
ELENCO DELLE PUBBLICAZIONI	96
RINGRAZIAMENTI	97

ABSTRACT

The aim of this work was the molecular characterization of gene PRB1 in three subjects with different salivary proteome profile. The PRB1 gene encodes a member of the heterogeneous family of Basic Proline rich Proteins (PRB), produced by the parotid gland and secreted in human saliva. The protein polymorphism resulting from post-translational modifications of the pre-protein PRB1, are further complicated by the gene allelic variants. Those variants involves the third exon of the gene, and are related to the presence or absence of a tandem repeat sequence 183 bp long. That tandem repeat may originate three different allelic forms, defined small, medium or large depending on the number of repeats inside the third exon.

The information reported in the literature about the class of PRP genes, and in particular PRB1 are, from the molecular point of view, fragmentary and difficult to reconstruct. In fact, the literature on this topic covers a time span ranging from the seventies (Azen & Oppenheim, 1973), when the associated peptides have been characterized for the first time, up to the end of the nineties (Stubbs et al., 1998). Since the eighties, there have been various attempts to characterize the PRB1 allelic variants, based on the resulting peptides. This approach has greatly complicated the classification of the gene, but also of his allelic variants and the deriving proteins (Maeda et al., 1985 ; Lyons et al., 1988a, b; Azen et al., 1993). Today we know that from PRB1, considering its allelic and splicing variants, are expressed six different proteins (Marconi et al., 2015). However, amongst the three allelic variants of this gene, the only of which we have the complete nucleotide sequence is the medium. Conversely, for the small and large we have so far only partial sequences, referring only to the third exon. Among the objectives of this study, there was the reconstruction of the complete sequences of the small and large variants, of which there have been previously characterized two putative bearers from the proteomic point of view. At the same time, we also sequenced the PRB1 gene from a putative homozygotic bearer of the medium variant.

While for the subject bearer of the small variant, we were able to identify and sequence the entire gene, the same was not possible for the subject bearer of the large variant. In fact, from our analysis this last subject has shown to be homozygous carrier of the medium variant. This result is incompatible with the presence in his saliva of Ps2 protein, which is characteristic of the large variant. In the absence of further mass spectrometric analysis, we cannot explain the reason for this discrepancy between

genomic and proteomic data. The analysis of the complete sequences did not allow us to understand why it has not been identified to the mass spectrometer the peptide relative to the splice variant classified by Maeda with the acronym cP5 (Maeda et al., 1985), while in all the subject has been identified the peptide and the relative cP4 splice variant. Pk-o protein has so far been reported to be expressed from the large variant only, although the first classification of splice variants had been made from an individual carrying the medium variant. Since in none of the three sequenced subjects the splice acceptor site in 3' of the cP5 variant results to be altered, it is possible that the relative peptide has not been identified to the mass spectrometer. Otherwise, specific splicing factors, that recognize Splicing Regulatory Elements (SRES; Hernandez-Imaz et al., 2015) in the neighborhood of the two molecular acceptor sites, prevented the transcription of the cP5 in favor of cP4. In order to shed light on this aspect, there are some possible strategies in perspective. Recently, in fact, studies have been published in which extensive *in vivo* screening for the detection of SRES is coupled with RNA affinity purification and mass spectrometry, which allow also the identification of the splicing factors which bound to the SRES (Wang Y & Z Wang, 2014; Wang et al., 2013). This strategy could be effectively integrated with the analysis of minigenes (Hernandez-Imaz et al., 2015), specifically designed on the model of the third exon of PRB1, where alternative splicing takes place and where are probably placed the SRES.

Molecular analysis has also enabled us to identify a set of Single Nucleotide Polymorphisms (SNPs), most of which have never been described so far, localized specially inside introns and some even within exons. In particular, for two of these, we have detected a non-synonymous substitution at the amino acid level, the role of which can only be clarified by means of appropriate functional studies. Although translationally silent, even synonymous SNPs may have an important function, especially in alternative splicing. In fact, these substitutions can determine the genesis or destruction of SRES, or strengthen cryptic donor /acceptor sites. Furthermore, they can result in the alteration of the secondary structure of the mRNA, important for the exons definition and the pause sites of RNA II. This could, in turn, alter the processivity of the polymerase, with possible consequences on the choice of splicing sites.

The reconstruction of the complete sequences and the comparison with the data available in the literature (Lyons et al., 1988) allowed us to produce an updated model, which allows to explain the generation of allelic variants small and large from the

medium variant. Hopefully, the reconstruction also of the large allelic variation, and possibly the very large, will lay the basis for the validation of our model and possibly also its extension to other allelic variants of the PRB gene class.

The genotypic characterization of allelic variants of PRB1 may be, in the future, an important predictive tool about the subjective susceptibility towards a series of oral pathogens. In fact, polymorphic variations in the PRB1 peptides can build the basis for the prediction of individual differences at the level of the oral *microflora*, with repercussions on the susceptibility to infections and diseases in this body part (Newman et al., 1993 and 1996; O'Sullivan et al., 2000).

INTRODUZIONE

Le Proteine Salivari Umane ricche in Prolina

Le Proteine Salivari Umane ricche in Prolina (PRPs) rappresentano le principali componenti proteiche della saliva umana (tra il 20-30% peso/peso; Manconi et al., 2015). Esse sono molto comuni nella saliva degli erbivori e dei primati, mentre risultano del tutto assenti nei carnivori (Boze et al., 2010). La loro sequenza aminoacidica è molto ripetitiva, caratterizzata da una prevalenza degli aminoacidi prolina (25-42%), glicina (16-22%) e acido glutammico/glutamina (15-28%; Maeda et al., 1985). La prolina è l'unico aminoacido la cui catena laterale ciclizza mediante legame del carbonio con l'azoto dello scheletro aminoacidico. Ciò limita le conformazioni possibili del residuo precedente nella catena peptidica, che adotterà più favorevolmente una conformazione estesa (Manconi et al., 2015; Boze et al., 2010). Infatti, le PRPs sono considerate proteine intrinsecamente non strutturate (*intrinsically unstructured proteins* o *IDP*; Tompa P, 2002; Boze et al., 2010). Esse costituiscono una famiglia di circa venti proteine (Stubbs et al., 1998) che, sulla base del punto isoelettrico (pI), sono comunemente divise in acide (aPRPs), basiche (bPRPs) e basiche glicosilate (gPRPs; Bennick et al., 1982). Tale classificazione è stata poi confermata per via di una serie di peculiarità strutturali e secretorie: le bPRP, infatti, sono secrete esclusivamente dalla parotide, le aPRP, invece, anche a livello delle ghiandole sottomandibolari e sublinguali (Messana et al., 2008).

Le PRP acide sono necessarie per mantenere l'omeostasi del calcio nel cavo orale, legando gli ioni Ca^{2+} e inibendo la formazione di cristalli di idrossiapatite (Bennick et al., 1981; Hay et al., 1979); le PRP basiche precipitano i tannini e prevengono l'assorbimento di questa potenziale tossina dal canale alimentare (Lu and Bennick, 1998); le PRP basiche glicosilate agiscono da lubrificanti del cavo orale (Hatton et al., 1985). Studi recenti (Messana et al., 2008) confermarono che le PRP basiche sono secrete solo dalla ghiandola parotide, sono scisse nel cavo orale da una serie di peptidasi, alcune delle quali ignote, per poi essere sottoposte a una serie di modificazioni post-traduzionali: queste includono la rimozione del peptide segnale e, in particolar modo, tagli convertasici, glicosilazione, fosforilazione e solfatazione (Manconi et al., 2015).

Le PRPs, sono codificate da un cluster di sei geni, strettamente associati sul cromosoma 12p13.2 e disposti nell'ordine PRB2, PRB1, PRB4, PRB3, PRH2, PRH1 (fig. 1; Sherer et al., 2006). I sei geni sono contenuti all'interno di un segmento di circa 0,5 Mbp, e tutti codificati sul filamento antisenso, ad eccezione di PRH2 (fig. 1; Sherer et al., 2006).

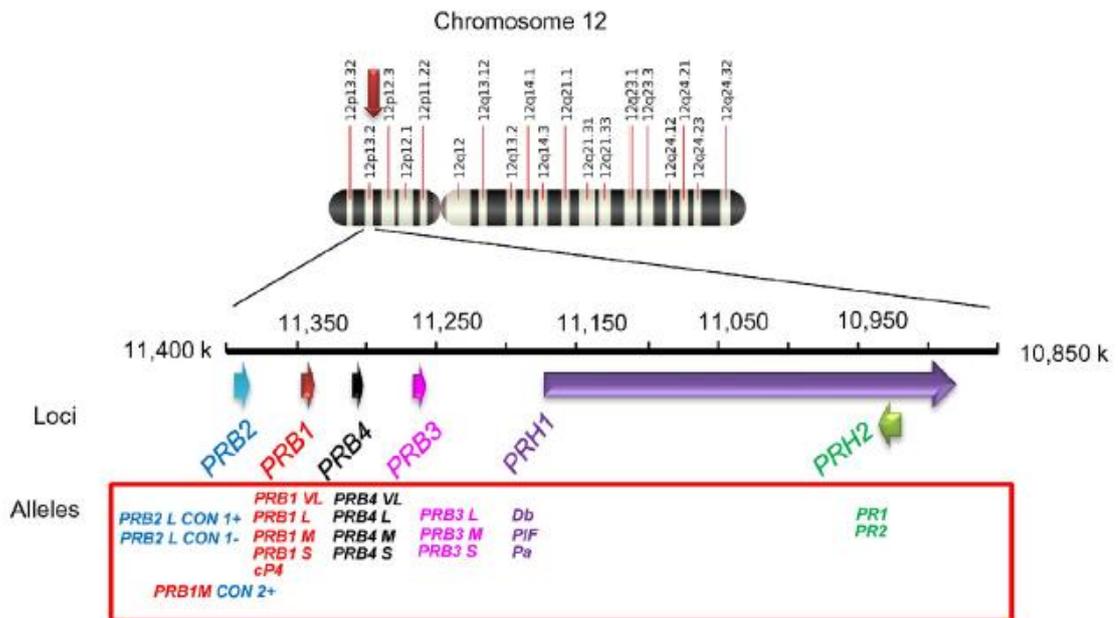


Fig. 1: rappresentazione schematica del cluster dei geni PRP (tratta da Manconi et al., 2015)

La classificazione e denominazione dei sei geni è stata fatta in base alla loro sensibilità rispetto a due diversi enzimi di restrizione. Sono denominati PRH1 e PRH2 i geni codificanti le PRP acide e sensibili all'azione di HaeIII, mentre i quattro geni PRB codificano le PRP basiche e contengono numerosi siti di restrizione per BstNI (Maeda et al., 1985; Lyons et al., 1988).

Studi proteomici hanno portato inizialmente all'ipotesi dell'esistenza di tredici loci differenti, ma geneticamente associati, che codificano le proteine salivari umane ricche in prolina. Oggi è noto che ciascun locus è poliallelico e produce molteplici PRPs (Manconi et al., 2015). Inizialmente, infatti, i loci attribuiti alle PRP acide erano: Pr, proline rich (Azen & Oppenheim, 1973); Pa, acidic protein (Friedman & Merritt, 1975); Db, double band (Azen & Denniston, 1974); PIF, parotid isoelectrofocusing variant

(Azen & Denniston, 1981); As (Minaguchi et al., 1990), Au, At e Aw (Shintani et al., 1990a). In realtà, come visto, le diverse aPRP presenti nella saliva derivano dall'espressione dei loci PRH1 e PRH2. Le pre-proproteine derivanti dai due loci aPRP danno origine a differenti isoforme, a causa di variazioni puntiformi del DNA e modificazioni post-traduzionali: PRH1 codifica per le isoforme PIF, Pa e Db; PRH2 invece codifica le proteine PRP-1 e PRP-2. Le aPRP sono necessarie per mantenere l'omeostasi del calcio nel cavo orale (Gron et al., 1976), consentendo di mantenere la sua concentrazione superiore ai livelli di saturazione, grazie ad un'interazione in cui sono coinvolti i gruppi fosforici (Bennick et al., 1981; Hay et al., 1979). Inoltre, le aPRP sono adsorbite selettivamente sulla superficie di idrossiapatite a livello dentale (Hay et al., 1979; Moreno et al., 1979) e concorrono alla formazione della pellicola acquisita dello smalto (Moreno et al., 1982, Bennick et al., 1983). È stata altresì evidenziata una possibile funzione nella colonizzazione batterica del dente (Gibbons et al., 1991). Infatti, da recenti osservazioni è emerso che la concentrazione delle aPRP risultano maggiore nei soggetti privi di carie, inoltre esse sono soggette a frammentazione massiccia in soggetti considerati a rischio, o affetti da questa patologia dentale, rispetto ai soggetti sani (Vitorino et al., 2005).

Anche per le bPRP (o PRB) era stato inizialmente postulato un numero di loci pari ai peptidi rilevati nella saliva. Le bPRP costituiscono la famiglia maggiormente complessa tra le proteine salivari, sia in termini di variabilità dei loci PRB corrispettivi, che in termini di modificazioni post-traduzionali a carico di ciascuna proteina espressa. Levine e colleghi furono i pionieri, nel 1973, della caratterizzazione strutturale di questa famiglia di proteine (Levine et al., 1973). In seguito (Kauffman et al., 1979) furono identificate undici PRB nella saliva parotidea, dieci delle quali vennero sequenziate e denominate col prefisso IB negli anni successivi. Le PRB classificate dal gruppo questo gruppo di ricerca sono: IB-1, IB-4, IB-5, IB-6, IB-7, IB-8a, IB-8b, IB-8c, IB-9 e le proteine II-1 e II-2. Queste ultime due presentano carattere basico meno marcato rispetto alle altre PRB. Tra le PRB individuate da Kauffman e colleghi ci sono elevate omologie di sequenza e piccole discrepanze riconducibili a mutazioni puntiformi (Kauffman et al., 1991). Parallelamente, in Giappone, furono isolate nove PRB dalla saliva intera, sette delle quali sequenziate e denominate da P-A a P-I, in ordine alfabetico (Saitoh et al., 1979). In realtà, i due gruppi di ricerca avevano isolato, contemporaneamente e in modo indipendente l'uno dall'altro, i medesimi peptidi,

attribuendo loro, tuttavia, nomi differenti. Infatti, sette delle proteine caratterizzate dal primo gruppo di ricerca corrispondono ad altrettante (su nove) tra quelle descritte dal gruppo di ricerca giapponese: P-C→IB-8b; P-D→IB-5; P-E→IB-9; P-F→IB-8c; P-G→IB-7, P-H→IB-5, P-I→IB-6. P-A e P-B sono distinte, e non coincidono con nessuno dei peptidi descritti da Kauffman e colleghi. Il peptide P-B, infatti, non è espresso dai geni PRB, bensì da SMR3B (PROL3). P-A corrisponde al frammento C-terminale di P-B, mentre P-C (Isemura et al., 1980) deriva dal taglio proteolitico di PRP-1 e PRP-2.

A queste proteine si aggiungono le Ps (parotid size variant) 1 e 2, descritte e parzialmente sequenziate dal gruppo di Azen (Azen et al., 1980) e attribuite inizialmente al locus PRB2 (Azen et al., 1993). In seguito, Ps1 è stato correttamente attribuito a PRB1M, mentre Ps2 all'allele PRB1L (Azen et al., 1993; vedi anche sezione seguente su PRB1). Altre tre varianti, Ps 3, 4, 5 sono state descritte nella popolazione giapponese (Minaguchi et al., 1988; Shintani et al., 1990). Grazie a studi successivi (Lyons et al., 1988a; Stubbs et al., 1998) il numero dei loci PRB è stato ridotto a quattro. Ciascuno di essi può presentarsi sotto almeno tre varianti alleliche, differenti per la lunghezza del gene e definite *small*, *medium* o *large* (Lyons et al., 1988a e b; Maeda et al., 1985) in base al numero crescente di ripetizioni in tandem. Tali ripetizioni codificano per sequenze ricche in Prolina, Glutamina e Glicina. Un ulteriore allele, VL (*very large*), è stato descritto per i loci PRB1, PRB2 (Lyons et al., 1988a) e PRB3 (Azen et al., 1990). Alcune PRB sono espresse da più di un allele, ad esempio il peptide II-2 è espresso da tutti gli alleli noti del locus PRB1. Tuttavia è bene notare che alcuni degli alleli conosciuti non sono stati ancora sequenziati (PRB1VL, PRB2S, M, VL, PRB3VL). Le funzioni fisiologiche attribuite alle PRB sono:

- il legame del calcio nel cavo orale; II-2 e IB-1 legano il calcio mediante le loro serine fosforilate (Kauffman et al., 1991) e compaiono con elevate concentrazioni nel cavo orale di soggetti non affette da carie (Vitorino et al., 2005);
- la difesa dai tannini e dai polifenoli alimentari, in sinergia con le PRP glicosilate (Baxter et al., 1997; Beeley et al., 2001);
- la difesa nei confronti della flora batterica.

La concentrazione di IB-7 risulta altresì più elevata nella saliva di soggetti sani rispetto a individui affetti da carie (Ayad et al., 2000). Peptidi derivanti dalla frammentazione di bPRP sono efficaci nell'inibire la replicazione dell'*Herpes simplex*, interferendo con la

sua penetrazione nelle cellule bersaglio e/o col processamento intracellulare del virus stesso (Gu et al., 1995).

Un ulteriore contributo all'eterogeneità delle PRB è dato dalle glicosilazioni, da cui originano appunto le PRB glicosilate (gPRPs). Esse sono codificate essenzialmente dai loci PRB3 e PRB4, che codifica altresì per il peptide P-D e le sue varianti, unico prodotto non glicosilato espresso a partire da questi due loci. Tuttavia, sono stati individuati siti di O-glicosilazione anche nei peptidi Ps1 e Ps2, espressi rispettivamente dagli alleli *medium* e *large* di PRB1 (Vitorino et al., 2007; Carpenter et al., 1999). La prima descrizione delle gPRPs risale al 1979 per opera del gruppo di Edwin Azen (Azen et al., 1979), che denominò la glicoproteina e il locus genico corrispondente Gl.

Le PRB glicosilate si differenziano dalle altre PRB per la massiccia glicosilazione e la tipologia di sequenze ripetute in tandem, come evidenziato nel cDNA da Maeda e colleghi (Maeda et al., 1985) mediante digestione dei quattro loci PRB con l'enzima di restrizione BstNI. Le gPRPs sono soggette sia ad O-glicosilazione, per la quale non è stata ancora individuata alcuna sequenza consenso (Varki, 2009), che ad N-glicosilazione, mediante legame del glicano all'azoto ammidico di un'asparagina posta all'interno della sequenza NXS/T (dove X rappresenta qualsiasi aminoacido, prolina esclusa; Gavel & Heijne, 1990). Queste proteine hanno elevato peso molecolare, con gli zuccheri che possono rappresentare oltre il 50% del peso complessivo. Le gPRPs costituiscono la prima barriera contro i polifenoli alimentari e le infezioni patogeniche all'interno del cavo orale (Mehansho & Carson, 1983). La loro elevata eterogeneità, e in particolare la struttura dei glicani, sono fondamentali per il riconoscimento dei vari microorganismi che popolano il cavo orale (Amano et al., 1998; Gillece-Castro et al., 1991).

Stubbs e il suo gruppo di lavoro formularono, per la prima volta, una descrizione completa del complesso di geni PRP e delle proteine da essi codificate (Stubbs et al., 1998). Mediante l'allineamento è stato possibile assegnare le regioni geniche e determinare la sequenza aminoacidica delle proteine secrete. Ciò risultò importante al fine di prevedere i siti di taglio proteolitico della proteina precursore. Essi sono bersaglio di enzimi appartenenti alla classe delle convertasi, in altre parole proteasi responsabili del processamento dei precursori proteici, fondamentali per determinare il

fenotipo delle PRP. La comprensione di tale processo è importante per la conoscenza delle implicazioni funzionali delle PRP, in particolare per le PRB (Stubbs et al., 1998). Le PRB sono presenti nella saliva come piccoli peptidi (7-21 residui aminoacidici), poiché totalmente trasformate da numerosi tagli convertasici prima di essere immagazzinate nei granuli secretori. Questo taglio si può accompagnare alla rimozione del residuo C-terminale.

In particolare, uno studio focalizzato sui siti di taglio convertasico ha analizzato la reazione di proteolisi con la furina di una pre-proproteina ricombinante derivata da PRB4S; l'enzima coinvolto nelle reazioni *in vivo* però è ancora ignoto (Chan & Bennick, 2001).

La capacità delle PRB di interagire con i batteri e precipitare i tannini potrebbe essere legata proprio al taglio delle PRP precursori (Stubbs et al., 1998).

L'evoluzione dei geni delle PRP salivari umane è avvenuta mediante duplicazioni, conversioni, e crossing over omologhi e ineguali (Lyons KM, 1988a), che hanno in primo luogo variato il numero di ripetizioni in tandem, in particolar modo all'interno del terzo esone delle PRB: pertanto, le sequenze della maggior parte delle bPRP risultano simili e molto ripetitive. Esse sono la conseguenza di ricombinazioni omologhe ineguali all'interno del terzo esone (Azen et al., 1984; Lyons et al., 1988) che risultano in duplicazioni o delezioni delle ripetizioni (codificanti sequenze ricche in prolina, glutammina e glicina). Tale variabilità allelica, unita a fenomeni di splicing alternativo e modificazioni post-traduzionali (Maeda et al., 1985), determina il gran numero di PRP espresse a partire da soli sei geni. A complicare ulteriormente questo quadro genetico, interviene anche lo splicing alternativo degli mRNA, che è stato osservato per PRB1 (Maeda N, 1985a). È interessante notare che alcuni degli alleli noti non sono ancora stati sequenziali (PRB1VL, PRB2S, M, VL, PRB3VL), pertanto non se ne conosce il prodotto proteico, mentre di PRB1L è stata completamente determinata la sequenza aminoacidica (P04280) e solo parzialmente quella nucleotidica (Azen et al., 1993). E' inoltre notevole il fatto che alcune bPRP sono espresse da più di un allele: ad esempio, il peptide II-2 viene espresso da tutti gli alleli noti del locus PRB1. La maggior parte dei 13 loci risultano polimorfici in tutte le razze studiate (Lyons et al., 1988a; Azen, 1993). Inoltre, alcuni di essi risultano essere alleli nulli, caratterizzati dall'assenza delle corrispondenti PRP negli omozigoti (Azen, 1993). A livello genico, sono stati rilevati alleli nulli a frequenze relativamente elevate nel caso delle PRP basiche, con la

sola eccezione del polimorfismo Pc (Azen et al., 1993) che è stato rinvenuto a bassa frequenza nella popolazione, mentre non sono risultati numericamente significativi nel caso delle PRP acide. Inoltre, il gruppo di Azen (Azen et al., 1992) ha identificato individui portatori di un gene di fusione PRB2/1, derivante da un crossing over ineguale avvenuto tra i loci PRB1 e 2; la sua espressione è quantitativamente scarsa o nulla. Si potrebbero ipotizzare numerose spiegazioni riguardanti la mancata espressione di un gene, che includono mutazioni del promotore, eventi di splicing e mutazioni nel processamento dell'RNA o riarrangiamenti genetici (con perdita d'informazione, nel caso delle delezioni) che ne possano alterare la stabilità stessa.

La complessità delle PRP è costituita quindi su più livelli, a partire dai geni, passando per i trascritti fino ad arrivare ai prodotti proteici finali. In particolare, la presenza di sequenze ripetute in tandem nella sequenza genica e le estese frammentazioni dei peptidi espressi, che hanno luogo sia prima che dopo la secrezione, fanno delle PRP un puzzle di ardua ricostruzione. Nelle prossime sezioni verranno approfonditi i loci PRB, in particolar modo PRB1 in quanto oggetto di questo studio.

PRB4

Per quanto riguarda la funzione delle proteine glicosilate espresse da PRB4, è noto che II-1 è in grado di chelare i tannini vegetali per via della sua componente aminoacidica ricca in prolina e glicina, in grado di stabilire numerosi punti di contatto con questi polifenoli (Boze et al., 2010).

La proteina II-1 fu identificata per la prima volta da Kauffman e Keller ed ha caratteristiche di basicità meno marcate rispetto alle PRB aventi prefisso IB, risulta essere fosforilata (Kauffman et al., 1979), inoltre contiene sette potenziali siti di N-glicosilazione (Kauffman et al., 1993).

Il gruppo condotto da Maeda ha analizzato per primo l'allele *small* di PRB4 (PRB4S), riportando la sequenza nucleotidica tradotta del clone di cDNA cP6. In seguito, Lyons e colleghi hanno sequenziato il terzo esone dei tre alleli del PRB4 (Lyons et al., 1988). Gli stessi ricercatori hanno attribuito all'allele *large* le proteine CON 1 e CON 2 erroneamente, in quanto queste sono prodotti dei polimorfismi di PRB2 (Azen et al., 1993).

L'allele *small* è stato sequenziato genomicamente nel 1993 anche da Kim e dai suoi collaboratori (Kim et al., 1993).

Sequenze aminoacidiche parziali attribuite a PRB4, sono state pubblicate nel 1993 (Kauffman et al., 1993). La sequenza nucleotidica tradotta del terzo esone dell'allele *medium* è stata pubblicata dal gruppo di Azen nel 1996, mentre nel 1998 Stubbs ha riportato la sequenza nucleotidica del terzo esone dell'allele *medium* nullo (Azen et al., 1996; Stubbs et al., 1998). Sequenze aminoacidiche parziali di prodotti di PRB4 sono state riportate da Kauffman e collaboratori. Confrontando la sequenza aminoacidica parziale della proteina II-1 (Kauffman et al., 1993) con la sequenza genomica tradotta del terzo esone dell'allele PRB4M, sebbene l'omologia sia molto elevata (Lyons et al., 1988), non è stata trovata piena corrispondenza. Come visto per PRB3, l'eterogeneità delle PRP glicosilate derivanti dai loci PRB4 è complicata anche dalla loro poli-allelia, che determina polimorfismi di lunghezza della proteina.

Le varianti di lunghezza di PRB4 codificano per le seguenti proteine:

- PRB4S: PGA, la glicoproteina più piccola (sei siti di N-glicosilazione), oltre alle varianti *wild type* e Pro→Ala del peptide P-D (Manconi et al., 2015);
- PRB4M: II-1, una glicoproteina con ventuno amminoacidi in più rispetto a PGA, con sette possibili unità oligosaccaridiche; P-D e le sue varianti (Kauffman et al., 1993; Stubbs et al., 1998);
- PRB4L: P-D e le sue varianti, oltre a CD-IIg, che rappresenta la proteina più grande e maggiormente glicosilata espressa da questo locus, con otto siti di N-glicosilazione; (Manconi et al., 2015).

PRB3

Il locus PRB3 esprime delle PRP glicosilate descritte nel 1979, quando fu individuata la prima glicoproteina salivare, denominata GI esattamente come il locus genico corrispondente (Azen et al., 1979).

Le PRB3 sono ricche di N-glicosilazioni estremamente eterogenee, presentando strutture complesse altamente fucosilate (da cinque a sette fucosi) e spesso dotate di acidi sialici terminali (legati in $\alpha 2 \rightarrow 3$ al Galattosio). Inizialmente, furono descritte

quattro varianti di Gl: tre a mobilità elettroforetica elevata, chiamate Gl1, Gl2 e Gl3 (in ordine di PM decrescente) e una a bassa mobilità, denominata Gl4.

Alle varianti Gl descritte da Azen se ne aggiunsero altre quattro descritte da Minaguchi, rilevate nella popolazione orientale: Gl 4, 5, 6 e 7 (Minaguchi et al., 1981).

A seguito della riorganizzazione delle PRB (Maeda et al., 1985) in un cluster formato da quattro loci, le Gl sono state attribuite tutte al locus PRB3 (Lyons et al., 1988). L'eterogeneità delle gPRP derivanti dai loci PRB3 è complicata anche dalla loro poli-allelia, che determina polimorfismi di lunghezza della proteina, come visto anche per i loci PRB1, PRB2 e PRB4 (vedere sezioni precedenti).

Il sequenziamento del terzo esone di PRB3S e PRB3L (Lyons et al., 1988) ha consentito di attribuire ai due alleli l'espressione di Gl1 e Gl3 e di evidenziare, inoltre, l'esistenza degli alleli *medium* e *very large*.

La PRP glicosilata classificata come Gl2 è più corta del prodotto Gl1 di PRB3L, ma più lunga rispetto a Gl3, prodotta a partire da PRB3S. I risultati di uno studio condotto da Stubbs e collaboratori, hanno mostrato che la PRP glicosilata Gl2 sarebbe codificata da PRB3M (Stubbs et al., 1998).

In uno studio successivo, sono stati sequenziati quattro frammenti di Gl2, la cui sequenza aminoacidica è risultata coincidente con la traduzione in silico del locus PRB3M (Hardt et al., 2005).

L'analisi genomica di un individuo tipizzato elettroforeticamente Gl0, in altre parole portatore di un allele nullo, ha consentito di attribuire l'assenza di Gl a codoni di stop prematuri (Azen et al., 1990).

I quattro alleli caratterizzati producono le quattro varianti di lunghezza di Gl:

- Gl1 è codificata da PRB3L (Lyons et al., 1988);
- Gl2 da PRB3M (Stubbs et al., 1998);
- Gl3, codificata dal locus PRB3S;
- Gl4, infine, rappresenta l'espressione di PRB3VL, del quale non è ancora stato eseguito il sequenziamento: si sa solo che presenta 126 paia di basi aggiuntive rispetto a quelle riscontrate nel *locus* PRB3M, ossia due *tandem repeat* da 63 basi ciascuno (Azen et al., 1993).

Azen e i suoi collaboratori hanno inoltre riportato una variante Arg→Cys di PRB3S, sostituzione che consente la dimerizzazione di GI3 mediante legame disolfuro tra le cisteine, a formare un omodimero chiamato GI8 (Azen et al., 1990).

I prodotti di espressione del locus PRB3 sono stati ampiamente studiati per quanto riguarda l'interazione con la flora batterica del cavo orale e con i tannini.

Per quanto concerne la funzione di detossificazione dei tannini alimentari, per alcuni autori essa sarebbe possibile solo per le PRP de-glicosilate (Bennick et al., 2002; Lu et al., 1998).

Le proteine derivanti dalle varianti alleliche del locus PRB3 differiscono non solo per il numero di *tandem repeats*, ma anche per mutazioni a singolo nucleotide manifeste a livello fenotipico in sostituzioni amminoacidiche.

PRB2

Il locus PRB2 produce le proteine CON 1 e CON 2 e la proteina glicosilata IB-8a, la cui espressione è stata attribuita all'allele *large* (Azen et al., 1996). In base al sequenziamento parziale, la regione codificante IB-8a è stata assegnata alla proteina CON1, così chiamata per la capacità di legare la lectina concanavalina A (Debray et al., 1981).

La caratterizzazione del peptide P-J ha consentito di attribuirne la provenienza genetica dal locus PRB2.

Così come la proteina IB-8a si ritrova nella saliva intera in seguito al taglio dell'ultimo amminoacido C-terminale (Arg), anche P-J si ritrova nella saliva intera privo dell'Arg C-terminale. Dal peptide P-J, dopo l'azione di una carbossi-peptidasi sulle due serine carbossi-terminali, si originerebbe il peptide IB-7.

Mediante mappatura di restrizione, si è visto che, in tutti gli alleli, le differenze di lunghezza si presentano a livello del terzo esone del gene (Maeda et al., 1985). Inoltre, i siti di restrizione fiancheggianti il terzo esone risultano identici tra gli alleli, tranne un unico sito riconosciuto dall'enzima di restrizione *PstI*, presente nell'allele *large* di PRB2 e assente negli alleli *small* e *medium*. Tale differenza è stata attribuita alla possibile presenza di un sito di restrizione polimorfico (Lyons et al., 1988).

Inoltre, è stato osservato che la maggior parte degli alleli polimorfici varia in lunghezza per un multiplo di circa 60 bp (*base pair*), che corrisponde approssimativamente alla lunghezza dei *tandem repeat*, compresi all'interno del terzo esone (Lyons et al., 1988). Sebbene non siano state determinate le sequenze nucleotidiche per i tre alleli di PRB2, le loro mappe di restrizione indicano che essi variano in lunghezza per multipli di tre ripetizioni: PRB2S appare essere circa 180 bp più corto rispetto a PRB2L, misura corrispondente a circa tre ripetizioni. Invece PRB2VL sembra essere 15 ripetizioni più lungo rispetto a PRB2S (15 x 63 pb, circa 900 bp). Ciò suggerisce che gli alleli di PRB2, come pure di PRB1, differiscono in lunghezza per multipli di tre ripetizioni (Lyons et al., 1988).

PRB1

Il locus PRB1, che esprime PRB fondamentalmente non glicosilate (Azen et al., 1993), codifica per: il peptide II-2 (Kauffman & Keller, 1979) a partire dagli alleli *small*, *medium* e *large*; il peptide P-E (o IB-9, o PmF); PmS (IB-6; Azen & Denniston, 1980), codificato da PRB1S, che funge da proteina recettoriale nei confronti dell'adesione di *Candida albicans* sulla superficie d'idrossiapatite (O'Sullivan et al., 1997); PmF (IB-9) (Kauffman D et al., 1982), anch'esso a partire dall'allele *small*; Ps1 e Ps2 (Azen, 1993), derivanti rispettivamente dagli alleli *medium* e *large*; P-Ko (Halgand et al., 2012), dalla variante di splicing cP4, attribuito all'allele *large*.

E' stato dimostrato in diversi studi come la proteina II-2 sia coinvolta nella sensibilità al gusto amaro della 6-n-propiltiouracile (PROP; Cabras et al., 2012; Melis et al., 2013). La produzione di IB-6 e IB-9, inizialmente, è stata attribuita all'allele *small* (Kauffman et al., 1991). Tale attribuzione è stata poi confermata dal gruppo di Stubbs (Stubbs et al., 1998). In seguito ai risultati ricavati dal sequenziamento. La proteina Ps1, descritta e parzialmente analizzata dal gruppo di Azen (Azen et al., 1993), è stata attribuita al locus PRB1M.

Le proteine Ps furono inizialmente assegnate a PRB2, sulla base della sequenza aminoacidica (Lyons et al., 1988). Studi successivi hanno permesso di identificarle come prodotti di PRB1 (Azen et al., 1993).

E' stato denominato Ps0, invece, l'allele nullo (Azen et al., 1993).

Sia l'allele *medium* sia il *large* codificano, come già descritto in precedenza, il peptide II-2. Ps1, diversamente da Ps2 (Uniprot accession: P04280) è stato caratterizzato solo parzialmente (Azen et al., 1993), e non ne è stata ancora depositata la sequenza aminoacidica completa.

Sono state descritte almeno sei diverse proteine Ps, differenti per lunghezza (Azen e Denniston, 1980; Shintani et al., 1990) in base al numero di ripetizioni in tandem presenti nel terzo esone del locus PRB1 (Azen et al., 1993). Allo stato dell'arte, si ipotizza che queste differenze di lunghezza siano dovute, con ogni probabilità, a *crossing-over* ineguali, modificazioni post-traduzionali o codoni di stop prematuri.

Il gruppo condotto da Lyons (Lyons et al., 1988b) aveva correlato i fenotipi PmF e PmS con le varianti di lunghezza PRB1, deducendo che tutti gli individui con il fenotipo PmF⁺ / PmS⁺ presentano l'allele S e nessun individuo con l'allele S è stato tipizzato come PmF⁻ o PmS⁻.

E' stato dimostrato, inoltre, che la maggior parte degli individui con alleli PRB1M mostrano il fenotipo PmF⁻ / PmS⁻, ma alcuni mostrano il fenotipo meno comune PmF⁺ / PmS⁻. Questo suggerisce che alcune copie PRB1M possano contenere siti di taglio nella proteina precursore per generare il peptide PmF (Lyons et al., 1988).

La mutazione è localizzata in corrispondenza di un presunto sito di taglio proteolitico, presente nell'allele *small* (Lyons et al., 1988b).

La funzione delle proteine Ps è molteplice:

1. è stato dimostrato il legame con la microflora del cavo orale sia per Ps 2 (O'Sullivan et al., 1997), che per Ps 1 (Newman et al., 1996);
2. la detossificazione di tannini e polifenoli alimentari, funzione caratteristica delle PRB;
3. come II-2, anche Ps1 è coinvolta nella sensibilità al PROP (Cabras et al., 2012; Melis et al., 2013); *Streptococcus gordonii* è legato preferenzialmente dalle proteine IB-1, IB-9 e Ps1 (Newman et al., 1996);
4. è stato dimostrato (O'Sullivan et al., 1997 e 2000) che Ps2 (e probabilmente anche IB-6 e IB-8a) è capace di legare la *Candida albicans*.
5. Ps1 e Ps2 hanno anche la capacità di legarsi alle cellule della mucosa buccale, suggerendo un loro possibile coinvolgimento nella adesione batterica alla superficie della mucosa (Beeley & Newmann, 1995).

6. IB-6 e Ps1 sono coinvolte nel legame alla proteina capsidica gp-120 del virus HIV-1, inibendo così l'interazione del virus con i potenziali ospiti (Robinovitch et al., 2001), sebbene l'effetto specifico di ciascun peptide non sia ancora chiaro.

Nel lavoro svolto dal gruppo di Maeda, furono identificati tre mRNA generati dalla trascrizione di PRB1M: uno deriva dal normale *splicing* degli introni (cP3), mentre gli altri due (cP4 e cP5) da *splicing* alternativi (Maeda et al., 1985). Infatti, allineando le sequenze nucleotidiche dei tre cloni, è stato osservato che cP3, cP4 e cP5 sono identici, sebbene cP4 e cP5 presentino una delezione pari a, rispettivamente, 399 e 459 bp rispetto alla porzione centrale di cP3. Il confronto tra le sequenze nucleotidiche dei tre cloni mostra che le estremità 5' e 3' delle delezioni di cP4 e cP5 sono omologhe alle sequenze consenso per i siti donatori di *splicing* (C / A) AG: GT (A /G) AGTA e i siti accettori (C / T)₁₁N (C /T) AG: G (Mount et al., 1982).

Nel complesso, in cP3, sono stati individuati due possibili siti donatori di *splicing* alternativo nelle posizioni nucleotidiche 347 e 529 e otto potenziali siti accettori nelle posizioni 379, 439, 499, 562, 622, 682, 745 e 805. cP4 sembrerebbe il risultato di uno *splicing* dal nucleotide 347 al 745, e cP5 il risultato di uno *splicing* dal nucleotide 347 al 805.

In base all'omologia della sequenza nucleotidica all'estremità 5' di cP3 in confronto con quelle dei cloni cP1 e cP2 (attribuiti rispettivamente a PRH1 e PRH2), il gruppo di Maeda ha tentato di prevedere la struttura aminoacidica dei peptidi derivanti da cP3, cP4 e cP5 (Maeda et al., 1985). Le tre proteine di secrezione ipotizzate avrebbero una regione N-terminale di diciassette amminoacidi (probabilmente un residuo di glutamina in posizione N-terminale), seguita da un numero diverso di ripetizioni "B1, B2, B3", riconosciute dall'enzima di restrizione *BstNI*. Le ripetizioni B1 e B2 sono costituite ciascuna da 63 bp, corrispondenti a 21 amminoacidi; le prime tre ripetizioni B3 sono costituite da 57 bp ciascuna, cioè 19 amminoacidi; infine, la quarta ripetizione B3 è costituita da 60 bp, quindi 20 amminoacidi. L'insieme delle ripetizioni "B1-B2-B3" compare quattro volte in cP3. La sequenza nucleotidica della regione 3' di cP3 diverge da quella delle unità ripetute, sebbene la sua sequenza aminoacidica, derivante dalla traduzione *in silico* della sequenza nucleotidica, sia comunque ricca in prolina.

La sequenza aminoacidica di cP3, corrisponde a una PRB di 315 amminoacidi. Ciò suggerì che alcune delle PRB sequenziate fossero prodotte da un taglio post-traduzionale, a livello del legame peptidico tra una arginina e una serina di una proteina

precursore “cP3-like”. Un residuo di arginina ricorre nella medesima posizione in tutte le ripetizioni, tranne che in una delle ripetizioni B3 della proteina precursore cP3. Da studi condotti da Azen e collaboratori (1985) emerge che, qualora il taglio proteolitico si fosse verificato in corrispondenza di tutti questi residui di arginina (nelle posizioni 75, 136 e 259), cP3 avrebbe prodotto quattro diverse PRB. Si sarebbero generati peptidi di 75 residui, (da N1 e la prima unità B1-B2-B3), di 61 residui (dalla seconda unità B2-B2-B3), di 123 residui (due unità B1-B2-B3 dovuta alla sostituzione di arginina con glutammina in posizione 197) e di 56 residui (l'ultima ripetizione B1 e la regione C-terminale). I cloni cP4 e cP5, seguendo la stessa logica, genererebbero gli stessi peptidi N- e C-terminale di cP3. Inoltre, cP4 produrrebbe una PRB da 50 residui, mentre cP5 basico una PRB da 30 residui.

Le sequenze aminoacidiche dedotte dalle quattro unità “B1-B2-B3” di cP3 sono molto simili, sebbene non identiche, alle sequenze amminoacidiche delle piccole PRP basiche. Infatti, i residui da 87 a 197 sono identici a IB-9 e P-E (Kauffman et al., 1982; Isemura et al., 1982), fatta eccezione per la glutammina in posizione 197 di cP3, sostituita da un'arginina in IB-9 e P-E. I residui dalla posizione 198 alla 259 di cP3 sono identici a quelli presenti nelle posizioni corrispondenti in P-F (IB-8c; Saitoh et al., 1983a), con tre eccezioni: Ala-249 e Gln-256 di cP3 sono sostituite da prolina e arginina, rispettivamente, in P-F; mentre Arg-259 non è presente in P-F.

La sequenza aminoacidica di P-H (IB-5; Saitoh et al., 1983b) corrisponde a quella di cP3 dall'amminoacido 260 al residuo 315. L'unica differenza è situata all'N-terminale di P-H, dove un'alanina prende il posto di una serina nella posizione equivalente (residuo 260) di cP3.

Dal momento che, come visto, il sistema PRP è estremamente polimorfico, alcune di queste piccole differenze tra le proteine sequenziate e quelle previste da cP3, potrebbero essere dovute a variazioni genetiche ereditarie individuali.

Nella saliva intera e nei granuli secretori parotidei è stata rilevata, mediante spettrometria di massa, una proteina avente Mav (*average molecular mass values*) pari a 10,433.5 Da, denominata P-Ko (Halgand F. et al., 2012). Tale peptide è codificato da cP4, in seguito ad un processo di *splicing* alternativo dell'allele PRB1L. La pro-proteina derivante da traduzione *in silico* del clone cP4 non presenta, nella struttura primaria, gli amminoacidi dalla posizione 106 alla posizione 299 di PRB1L (*accession number* P04280) e il suo taglio genera il peptide II-2 e la proteina P-Ko.

Sono stati identificati eventi di *crossing-over* intergenico ineguale, che coinvolgono i loci PRB1 e PRB2. L'analisi delle delezioni e duplicazioni a carico di questi loci ha suggerito che tali eventi intergenici non siano rari, tenendo conto dell'organizzazione cromosomica della famiglia multigenica PRP. Infatti, PRB1 e PRB2 sono fisicamente più vicini tra loro (70 kb) rispetto a uno qualsiasi degli altri loci (Lyons et al., 1988b).

Il gruppo di Azen (Azen et al., 1992) ha identificato individui portatori di un gene di fusione PRB2/1 derivante da un *crossing-over* ineguale avvenuto tra i loci PRB1 e 2. Un *crossing-over* tra i geni PRB1 e PRB2 fu rilevato per la prima volta dai membri dell'*Utah Genetic Reference Project Pedigrees* (O'Connell et al., 1987).

Il gruppo di Lyons stabilì che un individuo omozigote per il gene di fusione PRB2/1 possedeva la regione in 5' del terzo esone identica alla regione 5' di PRB2, mentre la regione 3' risultava omologa a quella di PRB1 (Lyons et al., 1988b).

Al fine di determinare le basi molecolari del gene di fusione PRB2/1, quest'ultimo è stato clonato da un soggetto omozigote PRB2/1 e la sua sequenza nucleotidica comparata con quella dei geni PRB1 e PRB2 (Azen et al., 1992). È stato ipotizzato che il *crossing-over* tra i geni PRB2 e PRB1 avvenga in corrispondenza di una porzione altamente omologa a cavallo del terzo introne dei due loci, al determinando la perdita dei geni intatti PRB1 e PRB2 su entrambi i cromosomi negli omozigoti PRB2/1.

Il *crossing-over* conduce anche alla perdita di gran parte della regione codificante per PRB1 nel gene PRB2/1, mentre la regione codificante PRB2 rimane pressoché intatta.

Si è supposto, pertanto, che il *crossing-over* si verificasse in una regione di 743 paia di basi con solo un *mismatch* di una coppia di basi nel terzo introne dei geni PRB1 e PRB2. Peraltro, tale regione di omologia quasi completa è la più vasta individuata tra i geni PRP.

Da *western blot* sui peptidi contenuti nella saliva intera, si è osservato che nella saliva di individui omozigoti per PRB2/1, le PRP Pe, PmF, PmS e Ps non vengono prodotte espresse (Azen et al., 1992). Uno studio successivo ha individuato un codone di stop prematuro [CGA (Arg) → TGA (stop)] in corrispondenza del residuo 61 nella sequenza genica di PRB1, che spiega i fenotipi Ps0, Pe- e Pm- (Azen et al., 1993).

In seguito, è stato osservato (Azen et al., 1993) che l'individuo portatore dell'allele PRB1 / 2S mostra il fenotipo PmF⁻ / PmS⁻, derivante, in seguito ad una mutazione del DNA, da un codone di stop prematuro. Il *crossing-over* omologo frequente e ineguale all'interno del cluster genico PRP conduce a frequenti varianti proteiche per dimensione

(ad esempio le proteine G1, in seguito ad eventi intragenici) e alla formazione del gene di fusione PRB2/1, come conseguenza di eventi intergenici (Lyons et al., 1988).

Il gruppo di Lyons ha definito le sequenze genomiche che comprendono le ripetizioni in tandem all'interno del terzo esone, le quali sono state allineate per consentire comparazioni in ciascun allele di un singolo individuo (Lyons et al., 1988).

Il confronto tra le sequenze genomiche degli alleli *small* e *medium* di PRB1, mostra che essi differiscono in lunghezza per 3 ripetizioni, in modo tale che l'ordine dei tre tipi di elementi ripetuti "B1, B2 e B3" venga mantenuto passando da un allele all'altro. In base a tale evidenza, è stato proposto un modello per la generazione di PRB1S da PRB1M, che prevede un *crossing-over* e sostituzioni di 6 nucleotidi (fig. 2; Lyons et al., 1988).

Questo modello sarà discusso più approfonditamente nella sezione Risultati e Discussione.

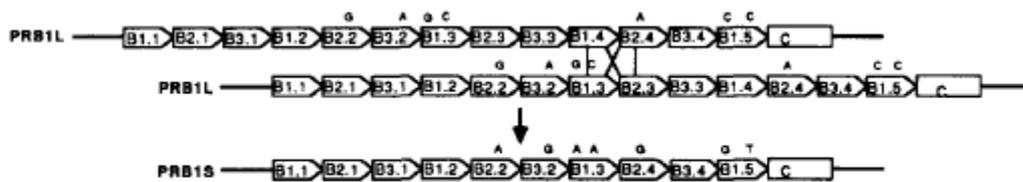


Fig. 2: modello per la generazione dell'allele *small* dalla ricombinazione omologa ineguale di due alleli di PRB1. Estratto dalla figura originale di Lyons et al., 1988a.

MATERIALI E METODI

Selezione dei soggetti

Gli individui, oggetto di questo studio, sono stati selezionati all'interno di un gruppo costituito da soggetti volontari, il cui profilo proteomico salivare è stato caratterizzato dal gruppo di ricerca della Prof.ssa Irene Messina mediante l'utilizzo delle tecniche di spettrometria di massa. Sulla base dei risultati emersi dal profilo proteomico relativo ai peptidi ascrivibili a PRB1 (Manconi et al., 2015), sono stati infatti ulteriormente selezionati tre individui (che chiameremo soggetto 1, soggetto 2 e soggetto 3), differenti dal punto di vista proteomico e rappresentativi delle tre varianti di lunghezza note per il gene oggetto di studio (tab. 1).

Nello specifico, il soggetto 1 presenta il seguente fenotipo biochimico: II-2, presente in tutti i soggetti in quanto espresso a partire da qualsiasi variante del gene (Manconi et al., 2015); Ps1, tipico della variante Medium (Azen et al., 1993); Pko, recentemente associato alla variante di splicing cp4 (Cabras et al. 2015; Maeda 1985); Ps2, tipico della variante Large (Azen et al., 1993). Questo profilo peptidico, sulla base dei dati riportati in letteratura, mette in evidenza una condizione di eterozigosi per le varianti Large e Medium.

Il soggetto 2 presenta i seguenti peptidi: II-2; Ps1 e Pko. In questo caso il profilo peptidico è attribuibile ad una condizione di omozigosi per la variante Medium.

Infine, il soggetto 3 presenta, oltre ai peptidi II-2, Ps1 e Pko, ricorrenti in tutti i soggetti analizzati, anche PE e IB6, tipici della variante Small (Azen et al., 1993). Sulla base dei dati presenti in letteratura il profilo peptidico è compatibile con una condizione di eterozigosi per la presenza delle 2 varianti Small e Medium.

	Profilo peptidico					Genotipo putativo
Soggetto 1	II-2	Ps1	Pko	Ps2		Medium/Large
Soggetto 2	II-2	Ps1	Pko			Medium/Medium
Soggetto 3	II-2	Ps1	Pko	PE	IB6	Medium/Small

Tabella 1: Profilo peptidico e genotipo putativo dei soggetti selezionati per la caratterizzazione molecolare

Estrazione DNA

Il DNA genomico dei soggetti selezionati è stato estratto da cellule della cavità buccale presenti nella saliva. Per ciascun soggetto, è stata effettuata la raccolta di 1 ml circa di saliva in apposite provette sterili da 1,5 ml. Il campione biologico, in attesa di essere processato, è stato conservato alla temperatura di 4°C tal quale, o sottoforma di pellet previa centrifugazione a 12000 rpm per 10'. Il DNA è stato purificato dalle cellule mediante l'utilizzo del *kit* commerciale *GeneMATRIX Tissue DNA Purification Kit* (EUR_X Ltd, Poland), seguendo un protocollo sperimentale in linea con le indicazioni fornite dal produttore, ma con piccole variazioni volte a migliorare la resa del processo. La metodica prevede l'utilizzo di *i*) reattivi in grado di produrre un lisato cellulare contenente molecole di RNA, polipeptidi, lipidi e contaminanti cellulari che vengono selettivamente eliminati, *ii*) colonnine cromatografiche ad adsorbimento munite di membrana di silice in grado di legare in maniera specifica le molecole di DNA.

In sintesi, dopo aver centrifugato il campione di saliva a 12000 rpm per 10', viene eliminato il surnatante ed il pellet, contenente le cellule, viene risospeso mediante vortex in presenza di 200 µl di *buffer* "Lyse T" (contenente sodio dodecil solfato) e 2 µl di RNasi A. Il campione viene quindi incubato per 5' a temperatura ambiente e successivamente trattato con 10 µl di Proteinasi K e 200 µl di *buffer* "Sol T", ed incubato a 70°C per 10' al fine di promuovere le condizioni sperimentali per la digestione e precipitazione dei contaminanti proteici e lipidici. Terminato il periodo di incubazione, il campione viene miscelato con 200 µl di etanolo 96%, e centrifugato per 1' a 12000 rpm al fine di sedimentare le macromolecole contaminanti presenti nel lisato. Il surnatante, contenente gli acidi nucleici, è successivamente trasferito nella

colonnina cromatografica ad adsorbimento (precedentemente equilibrata con l'apposito "buffer T"), la quale viene sottoposta a centrifugazione per 1' a 12000 rpm. Il DNA, legato alla membrana di silice, viene sottoposto a due serie di lavaggi, rispettivamente coi buffer "Wash TX1" e "Wash TX2" contenenti etanolo. Il passaggio finale dell'estrazione ha come finalità l'eluizione del DNA dalla colonnina ed il suo recupero in una provetta sterile. A tal fine si utilizza una soluzione contenente Tris-HCl 5mM, la quale è in grado di favorire il rilascio del DNA dalla fase fissa della colonnina. La concentrazione e purezza del DNA purificato è stata determinata spettrofotometricamente, misurando rispettivamente l'assorbanza a 260 nm (A 260) ed il rapporto tra i valori di assorbanza a 260 nm e 280 nm (A 260/280)

Primer Walking

Inizialmente, la sequenza selezionata come riferimento per la progettazione dei primers è stata quella depositata da E.A. Azen e collaboratori nel 1993 (S52986.1, presente in banca dati NCBI, fig. xx). In tabella 2 sono riportate le coppie di primers e le condizioni di amplificazione utilizzate nelle reazioni di PCR (*Polymerase Chain Reaction*). Le miscele di reazione sono state preparate utilizzando ≈ 20 ng di DNA purificato, *PCR Buffer 10x* contenente $MgCl_2$ 15 mM (5PRIME), dNTPs mix 200 μ M (SIGMA Aldrich), 20 pM di ciascun primer, 1.5 unità di DNA polimerasi (5PRIME PerfectTaq™ Plus DNA) in un volume finale di 25 μ l. Per ogni coppia di primers sono stati allestite almeno 3 reazioni utilizzando le medesime condizioni sperimentali, al fine di ottenere una idonea concentrazione del frammento di interesse, per successive purificazioni e reazioni di sequenziamento.

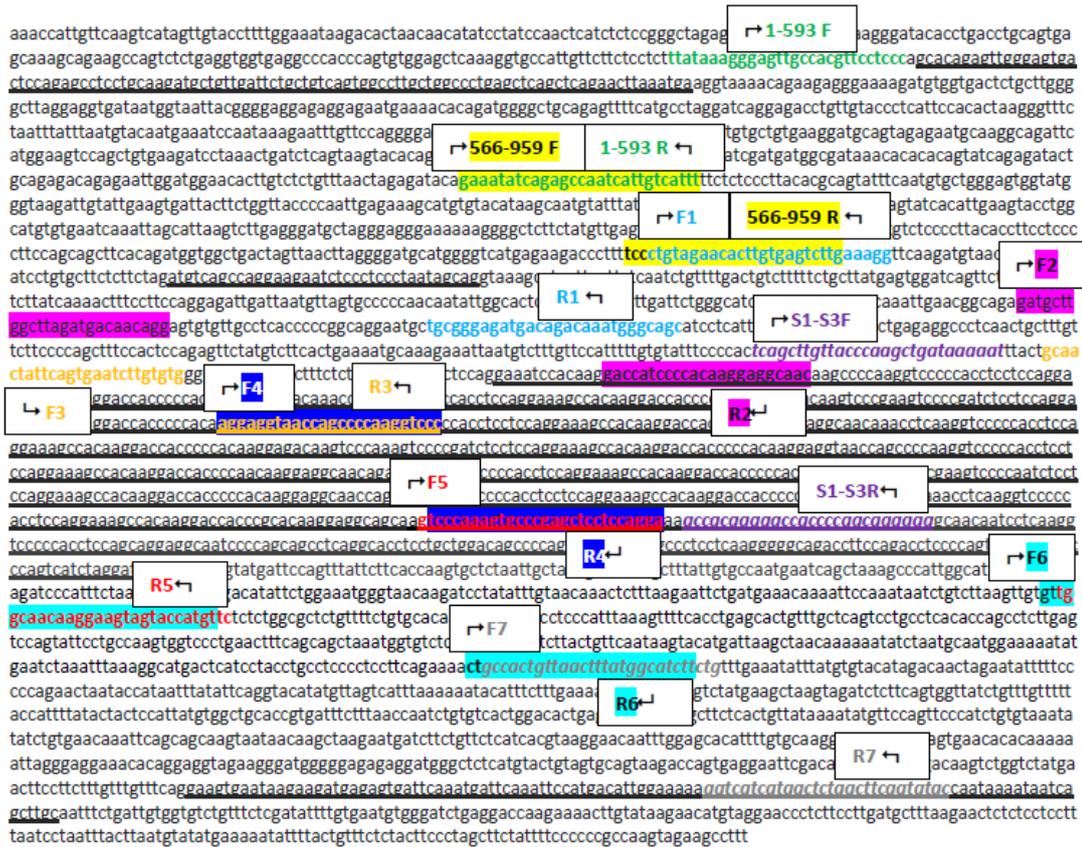


Figura. 2: posizione e orientamento dei primers sulla sequenza PRB1 di riferimento (*accession number* S52986.1). Le sequenze dei primers sono riportate nella tab. 2. I nomi dei primers sono indicati nei riquadri sopra o sotto le rispettive sequenze; le frecce rivolte a destra indicano i primers forward, quelle rivolte a sinistra i primers reverse. Le porzioni di sequenza sottolineate indicano gli esoni.

mediante un allineamento con Nucleotide Blast, la coerenza, in termini di contenuto nucleotidico, delle due sequenze depositate. Dal momento che i risultati dell'allineamento hanno evidenziato la presenza di porzioni nucleotidiche non perfettamente sovrapponibili, abbiamo deciso di procedere con la progettazione di un nuovo set di primers, disegnati sulla sequenza NT_187658.1 (Nature 2004, presente in banca dati NCBI, fig. xx), e utilizzabili in combinazione con il primerPRB F2 (precedentemente disegnati sulla sequenza S52986.1, fig. 2), perfettamente omologo sia alla sequenza S52986.1 che alla NT_187658.1. Le sequenze dei primers sono riportate nella tabella 3.

Coppie di primers	Condizioni di PCR	Frammento atteso
F1: ctgtagaacactgtgagctttaaagg R1: gctgccattgtctgtcatctcccga	94°C 3' 94°C 30'' 65°C 30'' 72°C 30'' 72°C 10'	385 bp
F2: gatgcttgcttagatgacaacagg R2: gttgctctctgtgggatggtc	94°C 3' 94°C 30'' 68°C 30'' 72°C 30'' 72°C 10'	359 bp
F3: gcaactattcagtaacttctgtg R3: gggacctggggctgttacctct	94°C 3' 94°C 30'' 62°C 30'' 72°C 30'' 72°C 10'	298 bp
F4: aggagtaaccagcccaaggtccc R4: tctggaggagctcggcactttggga	94°C 3' 94°C 30'' 68°C 30'' 72°C 30'' 72°C 10'	529 bp
F5: gtcccaagtccccgatctctccagga R5: tggcaacaaggaagttagtaccatgtc	94°C 3' 94°C 30'' 66°C 30'' 72°C 30'' 72°C 10'	482 bp
F6: gttggcaacaaggaagttagtaccatgt R6: ctgccactgtaactttatggcatctt	94°C 3' 94°C 30'' 67°C 30'' 72°C 30'' 72°C 10'	339 bp
F7: gccactgtaactttatggcatctctg R7: ggtcatcatagctctaactcaataac	94°C 3' 94°C 30'' 50°C 30'' 72°C 1' 72°C 10'	684 bp
S1-3F: tcagctgttaccacaagctgataaaaat S1-3R: ctcttgtgggtgtctctgtggc	94°C 3' 94°C 30'' 55°C 30'' 72°C 1' 72°C 10'	835 bp
1-593 F: ttataaaggagtgccacgttctccc 1-593 R: aaatgacaatgattggctctgatattc	94°C 3' 94°C 30'' 52°C 30'' 72°C 1' 72°C 10'	621 bp
566-959F: gaaatatcagaccaatcattgtcatt 566-959R: caagactcacaagttctacagga	94°C 3' 94°C 30'' 54°C 30'' 72°C 1' 72°C 10'	424 bp

Tabella 2: set dei primers usati per la prima fase del sequenziamento, disegnati sulla sequenza S52986.1 depositata nella banca dati NCBI. Per ciascuna coppia di primers sono riportate le rispettive condizioni di amplificazione e le dimensioni dei frammenti attesi. I primers forward sono abbreviati con F, analogamente i primers reverse sono abbreviati con R.

Le miscele di reazione, in presenza del nuovo set di primers (tab. 3), sono state allestite utilizzando i reattivi nelle concentrazioni e condizioni sperimentali, precedentemente descritte. Al fine di ottenere una idonea concentrazione del frammento di interesse, per le successive purificazioni e reazioni di sequenziamento, sono state preparate almeno 3 reazioni utilizzando le identiche condizioni sperimentali specifiche per l'amplificazione con ogni coppia di primers (tab 3).

Coppie di primers	Condizioni di PCR					Frammento atteso
PRBR1F: tctagaaattctctgtgtttctg PRBR2R: gaggacagcacactctgtgca	94°C 3'	94°C 30''	49°C 30''	72°C 20''	72°C 2'	213 bp
30 cicli						
PRBR3F: tgacacagtgtgatggcttgaca PRBR4R: aacaatggcacctttgagctccacac	94°C 3'	94°C 30''	58°C 30''	72°C 20''	72°C 2'	227 bp
30 cicli						
PRBR2F: tgcacagaagtgtgctctctc PRBR4R: aacaatggcacctttgagctccacac	94°C 3'	94°C 30''	50°C 30''	72°C 30''	72°C 5'	499 bp
30 cicli						
PRBR4F: gtgtggagctcaaagggtccattgt PRBR6R: taccataccactcccagcacattg	94°C 3'	94°C 30''	56°C 30''	72°C 45''	72°C 5'	706 bp
30 cicli						
PRBR6F: caatgtgctgggagtgtatgggta PRBR7R: ctagaagagaagcacaggatgatg	94°C 3'	94°C 30''	47°C 30''	72°C 30''	72°C 5'	418 bp
30 cicli						
F2: gatgcttgcttagatgacaacagg PRBR8R: gaacaaagacattaattatttgc	94°C 3'	94°C 30''	50°C 30''	72°C 20''	72°C 2'	203 bp
30 cicli						
PRBR8F: tgcaataaataatgtctttgtc PRBR9R: tctcctgtgatttctctg	94°C 3'	94°C 30''	47°C 30''	72°C 20''	72°C 2'	161 bp
30 cicli						
PRBR10F: ggggcagaccttcagacctccc PRBR11R: ttgaaaaatgagacaccatttagctgc	94°C 3'	94°C 30''	54°C 30''	72°C 30''	72°C 5'	480 bp
30 cicli						
PRBR9F: caggaaatccacaaggacca PRBR10R: gggaggctctggaaggtctgcccc	94°C 3'	94°C 30''	47°C 30''	72°C 20''	72°C 2'	710/893/1076 bp
30 cicli						
F2: gatgcttgcttagatgacaacagg PRBR10R: gggaggctctggaaggtctgcccc	94°C 3'	94°C 30''	64°C 30''	72°C 1'20''	72°C 10'	1031/1214/1397 bp
30 cicli						
PRBR11F: gcagctaaatggtgtctcattttcaa PRBR12R: gtattagtctggggaaaaatattc	94°C 3'	94°C 30''	47°C 30''	72°C 1'	72°C 2'	233 bp
30 cicli						
PRBR12F: gaatattttccccagaactaac PRBR13R: actgcactacagtacatgagagcc	94°C 3'	94°C 30''	50°C 30''	72°C 30''	72°C 5'	454 bp
30 cicli						
PRBR13F: ggctctcatgtactgtatgcagt PRBR14R: gaagtccagggaagtagagaaacag	94°C 3'	94°C 30''	49°C 30''	72°C 30''	72°C 5'	390 bp
30 cicli						

Tabella 3: set di primers utilizzati per la seconda fase del sequenziamento. I primers disegnati sulla sequenza NT_187658.1 sono siglati PRBR. Per ciascuna coppia di primers sono riportate le rispettive condizioni di amplificazione e le dimensioni dei frammenti attesi. I primers forward sono abbreviati con F, analogamente i primers reverse sono abbreviati con R.

I risultati ottenuti dall'analisi nucleotidica dei cloni di questa seconda serie di esperimenti ha consentito di ricostruire solo parzialmente le sequenze nucleotidiche del gene PRB1 dei tre soggetti. (vedi sezione risultati per i dettagli). Pertanto, al fine di conquistare le porzioni geniche mancanti, è stato utilizzato un nuovo approccio

sperimentale in cui la *Nested PCR* è stata abbinata al *Primer Walking*, utilizzando la sequenza di riferimento NC_000012.12 relativa al cromosoma 12 (assembly updated 12/05/2014; fig.xx). Da questa sequenza di riferimento, che costituisce l'*assembly* più recente e aggiornato prodotto dal consorzio genoma umano (Nature 2006, presente in banca dati NCBI), è stata estrapolata la sequenza nucleotidica codificante per PRB1 (nucleotidi da 11351823 a 11355590) e, ricorrendo al *tool primerBLAST* (tab. 4), è stato disegnato un nuovo set di primers in grado di ibridarsi a monte e a valle della porzione codificante del gene (tab 4, prima riga). Mediante reazioni di PCR è stato possibile ottenere un frammento la cui grandezza (4557 bp) rispecchiava quella attesa per l'intero gene, comprese una porzione non trascritta in 5' ed una in 3'. Il prodotto di PCR, è stato successivamente purificato da gel di agarosio e una volta concentrato, è stato utilizzato come substrato per le successive reazioni di *Nested PCR*, mediante la coppia di primers HOPE 1F-HOPE 2R (tabella 4). Per ogni coppia di primers sono stati allestite almeno 3 reazioni alle medesime condizioni, al fine di ottenere una concentrazione sufficiente per la successiva purificazione e sequenziamento del frammento di interesse. Ciascuna reazione è stata allestita esattamente coi medesimi volumi, reattivi e concentrazioni visti per il set di primers illustrato in tab. 2.

Coppie di primers	Condizioni di PCR	Frammento atteso
PRBR 1F: tctagaaatttcattggtgtttctg HOPE 4R: tcttgctcctcagatcccaca	94°C 3' 94°C 30'' 54°C 30'' 72°C 3'30'' 72°C 10' 30 cicli	4557 bp
HOPE 1F: tcagtgaattctgtgtgggct HOPE 2R: agatgactggggaggctgtc	94°C 3' 94°C 30'' 58°C 30'' 72°C 1' 72°C 10' 30 cicli	1154 / 971 / 788 bp
HOPE 4F: gaggaattcgacatttctgcc HOPE 4R: tcttgctcctcagatcccaca	94°C 3' 94°C 30'' 58°C 30'' 72°C 20'' 72°C 5' 30 cicli	232 bp

Tabella 4: set dei primers usati per la terza fase del sequenziamento. Il primer disegnato sulla sequenza NT_187658.1 è siglato PRBR, mentre i primers disegnati sulla sequenza NC_000012.12 sono siglati HOPE. Per ciascuna coppia di primers sono riportate le rispettive condizioni di amplificazione e le dimensioni dei frammenti attesi. I primers forward sono abbreviati con F, analogamente i primers reverse sono abbreviati con R.

gagatgacagaaatgggcagcatcctcattctgtctcctttatactgagaggccctcaactgcttggcttcccagctttccctccaggtttctatgtct
 tcaactgaaaatgcaataaataatgtctttgtccattttgtgtattccccactcagctgttaccgaagctgataaaa **HOPE 1F** attcagtgaa
tcttgtgtggctttactctgtctttctctctctttgtctccaggaatccacaaggaccatccccacaaggaggcaacaaggccccagggccccccacc
tctccaggaaagccacaaggaccacccccacaaggaggcaacaacctcaaggtccccacctccaggaaagccacaaggaccacccccacaaggg
gacaagtcccgaagtcccgatctcctccaggaaaaccacaaggaccacccccacaaggaggtaaccagcccccaaggtccccacctcctccaggaaag
ccacaaggaccacccccacaaggaggcaacagacctcaaggtccccacctccaggaaagccacaaggaccacccccacaaggagacaagtcccga
gtcctgatctcctccaggaaagccacaaggaccacccccacaaggaggtaaccaacccaaggtccccacctcctccaggaaagccacaaggaccac
ccccacaaggaggcaagaaacctcaggtccccacctccaggaaagccacaaggaccacccccacaaggagacaagtcccgaagtcccgaatctcctc
caggaaagccacaaggaccacccccacaaggaggcaaccagcccccaaggtccccacctcctccaggaaagccacaaggaccacccccacaaggagg
caacaacctcaaggtccccacctccaggaaagccacaaggaccacccgcacaaggaggcagcaagtcccgaagtcccgatctcctccaggaaagc
cacaaggaccacccccacaaggaggcaacaatcctcaaggtccccacctccagcaggaggcaatcccagca **HOPE 2R** tctctgtgacagc
cccaggaccacccacccctcctcaaggggcagacctccagacctcccagtgacagcctcccagtcata aggtatgattccag
 ttattctcaccgaagtgtctaatgtctacagctcctcagctttattgtgccaatgaatcagctaaaagcccattggcattgtatagctccagatcccatttct
 aaagattgtattgacatattctggaatgggtaacaagatcctatattgttaacaaactttaaaggaattctgatgttgaaacaaaattccaaataatc
 tgtcttaagtgtgtggcaacaaggagtagtaccatgtctctctggcctctgtttctgtgcacaaactgagagacctccatttaaagttttcacctgag
 cactgttctcagctcctcacaccagcctctgagtcagctattcctgccaagtggtccctgaacttcagcagctaaatgggtctcattttcaaattct
 tactgttcaataagtcacatgattaagctaacaaaaatcctaagcaatggaaaaatgaatctaaattaaaggcatgactcatctacctgcctcccc
 tcttcagaaaaactgccactgttaactttatggcactctctgtttgaaatattatgtgtacatagactactagaatattttccccagaactaataccataat
 ttatattcaggtacatattgttagctatttaaaaaatacatttttgaatattccacatacgtctatgaagtaagtagatctcctcagtggttatctgtttgtt
 ttaccattttatactactccattatgtggctgcaccgtgatttttaaccaatctgtgtcactggacactgagggtggtttcagcttctcactgttataaaat
 gttccagttcccactctgtgtaaatatctgtgaacaaattcagcagcaagtaataaagtaagaatgatcttctgtcttcatcacgtaaggaaacattg
 gagcacattttgtg **HOPE 4F** agagtgaacacacaaaaaattaggaggaaacacaggaggtagaagggatgggggagagaggatgggctc
tcatgtactgtagtg tgaggaaatcgacatttctgca atgtcaagctctgtctatgaacttctcttctgtttgttcaggaagtgaataa
gaagatgagagtgattcaaatgattcaaatccatgacattgaaaaaggtcatcatagctctaactcaatataccaataaaataatcagcttgc aattc
 tgaattgtgtgtctgttttcaatatttggaa gtgggatctgaggaccaagaa **HOPE 4R** acatgtagaacctcctcctgtatgctccaggaa
 ctctcctccttaactcaatattaccaggtgccatgaaaaatatttactgt

Fig. 4: posizione dei primers indicati in tab. 5 sulla sequenza NC_000012.12 (in *reverse-complement*). I nomi dei primers sono indicati nei riquadri accanto alle rispettive sequenze. Le porzioni di sequenza sottolineate indicano gli esoni, in questo caso sono riportati solo il terzo e quarto esone.

Gel elettroforesi

I frammenti di DNA ottenuti tramite PCR sono stati separati mediante elettroforesi su gel di acrilamide-bisacrilamide all'11% in TBE 10x, in presenza di marcatori di pesi molecolari noti. Le corse elettroforetiche sono state effettuate a 200 V costanti in TBE 1x, all'interno di una camera elettroforetica Labnet ENDURO VE10 E2010-PA.

La soluzione di TBE 10x, l'acrilamide al 99+%, la bis-acrilamide e la soluzione di TEMED al 99% sono state acquistate dalla SIGMA ALDRICH, mentre l'ammonio persolfato solido dalla GE Healthcare.

Al termine della corsa elettroforetica, previa incubazione del gel con un intercalante del DNA (bromuro di etidio, SIGMA ALDRICH), i frammenti di PCR sono stati visualizzati mediante transilluminatore UV (Uvifor Transilluminator, ELETTRIFOR) e

confrontati coi marcatori di peso molecolare utilizzati. Nello specifico, frammenti di dimensioni attese al di sotto delle 500 bp sono stati confrontati col plasmide puc18 digerito con l'enzima di restrizione *HaeIII*, quelli di dimensioni superiori alle 500 bp con il marcatore *Gene Ruler* 100bp (ThermoFisher). Il frammento di 4557 bp (tab. 4, prima riga), avendo dimensioni superiori al potere risolutivo del gel di acrilamide, è stato invece sottoposto a elettroforesi su gel d'agarosio (EUR_x) all'1,5% in TBE 1x, con aggiunta di SybrGreen (SIGMA ALDRICH) 1:10000 utilizzato come intercalante del DNA. La corsa elettroforetica è stata effettuata a 120V costanti in TBE 1x mediante camera elettroforetica Labnet ENDURO 10.10. Al termine della corsa elettroforetica, il gel d'agarosio è stato visualizzato mediante transilluminatore UV (Uvifor Transilluminator, ELETTROFOR) e confrontato con apposito marcatore di pesi molecolari noti, in questo caso il marker 1kp plus (ThermoFischer). Analogamente, sono stati valutati tutti i prodotti di PCR dopo la loro purificazione, condotta come illustrato nelle prossime sezioni.

Purificazione frammenti PCR

I frammenti di dimensioni attese sono stati purificati per eliminare i sali e i substrati non utilizzati nel corso della reazione di PCR (soluzione tampone, sali, deossiribonucleotidi, primers) ed eventuali prodotti aspecifici. Questa fase è propedeutica al successivo sequenziamento, ed è indispensabile per ottenere cromatogrammi privi di rumore di fondo.

I frammenti attesi, presenti in miscele prive di prodotti aspecifici, sono stati purificati mediante Exo-BAP Mix (EURx Ltd, Poland), che è impiegata per la purificazione enzimatica delle PCR propedeutica al sequenziamento. Questa consente infatti la rimozione rapida ed efficace di qualsiasi deossiribonucleotide e primer non consumato nella reazione di PCR. Il reattivo impiegato, costituito da specifici enzimi idrolitici, si aggiunge alla miscela di reazione PCR (2 µl di Exo-BAP ogni 5 µl di mix PCR). Dopo un periodo di incubazione a 37°C per 15', la temperatura va innalzata a 80°C per 15' al fine di inattivare gli enzimi idrolitici. La concentrazione dei frammenti purificati è stata determinata spettrofotometricamente, misurando l'assorbanza a 260 nm.

In presenza di frammenti aspecifici, i frammenti di taglia desiderata sono stati purificati mediante il GenElute Gel Extraction Kit (SIGMA ALDRICH) da gel d'agarosio in TBE 10x a diversa concentrazione (1.5% - 2%) a seconda delle dimensioni del frammento d'interesse. Le corse elettroforetiche sono state condotte a 120V costanti in TBE 1x mediante camera elettroforetica Labnet ENDURO 10.10. Al termine della migrazione, i gel sono stati visualizzati su transilluminatore UV (Uvifor Transilluminator, ELETTRFOR) e i frammenti d'interesse (individuati per confronto con opportuno marcatore di pesi molecolari noti) sono stati escissi dal gel ed inseriti in provette sterili da 2 ml, precedentemente pesate. La differenza tra il peso del volume vuoto e pieno della provetta costituisce il punto di riferimento per la determinazione del volume di Solubilization Buffer (utilizzato per distruggere la matrice del gel) da aggiungere in ciascuna provetta (3 volumi del buffer ogni 100 mg di agarosio). La miscela di agarosio e buffer deve essere incubata a 60°C fino a completa dissoluzione del gel. Successivamente, dopo aver aggiunto un volume di 2-propanolo ogni 100 mg di gel, la miscela viene caricata in colonnine da 700 µl (precedentemente equilibrate con apposito buffer), contenenti una membrana di silice in grado di legare il DNA, e centrifugata per 1' a 12000 rpm. A questo punto il frammento di interesse, legato alla membrana, viene sottoposto a due serie di lavaggi, con la "Wash Solution Concentrate G", per poi procedere ad eluizione del frammento purificato in Tris-HCl 5mM. Allo stesso modo è stato purificato anche il frammento da 4557 bp (tab. 4), utilizzato come substrato nelle reazioni di *Nested-PCR* illustrate nella sezione *Chromosome Walking*. La concentrazione dei frammenti purificati è stata determinata spettrofotometricamente, misurando l'assorbanza a 260 nm (A 260).

Sequenziamento

Nella prima fase del sequenziamento, ciascuno dei frammenti PCR della dimensioni attese, una volta purificato, è stato aliquotato in due diverse provette ed essiccato in presenza del primer forward e del primer reverse rispettivamente, al fine di ottenere la sequenza nucleotidica di entrambi i filamenti di DNA. Le reazioni di sequenziamento sono state eseguite dalla BMR Genomics (Padova). I frammenti purificati nella seconda e terza fase del nostro lavoro, sono stati inviati in fase liquida (anche in questo caso in

presenza dei rispettivi primers reverse e forward) alla Macrogene per il sequenziamento automatico. I cromatogrammi relativi a ciascun frammento sono stati analizzati mediante Chromas 2.4. Le sequenze dei soggetti, ricostruite mediante assemblaggio dei singoli frammenti di PCR sequenziati, sono state allineate tra loro mediante *Nucleotide Blast*, e confrontate mediante *Human Blast* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) con le rispettive sequenze di riferimento per ciascuna fase del sequenziamento.

Analisi in silico

L'analisi degli SNPs (*Single Nucleotide Polymorphism*), individuati nelle sequenze relative ai 3 soggetti, è stata effettuata con l'ausilio dello strumento grafico di *Nucleotide Blast* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). In sintesi, una volta impostata come query una sequenza comprendente circa 25 nucleotidi a monte e 25 a valle dello SNP che si vuole analizzare (oltre naturalmente alla sostituzione stessa), selezionata nel "Database setup" la voce "NCBI genomes (chromosome)" e impostato "Homo Sapiens" nella voce "Organism", si procede al *blast*. Nella pagina dei risultati, selezionando come di consueto dalla colonna "Descriptions" la voce riferita alla nostra query, si visualizza l'allineamento corrispondente. Successivamente, selezionando "Graphics" si viene rinviiati alla visualizzazione grafica della regione cromosomica corrispondente all'allineamento, e da qui alla voce "Configure" è possibile impostare la visualizzazione sia della nostra query di partenza che di qualsiasi eventuale variazione (siano esse SNPs, inserzioni o delezioni) registrata in banca dati per quella specifica regione.

La predizione di potenziali sequenze *Exon Splicing Enhancers* ed *Exon Splicing Silencers* è stata effettuata mediante lo strumento *Human Splicing Finder 3.0* (<http://www.umd.be/HSF3/>).

Le porzioni esoniche delle sequenze relative ai tre soggetti sono state tradotte virtualmente mediante l'apposito strumento del portale ExPASy (<http://www.expasy.org/>). Le sequenze aminoacidiche prodotte sono state allineate virtualmente mediante Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) con la sequenza di riferimento relativa al gene PRB1 P04280 dal portale UniProt (<http://www.uniprot.org/uniprot/P04280>).

RISULTATI E DISCUSSIONE

Amplificazione con il primo set di primers e purificazione dei cloni PCR

Il gene codificante per PRB1 consta di oltre 3,7 kbp, pertanto l'utilizzo del metodo enzimatico di Sanger, essendo efficace per frammenti di 1000 bp circa, non è applicabile utilizzando un solo passaggio. Per analizzare sequenze di dimensioni maggiori, si ricorre infatti a tecniche di *shotgun sequencing*, impiegate preferenzialmente per l'analisi di interi genomi. Considerando le limitazioni metodologiche della tecnica sviluppata da Sanger e al fine di ottenere la sequenza completa del gene di tutti i soggetti selezionati, si è scelto di utilizzare un approccio basato sulla metodica del *primer walking*. Grazie a tale metodica, è possibile suddividere la sequenza del gene di interesse dedotta dalla banca dati NCBI) in corte porzioni nucleotidiche (inferiori alle 1000 bp), adiacenti e parzialmente sovrapposte, in modo da poter ricostruire la sequenza intera del gene target per assemblaggio dei singoli cloni, in cui è stata artificialmente suddivisa. Nella metodica classica del *primer walking*, le prime coppie di primers sono progettate sulla base dei risultati ottenuti dall'analisi nucleotidica di uno o più cloni di sequenza non nota, isolati mediante l'utilizzo di sonde molecolari. A partire dalle regioni interne ed estendendosi a monte e a valle del frammento sequenziato, è possibile grazie all'impiego di nuove coppie di primers specifici, conquistare nuove porzioni nucleotidiche. Nel caso del gene PRB1, a differenza della metodica classica, non è stato necessario ricorrere a sonde molecolari, in quanto la sua sequenza è già stata pubblicata da Azen e collaboratori nel 1993 ed è disponibile in banca dati NCBI (accession: S52986.1). A causa dell'altissimo livello di omologia di sequenza intergenica, tipico della classe di geni PRP, oltre che dell'elevatissimo numero di sequenze ripetute in tandem intrageniche (Lyons et al., 1988a), non è stato possibile ricorrere alla progettazione, mediante software specifici, della maggior parte dei primers. Soltanto per gli oligonucleotidi riportati in tab.4, è stato possibile ricorrere agli strumenti bioinformatici e progettare coppie di primers con caratteristiche ideali, in termini di temperatura di melting e specificità, per il gene PRB1. Le condizioni ottimali di amplificazione per ciascuna coppia di primers, riportate in tabella 2, sono state ottenute ottimizzando opportunamente le condizioni sperimentali. In figura 5 sono mostrati i gel elettroforetici inerenti i cloni amplificati col primo set di primers utilizzato, disegnato sulla sequenza S52986.1 (fig. 2, tab. 2). Nel

pannello A sono riportati i risultati delle PCR per il soggetto 1, nel pannello B per il soggetto 2 e infine nel pannello C per il soggetto 3. La coppia di primers F1-R1, che copre la zona a monte e a valle del secondo esone (fig. 2), ha prodotto, nel caso dei soggetti 1 e 2 (fig. 5, rispettivamente lane 1A, 1B), un unico clone di 385 bp, le cui dimensioni rispecchiavano quelle del prodotto atteso. Vista l'assenza di prodotti aspecifici nelle due miscele di reazione, il clone d'interesse relativo ai due soggetti, è stato purificato con il reattivo *Exo-BAP Mix*. Nel caso del soggetto 3, il profilo elettroforetico ha evidenziato oltre al prodotto atteso di 385 bp anche un clone di dimensioni superiori, pertanto il prodotto di interesse è stato purificato da gel di agarosio al 2% (fig. 6A, lane 3). Per la coppia F2-R2, che delimita una regione di 359 bp a valle del secondo esone fino al 5' del terzo esone (fig. 2), relativamente ai soggetti 1 e 2 (fig. 5, rispettivamente lane 2A e 2B) è stato possibile osservare sia il clone di dimensioni attese che prodotti di dimensioni superiori e inferiori rispetto a quelle attese.

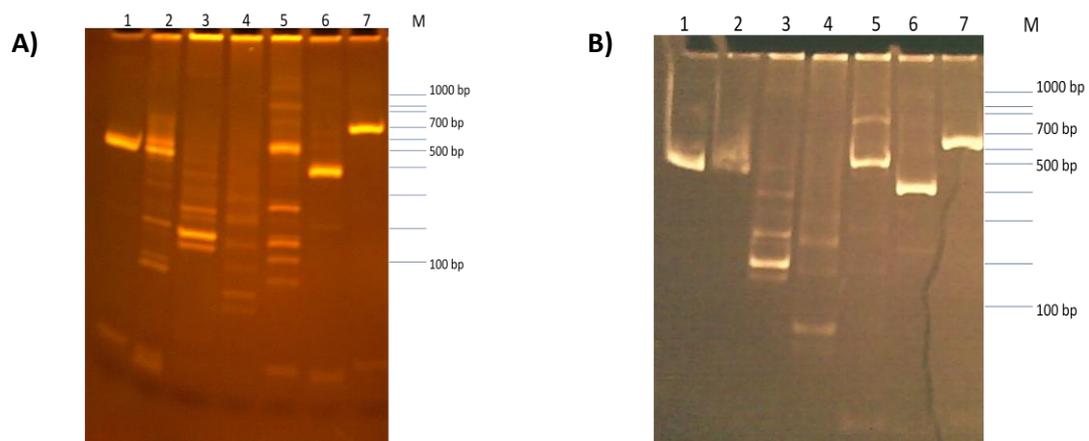


Fig. 5: PCRs PRB1, set di primers disegnati sulla sequenza S52986.1; elettroforesi su gels di poliacrilamide all'11%; **A)** soggetto 1: lane 1 = F1-R1, clone atteso 385 bp; lane 2 = F2-R2, clone atteso 359 bp; lane 3 = F3-R3, clone atteso 298 bp; lane 4 = F4-R4, clone atteso 529 bp; lane 5 = F5-R5, clone atteso 482 bp; lane 6 = F6-R6, clone atteso 339 bp; lane 7 = F7-R7, clone atteso 684 bp; marcatore (M) = *Gene Ruler* 100bp (ThermoFisher). **B)** soggetto 2; lane 1 = F1-R1, clone atteso 385 bp; lane 2 = F2-R2, clone atteso 359 bp; lane 3 = F3-R3, clone atteso 298 bp; lane 4 = F4-R4, clone atteso 529 bp; lane 5 = F5-R5, clone atteso 482 bp; lane 6 = F6-R6, clone atteso 339 bp; lane 7 = F7-R7, clone atteso 684 bp; marcatore (M) = *Gene Ruler* 100bp (ThermoFisher).

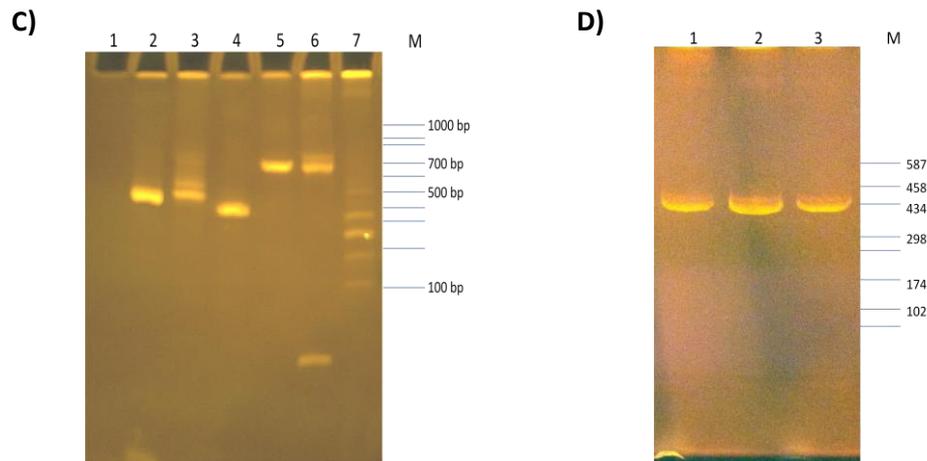


Fig. 6: PCRs PRB1, set di primers disegnati sulla sequenza S52986.1: elettroforesi su gels di poliacrilamide all'11%; **A)** soggetto 3; lane 1 = F4-R4, clone atteso 529 bp; lane 2 = F2-R2, clone atteso 359 bp; lane 3 = F1-R1, clone atteso 385bp; lane 4 = F3-R3, clone atteso 298 bp; lane 5 = F7-R7, clone atteso 684 bp; lane 6 = F5-R5, clone atteso 482 bp; lane 7 = F6-R6, clone atteso 339 bp; marcatore (M) = *Gene Ruler* 100bp (ThermoFisher). **B)** PCR PRB 566-959 F/R, il clone atteso ha dimensioni pari a 422 bp; lane 1= soggetto 1; lane 2= soggetto 2; lane 3= soggetto 3; il marcatore utilizzato (M) è il puc18 *HaeIII digested*

Quindi, il clone da 359 bp è stato in entrambi i casi purificato da gel di agarosio al 2%; per il soggetto 3 (fig. 6A, lane 2) non abbiamo invece cloni aspecifici, perciò in questo caso è stato possibile purificare mediante Exo-BAP Mix. Per la coppia F3-R3, che copre la regione a monte del terzo esone e una porzione dell'esone stesso (fig. 2), non è stato possibile purificare il clone di taglia attesa (298 bp), in quanto per tutti e tre i soggetti la PCR ha prodotto solo cloni aspecifici (fig. 5, lane 3° e 3B; fig. 6A, lane 4). Per quanto riguarda la coppia F4-R4, che comprende una porzione di 529 bp all'interno del terzo esone (fig. 2), la PCR non ha prodotto risultati per il soggetto 3 (fig. 6A, lane 1) e solo cloni aspecifici per i soggetti 1 e 2 (fig. 5, lane 4A e 4B), per cui, anche in questo caso, non è stato possibile isolare il clone di interesse. Nel caso dei primers 5F- 5R, che includono dalla regione 3' del terzo esone fino a una porzione del successivo introne (fig. 2), è stato ottenuto il clone di taglia attesa per i soggetti 1 e 2 (fig. 5, lane 5A e 5B rispettivamente) oltre ad alcuni cloni aspecifici, dunque il clone specifico da 482 bp è stato in entrambi i casi purificato da gel di agarosio al 2%. Per quanto riguarda il soggetto 3, la PCR ha prodotto invece solo cloni aspecifici (fig. 6A, lane 6), quindi non è stato possibile isolare il clone atteso. I primers F6-R6, comprendenti una regione all'interno del terzo introne (fig. 2), hanno prodotto il clone di taglia attesa (339 bp) per

tutti e tre i soggetti (rispettivamente, fig. 5, lane 6A e 6B; fig. 6A, lane 7), e in tutti e tre i casi la purificazione è stata fatta da gel di agarosio al 2% vista la presenza di cloni aspecifici. La coppia F7-R7, che copre la regione 3' del terzo introne fino all'ultimo esone (fig.2), ha prodotto il clone di taglia attesa (684 bp) per tutti e tre i soggetti (rispettivamente, fig. 5, lane 7A e 7B; fig. 6A, lane5), e in tutti e tre i casi, la purificazione è stata eseguita con Exo-BAP Mix. Le coppie di primers 1-593 F/R (comprendenti la regione immediatamente a monte del primo esone, fino a circa metà del primo introne, fig. 2) e S1-3F/R (che coprono una vasta porzione del terzo esone, fig 2) non hanno prodotto alcun risultato in nessuno dei soggetti presi in esame (gel non mostrati). Infine, la coppia PRB 566-959 F/R (che delimita una porzione da 424 bp del primo introne, fig. 2) ha prodotto il clone desiderato per tutti e tre i soggetti (fig. 6B, lane 1-3) che, vista l'assenza di cloni di taglia inattesa, è stato possibile purificare mediante Exo-BAP Mix.

Dai risultati ottenuti è immediatamente intuibile il fatto che questo set di primers non ci consenta di ricostruire la sequenza completa per nessuno dei tre soggetti, poiché il set di primers utilizzato non ci ha consentito di definire sia la porzione nucleotidica che si estende dalla regione 5'-UTR a circa metà del 1° introne, che quella includente una vasta porzione nucleotidica del 3° esone. Infatti, non è stato possibile amplificare e conseguentemente sequenziare alcuni dei cloni corrispondenti: F3-R3 (fig. 5, lane 3A e 3B; fig. 6, lane 4); F4-R4 (fig. 5, lane 4A e 4B; fig. 6, lane 1); 1-593 F/R e S1-3F/R (dato non mostrato). Nel caso del soggetto 3, la ricostruzione è stata ulteriormente complicata dall'impossibilità di definire, mediante l'utilizzo della coppia di primers F5-R5, la porzione 3' del 3° esone e la porzione 5' del 3° introne ad esso adiacente (fig. 3). E' plausibile ipotizzare che in corrispondenza della regione 5'-UTR fino alla metà del 1° introne del gene di tutti i soggetti analizzati (fig. 3), si sia verificata una mancata ibridazione del 3'-OH dei primers. Ciò a causa di un *mismatch* con lo stampo di DNA, tale da compromettere la fase di estensione da parte della DNA polimerasi. Tali *mismatch* sono verosimilmente determinati dalla presenza di polimorfismi a singolo nucleotide (SNPs, Single Nucleotide Polymorphisms) in corrispondenza delle regioni geniche complementari ai primers pocanzi descritti. Diverso il discorso, invece, per quanto riguarda le coppie F3-R3, F4-R4 e S1-3F/R (fig. 3), oltre che nella regione 3' del 3° esone e 5' del 3° introne per il soggetto 3 (PRB 5F-5R, fig. 3). In questo caso, bisogna tener conto del fatto che tutte e tre le coppie sono complementari alla zona del gene PRB1 a maggior contenuto di ripetizioni in tandem e di dinucleotidi GC, tutti

fattori che inevitabilmente influenzano in modo negativo l'ibridazione specifica dei primers e la fase di estensione del filamento da parte della DNA Polimerasi. Anche in questo caso, ovviamente, non si può escludere l'influenza di eventuali SNPs nella mancata o aspecifica ibridazione dei suddetti primers. La presenza di prodotti aspecifici in molte delle reazioni di PCR ottenute con questo set di primers (fig. 5 e 6), potrebbe essere inoltre ricondotta sia all'elevata omologia intergenica dei membri della classe PRB1 (Lyons et al., 1988a,b), che a ibridazioni aspecifiche in altre regioni del DNA.

Sequenziamento dei cloni amplificati col primo set di primers

Di seguito sono riportate le sequenze nucleotidiche del gene PRB1 per i tre soggetti, ricostruite mediante l'assemblaggio dei singoli cloni di PCR ottenuti mediante le amplificazioni con i primers riportati in tabella 2. Le sequenze risultano parziali, giacché non tutte le coppie di primers impiegate hanno permesso di ottenere i cloni attesi (vedere sezione precedente per i dettagli), questo anche a causa dell'impossibilità di analizzare alcuni dei cromatogrammi. In figura 7 è riportata la sequenza parziale di PRB1 relativa al soggetto 1, ricostruita dopo analisi dei cromatogrammi relativi al sequenziamento dei cloni PRB 566-959 F/R, F1-R1, F2-R2, F5-R5, F6-R6 e F7-R7. In una porzione dei cromatogrammi relativi ai cloni F2-R2 e F5-R5 non è stato possibile decodificare i nucleotidi corrispondenti, che in figura sono indicati con una "n" evidenziata in grigio, rispettivamente in posizione 2275 e 3431. Sulla sequenza parziale del soggetto 1 sono stati individuati una serie di SNPs, evidenziati in giallo (in totale 21),

Soggetto 1 (ref. S52986.1)

```
1293 gaaatatcagagccaatcattgtcatttttctctcccttacacgcagtatttcaatgtgctgggagtggtatgggtaa
1371 gattgtattgaagtgattacttctggttaccccaattgagaaagcatgtgtacataagcaatgtattataggaA/Ga
1447 tggagggcataagaacacccaaaA/GtatcacattgaagtacctggcatG/Ttgtgaactaaattagcattaagtcttg
1521 agggatgctagggagggaaaaaaggggctcttctatggttgagttcatggctggtgctctgtcataacaacctgT/Cc
1597 tcccttacaccttctccctccacagcagcttcacagatgggtggctgatgagttaacttaggggatgcatgggggtg
1675 ggtgagaagA/Tctgtttccctgtagaacacttgtagtcttgaagggttaagatgtaactgttcccatcatcctg
1751 tgcttctcttctagatgtcagccaggaagaatctccctccctaatagcaggtaaaT/GcctgattcA/Gttctcaatc
```

1825 tgtt **T/C**tgactgtctttttctgcttatga **A/G**tggatcagttctccagtgtcttcttatcaaaactttcctttc **A/G**
 1897 gga **A/G**ttgattaatgtagtgcccctaa **T/C**aatataggcactcttc **A/G**tgcaaccttgattctgggcatcatgag
 1969 caggccaccaaatgaacggcagagatgcttggttagatgacaacaggagtg **G/T**gttacctcatcccc **C/T**ggcc
 2043 agga **G/A**tgctgctgggagatgacagaaatgggcagcatcctcattctgtctcctctttataatgagaggcccttca
 2119 actgctttgttctccccagctttccactccagagttttatgttttcactgaaaaatgcaagaaattaatgtctttgt
 2197 tccatttttgtgtatttccccactcagcttgttacccaagctgataaaaatttactgcaactattcagtgaaatcttgt
 2275 ntgggcttttactcatgtcttcat **C/G**tctctctctttgtcctccaggaaatccacaaggaccatccccacaaggagg
 2351 caac
 3022 gtcccaagtgccgatctcctccaggaaagccacaaggaccaccccaacaagaaggcaacaatcctcaagggtcccc
 3100 acctccagcaggaggcaatccccagcagcctcaggcacctcctgctggacagccccagggaccaccacgcctcctca
 3178 agggggcagacctccagacctcccagtgacagcctcccagtcatttaggattcaatgacaggtatgattccagtt
 3256 tattcttcaccaagtgtcttaattgtctacagctctccagctttattgtgccaatgaatcagctaaaagccattggca
 3334 ttgtatagtcccagatcccatttctaaagatttgtattgacatattctggaaatgggtaacaagatcctatatttgt
 3412 acaaactctttaaggaatt **n**tgatggtgagaaacaaaattccaaataatctgtcttaagttgtgttgcaacaaggaa
 3490 gtagtaccatgttctctctggcgtctgttttctgtgcacaaactgagagacctccatttaaagttttcacctgagc
 3568 actgtttgctcagtcctgcctcacaccagcctcttgagtcagatattcctgccaagtggccttgaactttcagcagc
 3646 taaatgggtgtctcatttttcaaattcttactgttcaataagtacatgattaagctaacaacaaaaatctaatg **G/C**aa
 3722 tggaaaaatgatgaatctaaatttaaggcatgactcatcctacctgcctcccctcctcagaaaaactgccactgttaa
 3800 ctttatggcatcttctgtttgaaatatttatgtgtacatagac **A**actagaatatttttccccagaactaatactata
 3878 atttatattcagtgacatattgttagcttttaaaaaatacatttctttgaaaatttccacatacgtctatgaagctaa
 3956 gtagatctcttcagtggttatctgtttgtttttaccattttatactactccattatgtggctgcaccgtgatttcttt
 4034 aaccaatctgtgtcactggacactgagggtggtttcagcttctcactgttataaaatagttccagttccatctgtg
 4112 taaatatactgtgaacaaattcagcagcaagtaata **C/T**aagctaagaatgatcttctgtcttcatcacgtaaggaa
 4188 acaatttgagcacattttgtgcaagggcatccaaagagtgaacacacaaaaaattagggaggaaacacaggaggtag
 4266 aagggat **G/A**ggggagagaggatgggctctcatgtactgtagtgcagtaagaccagtgaggaattcgacatttctgc
 4342 catgtcaagtctggtctatgaacttcttctttgtttgtttcaggaagtgaataagaagatgagagtgattcaaatga
 4420 ttcaaattccatgacattggaaaaaggtcatcatagctctaacttcaatatac

Fig. 6: sequenza del gene PRB1 relativa al soggetto 1; sono indicate le posizioni, rispetto alla sequenza di riferimento con *accession number* S52986.1, delle porzioni di sequenza ricostruita; le porzioni esoniche sono riportate sottolineate; col carattere “n” ed evidenziati in grigio sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/).

la maggior parte dei quali è presente in eterozigosi e nessuno localizzato a livello esonico. Da sottolineare il fatto che le sequenze esoniche ricostruite sono parziali per il terzo esone (come indica la lacuna nucleotidica illustrata in fig. 7 in posizione 2017-3022) ed il 4° esone (interrotto a partire dal nucleotide 4465, fig. 8), mentre il 1° esone è del tutto assente (fig. 2).

Analogamente a quanto visto per il soggetto 1, in figura 8 è riportata la sequenza parziale di PRB1 relativa al soggetto 2, ricostruita mediante l'analisi dei cromatogrammi relativi al sequenziamento dei cloni PRB 566-959 F/R, F1-R1, F2-R2, F5-R5, F6-R6 e F7-R7. All'interno del secondo introne, nella regione di sovrapposizione tra i cloni F1-R1 e F2-R2 (fig. 2), l'analisi dei cromatogrammi ottenuti dal sequenziamento dei cloni F1-R1 e F2-R2 non ci ha permesso di assegnare ad un picco di fluorescenza il corrispondente nucleotide (fig. 8, "n" evidenziata in grigio). Analogamente, dall'analisi dei cromatogrammi relativi al clone F2-R2, non è stato possibile assegnare a tre picchi di fluorescenza i corrispondenti nucleotidi, sempre all'interno del 2° introne, sono stati individuati nei cromatogrammi relativi al clone F2-R2 (fig. 8). Inoltre, nel terzo introne sono presenti altri quattro nucleotidi che non è stato possibile assegnare al picco corrispondente: uno tramite individuato nel corso dell'analisi dei cromatogrammi relativi al clone F5-R5, gli altri tre individuati nei cromatogrammi relativi al clone F7-R7 (fig. 8). Anche nella sequenza parziale del soggetto 2 sono stati individuati una serie di SNPs (in totale 18), tutti in eterozigosi: 17 sono intronici ed uno è localizzato all'interno del quarto esone (fig. 8). Così come discusso per il soggetto 1, le sequenze esoniche ricostruite sono parziali per il terzo esone (come indica la lacuna nucleotidica presente in fig. 8 tra i nucleotidi 2017 e 3022) ed il 4° esone (interrotto a partire dal nucleotide 4465, fig. 8), ed inoltre la sequenza nucleotidica relativa al 1° esone è assente (fig. 2).

La sequenza parziale di PRB1 relativa al soggetto 3 è riportata in figura 9. Come osservato per gli altri due soggetti, anche in questo caso la ricostruzione è stata fatta in analizzando i cromatogrammi relativi al sequenziamento dei cloni PRB 566-959 F/R, F1-R1, F2-R2, F6-R6 e F7-R7. Per quanto concerne il clone F5-R5 (fig. 5C, lane 6), non avendo ottenuto risultati nel corso delle reazioni di PCR, non è stato possibile procedere con le reazioni di sequenziamento. In due punti dei cromatogrammi relativi al clone F2-R2 sono stati individuati due picchi di fluorescenza non ben definiti e pertanto non attribuibili a nessun nucleotide. (fig. 9, "n" evidenziata in grigio).

Soggetto 2 , a partire dal nucleotide 1293 (ref. S52986.1)

1293
gaaatatcagagccaatcattgtcatttttctctcccttacacgcagttttcaatgtgctgggagtggtatgggtaa
1371
gattgtattgaagtgattacttctggttacccaattgagaaagcatgtgtacataagcaatgtatttatagga **A/G**
1447
tggagggcataagaacacccaaa **G/A** taccacattgaagtacctggcat **G/T** tgtgaaactaaattagcattaagtcttg
1521
agggatgctagggagggaaaaaaggggctcttctatggttgagttcatgggctgttgcctctgtcacaacacctg **T/C**
1597
tccccttacaccttctcccctccagcagcttcacagatgggtggctgac **T/G** agttaacttaggggatgcatgggg
1673
gtg **A/G** tgagaag **A/T** ctgttttccctgtagaacactgtgagctctgaaaggtcaagatgtaacttttccatcat
1747
cctgtgcttctctcttagatgtcagccaggaagaatctcccctccctaataagcaggtaaagcctgattcgttctcaatc
1825
tgttttgactgtcttttctgcttatgagtggtacagttctccagtgcttcttatcaaaactttcctttcaggagtt
1903
gattaatgtagtgcccctaacaatataggcactcttcgtgcaaccttgattctgggcatcatgagcaggccacccaaa
1981
ttgaa **C/T** tggcagagatgcttggcttagatgacaacaggagtggttgctca **n** cccccggccaggaatgcctgctg
2057
ggagatgacagaaatgggcagcatcctcattctgttctctttataatgagaggccctcaactgctttgtttttccc
2135
cagcctttccattcagagttttatgttttactgaaaatgcaaagaaattaatgtctttgttccatttttgtgtatt
2213
cccactcagcttgttacccaagctgataaaaaatt **T/C** agtgca **A/G** ctattcagtgaaatcttgt **n** **A/T** gggctttta
2285
ctctgtct **C/T** actc **n** **n** tctctttgtcctccaggaaatccacaaggaccatccccacaaggaggcaac

3022
gtcccaaagtgccgatctcctccaggaaagccacaaggaccaccccaacaagaaggcaacaatcctcaagggtcccc

3100
acctccagcaggaggaatccccagcagcctcaggcactcctgctggacagccccaggaccaccagcctcctca

3178
agggggcagacctccagacctcccagtgacagcctcccagtcatttaggattcaatgacaggtatgattccagtt

3256
tattcttcaccaagtgtcttaattgtctacagctctccagctttattgtgccaatgaatcagctaaaagccattggca

3334
ttgtatagtccca **n** atcccatttctaagatttgtattgacatattctggaaatgggtaacaagatcctatatttgt

3412
acaaactctttaaggaattctgatgttgagaacaaaattccaaataatctgtcttaagttgttggcaacaaggaa

3490
gtagtaccatgttctttctggcgtctgttttctgtgcacaaactgagagacctccatttaaagttttcacctgagc

3568
actgtttgctcagtcctgcctcacaccagcctcttgagtcagtttctctgccaagtggtccctgaactttcagcagc

3646
taaatggtgtctcatttttcaaatcttactgttcaataagtacatgattaagctaacaaaaaatatctaagcaatg

3724
gaaaaatatgaatctaaatttaaaggcatgactcatcctacctgcctcccctcccctcaaaaaactgccactgttaact

3802
ttatggcatcttctgtttgaaatatttatgtgtacatagactactagaatatttttccccagaactaatactataat

3880
ttatattcaggtacatatgttagtcatthaaaaatacatttctttgaaaatttccacatacgtctatgaagctaagt

3958
agatctctcagtggttatctgtttgtttttaccatthtataactactccattatgtggctgcaccgtgatttctttaa

4036
ccaatctgtgtcactggacactgaggggtggtttcagcttctcactgttataaaatggttccagttccatctgtgta

4114
aatatatctgtgaacaaattcagcagcaagtaata **T/C** agctaagaatgatcttctgtcttcatcacgtaaggaa

4190
aatttggagcacattttgtgcaagggcatccaaagagtgaacacacaaaaaattaggaggaaacacaggaggtagaa

4268
qqat **G/A** gggggagagaggatgggctctcatgtactgt **n** tgcagt **A/T** agaccagtgaggaattcaacatttcc **T/**
4340
Ggccatgtcaagtctggctcatgaacttctt **G/C** tttgtttgtttcaggaagtgaata **n** aggatgagagtgatt **n**

4340

aatgattcaaatG/Tccatgacattggaaaaaggtcatcatagctctaacttcaatatac

Fig. 7: sequenza del gene PRB1 relativa al soggetto 2; sono indicate le posizioni, rispetto alla sequenza di riferimento con *accession number* S52986.1, delle porzioni di sequenza ricostruita; le porzioni esoniche sono riportate sottolineate; col carattere “n” ed evidenziati in grigio sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/).

Soggetto 3, a partire dal nucleotide 1293 (ref. S52986.1)

1293

gaaatatcagagccaatcattgtcatttttctctcccttacacgcagtatttcaatgtgctgggagtggtatgggtaa

1371

gattgtattgaaA/Gtgattacttctggttacccaattgagaaagcatgtgtacataagcaatgtatttataggaga

1447

tggagggcataagaacacccaaaA/GtatcacattgaagtacctggcatT/Gtgtgaactaaattagcattaagtcttg

1521

agggatgctagggagggaaaaaaggggctcttctatggttgagttcatggctggtgctctgtcataacaacctgC/Tc

1597

tccccttacaccttctcccctccagcagcttcacagatgggtgctgatgagttaacttaggggatgcatgggggtg

1675

cG/AtgagaagA/Tctgtttccctgtagaaccttgtagtcttgaaaggttcaagatgtaactgttcccatcatcc

1749

tgtgcttctcttctagatgtcagccaggaagaatctccctccctaataagcaggtaaaT/GcccgatcA/Gtttctcaa

1823

tctgttC/TtgactgtcttttctgcttatgaA/GtggatcagttctccagtgcttcttatacaaaactttcctttcG

1897

Y/AggagttgattaatggttagtgcccctaaT/CaatataggcactcttcA/Gtgcaaccttgattctgggcatcatgag

1969

cagccaccaaatgaacggcagagatgcttgcttagatgacaA/Gcaggagtggtgctcctcatcccccgccag

2045

gaAtgcctgctgggagatgacagaaaatgggcagcatcctcattctgtctcctctttataatgagagccctcaactgc

2123

tttgcttcccaggcggtccactccagagttntatgtcttctactgaaatgcaaagaaattaatgtcttggtttcatt

2201

tttggtgtttcccactcagcttgttaccctaaActgataaaaattC/TactgcaactatT/AcagagaatcttG/Ct

2273

cagggatntactcagtcctcAtctctctctctttgtcctccaggaaatccacaaggaccatccccacaaggaggcaac

3467

gttgccaacaaggaagtagtaccatgttctctctgctgctctgttttctgtgcacaaaactgagagacctcccatttaa

3545

agttttcacctgagcactgtttgctcagtcctgcctcacaccagcctcttgagtcagatattcctgccaagtggtccc

3623

tgaactttcagcagctaaatgggtgtctcatttttcaaattcttactgttcaataagtagatgattaagctaacaaaaa

3701

atatctaataatgcaatggaaaaatatgaatctaaatttaaaggcatgactcatcctacctgcctcccctccttcagaaaa

3779

ctgccactgtaaactttatggcatcttctggttgaaatatttatgtgtacatagacactagaatatttttccccaga

3857

actaatacC/Tataatttatattcaggtacatatgtagtcatttaaaaaatacatttctttgaaaatttccacatac

3933

gtctatgaagctaagtagatctctcagtggttatctggttggttttaccattttatactactccattatgtggctgc

4011

accgtgatttctttaaccaatctgtgtcactggcactgagggtggtttcagcttctcactgtttataaaatagtctcc

4089

agttcccatctgtgtaaatatatactgtgaacaaattcagcagcaagtaataaT/Caagctaagaatgatcttctgtct

```

4165 tcatcacgtaaggaacaatttggagcacattttgtgcaagggcatccaagagtgaacacacaaaaattagggagga
4243 aacacagggaggtagaagggatG/AggggagagaggatgggctctcatgtactgT/Gaatgcagtaagaccagnagga
4317 attcnacatttctctgG/Ccatggcaagtctggtctatgaacttccttctttggttgtttcaggaagtgaataagaaga
4393 tgagagtgattcaaatgattcaaatccatgacattggaaaaaggtcatcatagctctaacttcaatatac

```

Fig. 8: sequenza del gene PRB1 relativa al soggetto 3; sono indicate le posizioni, rispetto alla sequenza di riferimento con *accession number* S52986.1, delle porzioni di sequenza ricostruita; le porzioni esoniche sono riportate sottolineate; col carattere “n” ed evidenziati in grigio sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/).

Altri due picchi dello stesso tipo sono presenti nei cromatogrammi inerenti il clone F7-R7. Anche nella sequenza parziale del soggetto 3 sono presenti una serie di SNPs (in totale 24), la maggior parte presenti in eterozigosi e nessuno di questi localizzati a livello esonico.

In assenza di informazioni relative al sequenziamento del clone F5-R5, la porzione del gene PRB1 ricostruita per questo soggetto risulta essere ancora più corta rispetto alle sequenze degli altri due soggetti. Infatti, in questo caso, la lacuna nucleotidica in corrispondenza del terzo esone si estende dal nucleotide 2017 al 3467 (fig. 9). Anche in questo caso, il quarto esone risulta incompleto (interrotto a partire dal nucleotide 4465, fig. 9), e le informazioni relative al primo esone sono del tutto assenti (fig. 2).

Amplificazione con il secondo set di primers e purificazione dei cloni PCR

Per completare la sequenza dei tre soggetti si è quindi reso necessario disegnare un secondo set di primers, stavolta scegliendo come riferimento la sequenza con *accession number* NT_187658.1 (Nature, 2004). Confrontando quest’ultima sequenza con la S52986.1, mediante l’utilizzo di *Nucleotide Blast* (<https://blast.ncbi.nlm.nih.gov>; dato non mostrato), sono emerse delle differenze dovute ad inserzioni e delezioni che ne influenzano l’omologia, la quale è risultata pari al 94%. Queste differenze sono probabilmente risultate determinanti ai fini dell’amplificazione, in quanto possono aver precluso l’ibridazione corretta di alcune delle coppie di primers progettate sulla sequenza nucleotidica di riferimento S52986.1 (fig. 2). Si è quindi optato per la

progettazione di un nuovo set di primers, che ci consentisse di amplificare tutto il gene, e non solo le lacune nucleotidiche, frutto dell'insuccesso delle precedenti amplificazioni. Sono stati progettati, inoltre, nuovi primers da utilizzare in associazione con alcuni oligonucleotidi che, pur essendo stati disegnati sulla sequenza S52986.1, risultassero perfettamente omologhi alla nuova sequenza di riferimento NT_187658.1 (fig. 3). Analogamente a quanto visto per il set di primers riportato in tabella 2 e figura 2, le condizioni ottimali di amplificazione per ciascuna coppia di primers (tab. 3) sono state ottenute in seguito a una serie di esperimenti di PCR condotti con parametri sperimentali differenti.

Nella figura 9 sono mostrati i profili elettroforetici relativi ai prodotti delle reazioni di PCR realizzate coi primers PRBR1F-PRBR2R (pannelli A e B) e PRBR3F-PRBR4R (pannello C). Entrambe le coppie delimitano la regione a monte del primo esone del gene PRB1 (fig. 3). In tutti e tre i soggetti sono stati ottenuti i cloni di dimensioni attese per entrambe le coppie di primers e, vista la presenza anche di cloni aspecifici, è stato necessario purificare in entrambi i casi il frammento d'interesse da gel di agarosio al 2%. Nella figura 10 sono riportate i profili elettroforetici relativi ai cloni PRBR2F-PRBR4R (pannelli A e B) e PRBR4F-PRBR6R (pannelli C e D). La prima coppia di primers, analogamente alle coppie PRBR1F-PRBR2R e PRBR3F-PRBR4R, delimita la regione a monte del primo esone (fig. 3) ed ha permesso di amplificare il clone di dimensioni attese nel caso di tutti e i tre soggetti (fig. 10, lane 1A, 2A e 1B). Anche in questo caso, la presenza di cloni aspecifici ha richiesto la purificazione del frammento d'interesse da gel di agarosio all'1,8%. Pure la coppia PRBR4F-PRBR6R, comprendente il primo esone e parte del primo introne (fig. 3), ha prodotto come risultato il clone atteso di 706 bp in tutti e tre i soggetti. Nel caso dei soggetti 2 e 3 (fig. 10, lane 1D e 2D) i gel elettroforetici presentano bande aspecifiche (fatta eccezione per quelle relative ai primers visibili al di sotto delle 100 bp), mentre nel caso del soggetto 1 (fig. 10, lane 1C) sono presenti anche prodotti di dimensioni non attese.

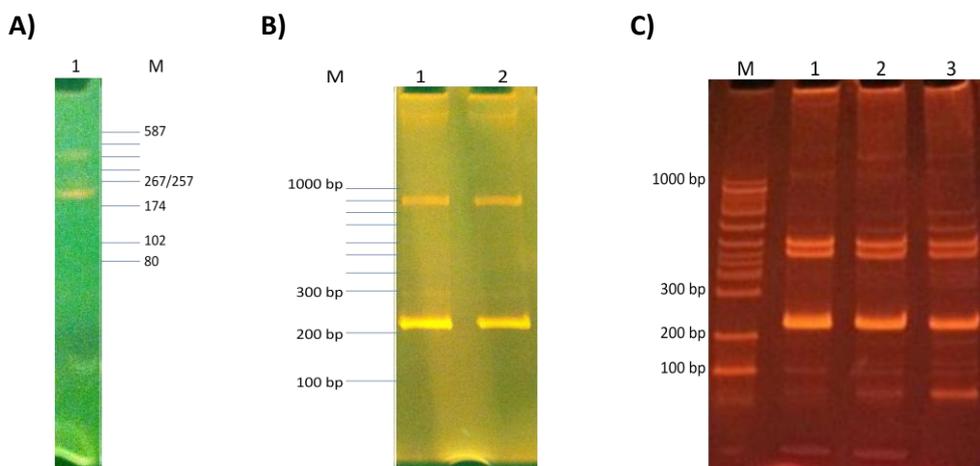


Fig. 9: PCRs PRB1, set di primers disegnati sulla sequenza NT_187658.1; elettroforesi su gels di poliacrilamide all'11%; **A)** PCR PRBR1F-PRBR2R per il soggetto 1, clone di dimensioni attese pari a 213 bp; il marcatore di pesi molecolari utilizzato (M) è il puc18 HaeIII digested. **B)** PCR PRBR1F-PRBR2R per i soggetti 2 (lane 1) e 3 (lane 2), clone di dimensioni attese pari a 213 bp; il marcatore di pesi molecolari utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer). **C)** PCR PRBR3F-PRBR4R, il clone atteso ha dimensioni pari a 227 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer).

Per quanto concerne il soggetto 1, il clone delle dimensioni di 706 bp è stato purificato da agarosio al 1,8%, mentre per i soggetti 2 e 3 è stato possibile eseguire la purificazione mediante Exo-BAP Mix. Anche in questo caso, la presenza di prodotti aspecifici potrebbe essere dovuta sia all'elevata omologia intergenica dei membri della classe PRB che ad ibridazioni totalmente impreviste in altre regioni del DNA. Le PCR condotte in presenza dei primers PRBR6F-PRBR7R (fig. 11 A e B) e PRB 2F-PRBR8R (fig. 11C) hanno permesso l'amplificazione dei cloni attesi (rispettivamente, 418 e 203 bp) per tutti i soggetti. La coppia PRBR6F-PRBR7R delimita una parte del primo introne (fig. 3), fin subito a monte del secondo esone, mentre la PRB 2F-PRBR8R comprende parte del secondo introne (fig. 2 e 3). La presenza di prodotti aspecifici nella miscela di reazione dei tre soggetti ha richiesto la purificazione di entrambi i cloni specifici da gel di agarosio al 2%.

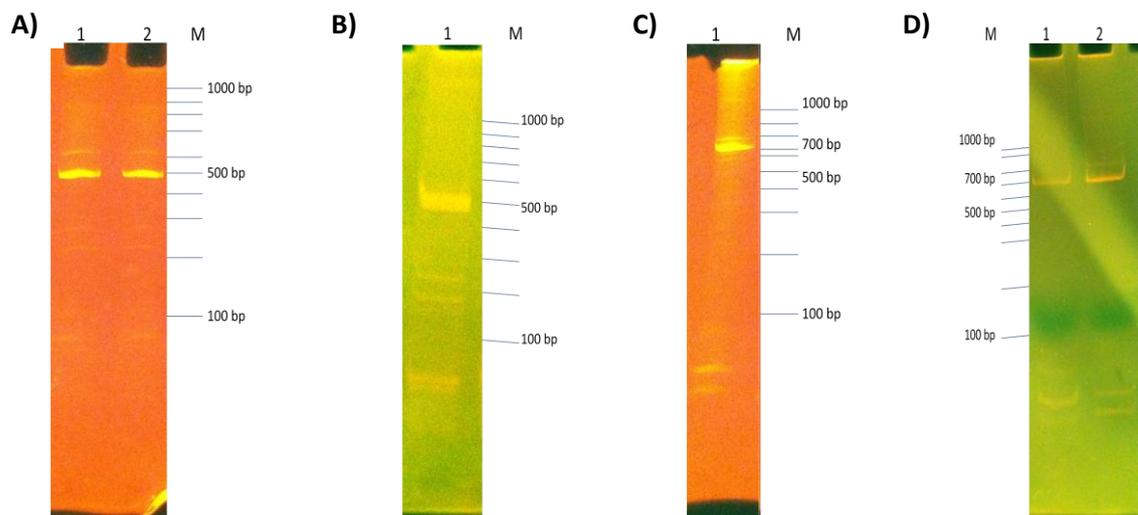


Fig. 10: PCRs PRB1, set di primers disegnati sulla sequenza NT_187658.1 elettroforesi su gels di poliacrilamide all'11%; **A)** PCR PRBR2F-PRBR4R per i soggetti 1 (lane 1) e 3 (lane 2), clone di dimensioni attese pari a 499 bp; il marcatore di pesi molecolari utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer). **B)** PCR PRBR2F-PRBR4R per il soggetto 2 (lane 1), clone di dimensioni attese è pari a 499 bp; il marcatore di pesi molecolari utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer). **C)** PCR PRBR4F-PRBR6R per il soggetto 1 (lane 1), clone di dimensioni attese è pari a 706 bp; il marcatore di pesi molecolari utilizzato (M) è il *marker* 1kp plus (ThermoFischer); **D)** PCR PRBR4F-PRBR6R per i soggetti 2 (lane 1) e 3 (lane 2), clone di dimensioni attese è pari a 706 bp; il marcatore di pesi molecolari utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer).

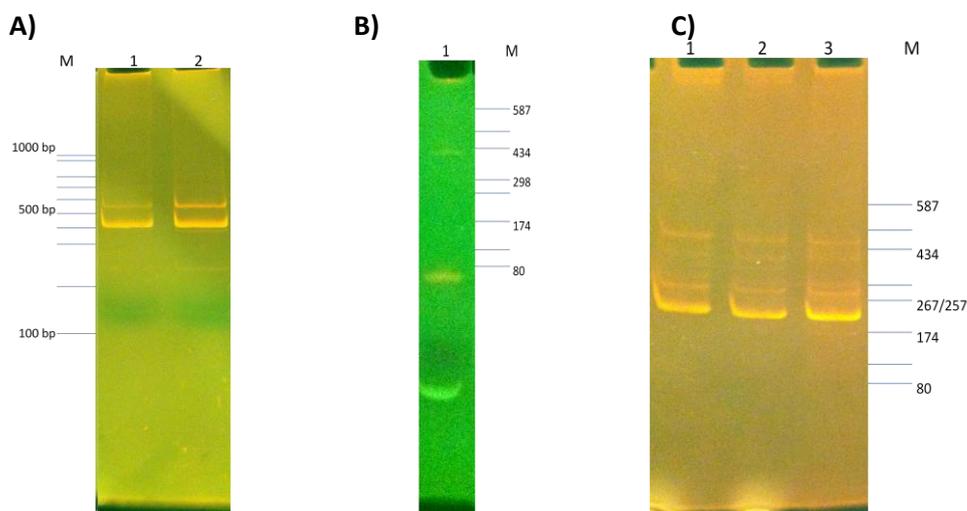


Fig. 11: PCRs PRB1, set di primers disegnati sulla sequenza NT_187658.1; elettroforesi su gels di poliacrilamide all'11%; **A)** PCR PRBR6F-PRBR7R per i soggetti 2 (lane 1) e 3 (lane 2), clone di taglia attesa 418 bp; il marcatore di pesi molecolari utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer); **B)** PCR PRBR2F-PRBR4R per il soggetto 1 (lane 1), clone di dimensioni attese pari a 418 bp; il marcatore di pesi molecolari utilizzato (M) è il puc18 *HaeIII* digested. **C)** PCR PRB 2F-PRBR8R, il clone atteso ha dimensioni pari a 203 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il puc18 *HaeIII* digested

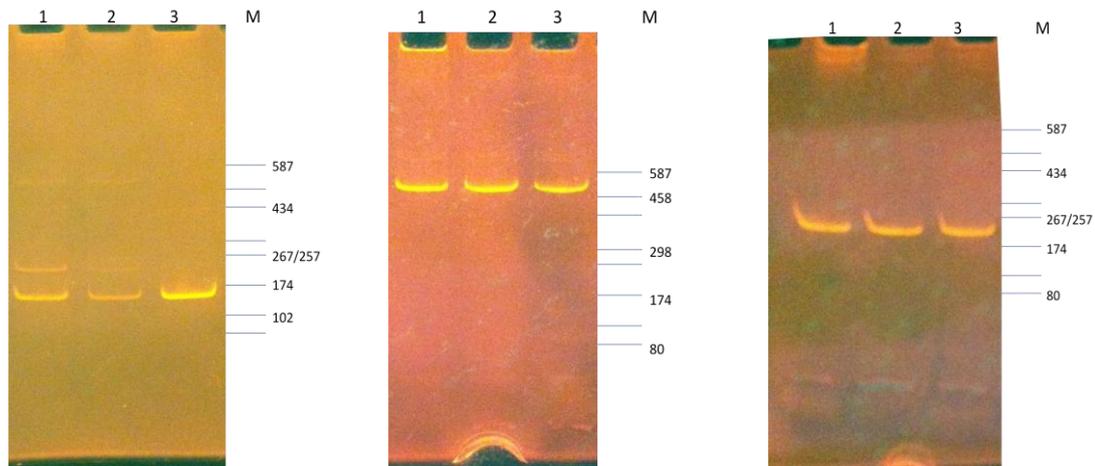


Fig 12: PCRs PRB1, set di primers disegnati sulla sequenza NT_187658.1; elettroforesi su gels di poliacrilamide all'11%; **A)** PCR PRBR8F-PRBR9R, il clone atteso ha dimensioni pari a 161 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il puc18 *HaeIII* digested. **B)** PCR PRBR10F-PRBR11R, il clone atteso ha dimensioni pari a 480 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il puc18 *HaeIII* digested. **C)** PCR PRBR11F-PRBR12R, il clone atteso ha dimensioni pari a 233 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il puc18 *HaeIII* digested.

La figura 12 mostra i profili elettroforetici relativi alle reazioni di PCR condotte con i primers PRBR8F-PRBR9R (fig. 12A), PRBR10F-PRBR11R (fig. 12B) e PRBR11F-PRBR12R (fig. 12C). La prima coppia delimita una regione da 161 bp a partire dal 3' del secondo introne fino al 5' del terzo esone; la seconda coppia comprende una porzione di 480 bp che si estende dal 3' del terzo esone e termina all'interno del terzo introne; l'ultima delle tre coppie copre una porzione da 233 bp all'interno del terzo introne (fig. 3). Tutte e tre le coppie di primers hanno permesso l'amplificazione dei cloni di dimensioni attese per i tre soggetti analizzati. Tuttavia, nel caso dei soggetti 1 e 2 (fig. 12A, lane 1 e 2), le reazioni di PCR PRBR8F-PRBR9R hanno prodotto come risultato anche l'amplificazione di cloni di dimensioni superiori a quelle previste, rendendo necessaria la purificazione da gel di agarosio al 2%. Vista l'assenza di aspecifici nella miscela di reazione del soggetto 3, è stato possibile procedere mediante purificazione con *Exo-BAP Mix*. Le reazioni di PCR condotte rispettivamente con le coppie di primers PRBR10F-PRBR11R (fig. 12B) e PRBR11F-PRBR12R (fig. 12C), hanno prodotto come risultato l'amplificazione esclusiva dei cloni di dimensioni attese che, analogamente a quanto descritto in precedenza, sono stati purificati facendo ricorso al reattivo *Exo-BAP Mix*.

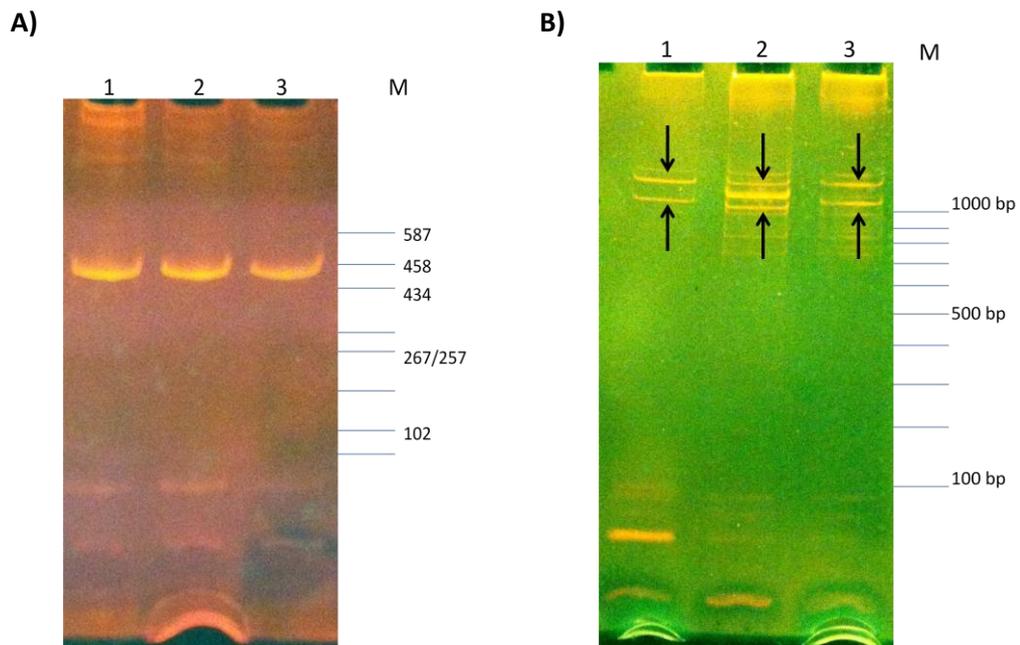


Fig 13: PCRs PRB1, set di primers disegnati sulla sequenza NT_187658.1; elettroforesi su gels di poliacrilamide all'11% **A)** PCR PRBR12F-PRBR13R, il clone atteso ha dimensioni pari a 454 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il puc18 *HaeIII* digested. **B)** PCR F2-PRBR10R, il clone atteso ha dimensioni pari a 1028/1211/1394 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3; le frecce rivolte in alto indicano i cloni di dimensioni pari a circa 1028 bp; le frecce rivolte in basso indicano i cloni di dimensioni pari a circa 1211 bp. Il marcatore utilizzato (M) è il *Gene Ruler* 100bp (ThermoFischer).

In figura 13 sono illustrati i risultati delle reazioni di PCR eseguite con i primers PRBR12F-PRBR13R e PRB 2F-PRBR10R. La prima coppia delimita una regione da 454 bp del terzo introne (fig. 3), mentre la seconda copre tutto il terzo esone, a partire dal secondo introne fino al 3' dell'esone stesso. Le reazioni di PCR condotte con la coppia PRBR12F-PRBR13R ha permesso di amplificare i cloni di dimensioni attese per i tre soggetti (fig. 13A). La presenza di prodotti aspecifici nella miscela di reazione, ha richiesto la purificazione da gel di agarosio all'1,5%. In realtà, sebbene non sia mostrato in figura 13, si è scelto di associare al primer reverse PRBR10R due diversi primers forward: PRBR9F, complementare al 5' dell'esone (fig. 3), e F2, appena descritto, che intercetta il 2° introne (fig. 2, tab. 3). Sulla base dei dati riportati in letteratura per le varianti di lunghezza del gene PRB1 (Lyons et al., 1988a,b), abbiamo stimato che le dimensioni attese dei prodotti di PCR per la coppia PRBR9F- PRBR10R dovessero essere 710 bp, 893 bp o 1076 bp, qualora il clone risultante fosse riconducibile rispettivamente alla variante allelica *small*, *medium* o *large*, e 1031 bp, 1214 bp o 1397 bp per la coppia PRB 2F-PRBR10R (tab. 3). Ciò comporta che molti dei cloni attesi per le suddette due coppie di primers, abbiano dimensioni ai limiti, o al di sopra dei limiti,

affinché la lettura dei cromatogrammi risultanti dalle reazioni di sequenziamento, possa essere ritenuta affidabile. Tuttavia, si è preferito comunque procedere all'amplificazione utilizzando queste coppie di primers, al fine di ovviare alle problematiche derivanti dalla presenza di ripetizioni in tandem all'interno di questa regione del gene PRB1 (Lyons et al., 1988a,b.). Le due coppie di primers sono state scelte in virtù del fatto che permettono l'amplificazione del terzo esone nella sua interezza, piuttosto che suddividerlo in porzioni più piccole come nella strategia sperimentale adottata nel corso del primo sequenziamento (fig.2), strategia che, sebbene più efficace ai fini dell'analisi nucleotidica, non ci ha consentito di ottenere tutti i cloni desiderati nella prima serie di amplificazioni (fig. 5, tab.2). La reazione di PCR condotta con i primers PRBR9F-PRBR10R non ci ha permesso di amplificare nessuno dei cloni desiderati (dato non mostrato), mentre l'utilizzo della coppia F2-PRBR10R ha prodotto come risultato una serie di cloni di difficile interpretazione. Sorprendentemente infatti, per tutti e tre i soggetti è possibile osservare una banda su gel elettroforetico di dimensione apparente pari a quella prevista per la variante *small* (fig. 13B, frecce rivolte in alto), sebbene noi ipotizzassimo di osservarla solo nel caso del soggetto 3 (tab. 1). Altresì, sembra essere del tutto assente il clone corrispondente alla variante *large*, che ci saremmo aspettati di osservare nel caso del soggetto 1 (fig. 13B, lane 1). Non sorprende affatto, invece, la presenza del clone corrispondente alla variante *medium* in tutti e tre i soggetti (fig. 13B, frecce rivolte in basso), dal momento che il soggetto 1 e 3 presentano i peptidi riconducibili ad eterozigosi per tale variante del gene PRB1, mentre il soggetto 2, sempre in base al profilo peptidico, ci aspettavamo fosse omozigote per la variante *medium* (tab. 1). Quest'ultimo soggetto presenta inoltre una banda intermedia tra le due appena descritte (fig. 13B, lane 2) non attesa. Per fornire una spiegazione ai dati osservati, si è provveduto a purificare da gel d'agarosio all'1,5% le bande riconducibili alla dimensione attesa per la variante sia *small* che *medium* in tutti e tre i soggetti. Inoltre, per il soggetto 2, è stata purificata al contempo anche la banda di dimensione intermedia tra le due (fig. 13B, lane 2). Nella prima serie di reazioni di PCR, come visto in precedenza, l'amplificazione dell'ultimo esone risultava essere incompleta, ragion per cui è stata individuata la coppia di primers PRBR13F-PRBR14R, teoricamente in grado di ibridarsi a monte e a valle del quarto esone (fig. 3). Tuttavia, la PCR condotta con la suddetta coppia di primers non ci ha permesso di amplificare il clone di taglia attesa.

Sequenziamento dei cloni amplificati col secondo set di primers

Le sequenze relative ai tre soggetti sono state ricostruite mediante l'assemblaggio dei singoli cloni di PCR ottenuti dalle amplificazioni con i primers riportati in tabella 3, confrontando i cromatogrammi risultanti da questo sequenziamento con quelli relativi alle medesime regioni già analizzate e sequenziate in precedenza (fig. 6, 7, 8). Le sequenze risultano essere, anche in questo caso parziali, in quanto le coppie di primers PCR PRBR9F-PRBR10R e PRBR13F-PRBR14R non hanno permesso l'amplificazione dei cloni attesi (vedere sezione precedente per i dettagli). Inoltre, l'analisi di alcuni dei cromatogrammi non è stata possibile, per via della presenza di eccessivo rumore di fondo o per l'assenza di picchi di fluorescenza leggibili in modo chiaro e non equivoco.

In figura 14 è riportata la ricostruzione della sequenza di PRB1 relativa al soggetto 1, comprendente i primi due esoni al completo e mancante di una porzione importante (675 nucleotidi) del terzo esone, oltre a una breve porzione (20 bp) del quarto esone (a causa della mancata amplificazione del frammento di DNA delimitato dai primers PRBR13F-PRBR14R). Pertanto, il sequenziamento dei due cloni di dimensioni differenti rispetto a quelle attese con l'impiego dei primers PRB 2F-PRBR10R (fig 13B, lane 1) non ha permesso di ricostruire in modo completo il terzo esone. Questo perché, nel caso della banda di 1028 bp (fig 13B, lane 1, freccia rivolta in alto) i cromatogrammi relativi alle reazioni di sequenziamento sono risultati del tutto illeggibili. Per quanto concerne invece la banda di dimensioni superiori (fig 13B, lane 1, freccia rivolta in basso), il cromatogramma in forward risulta essere illeggibile, mentre il cromatogramma in reverse ha consentito la ricostruzione solo di una breve porzione di sequenza a monte del primer PRBR10R (fig 3), perfettamente sovrapponibile alla sezione del terzo esone ricostruita nel precedente sequenziamento, a partire dal clone F5-R5 (fig. 6).

PRB1 soggetto 1 (ref. NT_187658.1)

551282

tttctagaaatttcatggtggtttctgtgtctactttgcA/Gttaataaccctaaaaatactttagaatacaaaatt
551206

ttagacttctgtgtgttacaggaccaagccttgccctcagctgagtgagatccacagaaacttgaacaatagccctat
551128

cctgagcaggctgcggggcagcctttctgtcatctatgtgtgcacagaagtgtgctgtcctcaca (nnnnnnnnnnnn
551050

n) cagaaacctctctatcacatcaaaaattgagctaaatatgaaaatatctgtA/Gcacacacatcacaagagatgg
550974

aaataqggcataaaaaG/TatC/GcagagatattgaccatT/CttcactgctggT/Cttagcatcgctgctcacaca
550904

C/GacataaacagaaataaatagcacagtggaaagtgatG/Aatagtcctacctaagattgaaactctaatagata
550834

tctaaagttcagtgatgacacagtgctgatggccttgacacagttccttcaaacattgttcaagtcataagttgtac
550752

cttttqaaataagacactaacaacatctctatccaactcatctctctgggctagagtctcaaagaaaaataaggga
550674

tacacctgacctgagtgagcaaacgagaagccagtcctctgaggtggtgaggccaccagtgaggagctcaaag (nn
550596

nnnnnnnnnnnn) tctctttataaaggagttgccacgttctccca^{gcacagagttgggagtgactccagagcct}
550518

^{ccT/A}gcaagatgctggtgattctgctgctcagtgccT/CtctgctggcctgagctcagctcagaG/Acttaaatgaag
550446

qtaaacacagaagaggaaaagatggtgactctgcttggggcttaggaggtgataatggtaattacggggaggagag
550368

gagaatgaaaacacagatggggctgcagagtttcatgcctagatcaggagacC/Gtggtgtaccctcattccacac
550292

taaqqqtcttaatttatttaagtacaatgaaatccaataaagaattgttccaggggaatgagaaggaagatttg
550214

catttatagatagaactgtgctgtgaaggatgcagtagagaatgcaaggcagattcatggaagtccagctgtgaagat
550136

cctaaactgatctcagtaagtacacagggA/Ctgatggtggccttctgtgtacagtgcatcagcatcgatgagcgt
550060

aaacacacacagatcagagatactgcagagacagagaattggatggaacacttgtctctgtttaactagagatacag
549982

aaatatcagagccaatcattgtcatttttctctcccttacacgcagatattcaatgtgctgggagtggtatgggtaag
549904

attqtattgaagtgattacttctggttaccccaattgagaaagcatgtgtacataagcaatgtatttataggaA/Gat
549828

gagggcataagaacacccaaaA/GtatcacattgaagtacctggcatG/Ttgtgaactaaattagcattaagtcttga
549750

gggatgctagggagggaaaaaagggtccttctatggtgagttcatggctggtgctctgtcataacaaccctgT/Cct
549674

cccttacaccttctcccttccagcagcttccagatggtggctgacT/Gagttaacttaggggatgcatgggggtg

549598

tcatgagaagA/TccC/GttttccctgtagaacacttgtgagcttgaaggttcaagatgtaactG/Tttcccatca

549526

tctgtgcttctctcttag^{atgtcagccaggaagaatctccctccctaatagcag}gtaaagcctgattcG/Attctca

549450

atctgttT/Ctgactgtctttttctgcttatgagtgatcagttctccagtgcttcttatcaaaacttctctttcag

549374

gagttgattaatgtagtgcccctaacaatataggcactcttcG/Atgcaaccttgattctgggcatcatgagcaggc

549298

caccaaattgaacggcagagatgcttggccttagatgacaacaggagtggttA/GcctcatccccC/TggccaggaA

549225

/GtgctgctgggagatgacagaaatgggcagcatcctcattctgtctcctctttataA/Ctgagaggcctcaactg

549150

ctttgttcttccccagctttccactccagagttctatgtcttctactgaaaatgcaaaGaaattaatgtctttgttcca

549072

ttttgtgtatttcccactcagcttgttaccacagctgataaaaatttactgcaactattcagtgatcttgtgtgg

548994

gcttttactctgtctttctctctctttgtcctccag^{gaaatccacaaggaccatccccacaaggaggcaac}

548256
gtcccaagtgccgatctcctccaggaaagccacaaggaccacccaacaagaaggcaacaatcctcaagggtcccc
 548178
acctccagcaggaggcaatccccagcagcctcaggcaacctctgtgtggacagccccaggaccaccacgcctctca
 548100
agggggcagaccttccagacctccccagtgcagcctccccagtcatttaggattcaatgacaggtatgattccagt
 548022
 tattcttcaccaagtgtcttaattgtctacagctctccagctttattgtgccaatgaatcagctaaaagccattggca
 547944
 ttgtatagtcccagatcccatttctaaagatttgtattgacatattctggaaatgggtaacaagatcctatatttga
 547866
 acaaactctttaaggaattctgatgttgagaaacaaaattccaaataatctgtcttaagttgtgttggaacaaggaa
 547788
 gtagtaccatgttctctctggcgtctgttttctgtgcacaaactgagagacctccatttaaagtttccacctgagc
 547710
 actgtttgtcagtcctgcctcacaccagcctcttgagtcagtcatttctgccaagtggtcctgaactttcagcagc
 547632
 taaatgggtgtctcatttttcaaattcttactgttcaataagtacatgattaagctaacaaaaaatatctaagcaatg
 547554
 gaaaaatatgaatctaaatttaaaggcatgactcatcctacctgcctccccctcttcagaaaaactgccactgttaact
 547476
 ttatggcatcttctgtttgaaatatttatgtgtacatagacAactagaatatttttccccagaactaataccataat
 547398
 ttatattcaggtacatattgttagtcatttaaaaaatacatttctttgaaaatttccacatacgtctatgaagctaagt
 547320
 agatctcttcagtggttatctgtttgtttttaccattttataactactccattatgtggctgcaccgtgatttctttaa
 547242
 ccaatctgtgtcactggacactgaggggtggtttcagcttctcactgttataaaatattgttccagttcccatctgtgta
 547164
 aatatactgtgaacaaattcagcagcaagtaataaC/Taagctaagaatgatcttctgtcttcatcacgtaaggaa
 547088
 aatttgagcacattttgtgcaagggcatccaaagagtgaacacacaaaaaattaggaggaaacacaggaggtagaa
 547010
 gggatgggggagagaggatgggctctcatgtactgtagtgcagtaagaccagtgaggaattcgacatttctgccaatg
 546932
 tcaagtctggtctatgaacttcttctttgtttcaggaagtgaataagaagatgagagtgattcaaatgattca
 546854
aattccatgacattggaaaaaggtcatcatagctctaacttcaatatac

Fig. 14: sequenza del gene PRB1 relativa al soggetto 1, riferimento: NT_187658.1; le porzioni esoniche sono riportate in colore rosso sottolineate; tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/).

All'inizio delle reazioni di sequenziamento, i filamenti di DNA neosintetizzati dalla DNA polimerasi terminano con un nucleotide fluorescente, come previsto. Tali filamenti però sono molto corti: questo favorisce una loro concentrazione eccessiva e un effetto “sovraccarico” del segnale. In altri casi, può verificarsi che il segnale sia comunque confuso, a causa della presenza di piccole molecole contaminanti. Nonostante la relativamente bassa qualità della sequenza all'estremità in cui è presente il primer reverse, questa è stata l'unica porzione del clone PRB 2F-PRBR10R “decodificabile”, mentre il resto del tracciato cromatografico è risultato del tutto illeggibile. Ciò risulta essere in accordo con la presenza, in questa regione, di numerose

ripetizioni in tandem, di tratti omopolimerici, e tratti ad elevata omologia intergenica della classe PRB, tutti fattori che possono determinare slittamenti nell'appaiamento della DNA polimerasi. A monte del primo esone, a partire dalla posizione 550749, è presente un altro breve tratto di sequenza, che non è stato possibile identificare, mediante il sequenziamento del clone PRBR4F-PRBR6R (fig. 14 e 3). Questo tratto si trova in corrispondenza del primer forward e di conseguenza difficilmente analizzabile, in quanto le estremità 5' e 3' dei cromatogrammi sono spesso caratterizzate dalla presenza di picchi di fluorescenza di bassa intensità, difficilmente leggibili. Per le stesse motivazioni, il sequenziamento del frammento con il primer *reverse* non ha consentito di ricostruire il tratto di sequenza in 5'. Un ulteriore breve tratto di sequenza caratterizzato da picchi non leggibili, si trova in corrispondenza dell'estremità 5' dei tracciati cromatografici relativi al clone PRBR2F-PRBR4R (a partire dal nucleotide 551192) e, nemmeno in questo caso, l'analisi dei cromatogrammi in *forward* e *reverse* ha permesso l'identificazione dei rispettivi nucleotidi. Quest'ultimo clone e quello corrispondente alle coppie di primers PRBR4F-PRBR6R presentano, sebbene in misura inferiore, le stesse problematiche descritte per il clone PRB 2F-PRBR10R, ovvero omologia intergenica nei confronti dei membri della classe PRB, presenza di piccole ripetizioni in tandem oltre a brevi tratti omopolimerici, caratteristiche molecolari che nell'insieme possono aver contribuito alla mancata ricostruzione dei due brevi tratti di sequenza appena discussi. Il confronto tra i cromatogrammi relativi ai cloni PRBR8F-PRBR9R e PRBR10F-PRBR11R (fig. 3), rispettivamente coi cromatogrammi relativi ai cloni F2-R2 e F5-R5 (fig. 2), ci ha permesso di identificare i due nucleotidi non rilevati nel corso del precedente sequenziamento (fig. 6, rispettive posizioni: 2275 e 3431). Al fine di confermare gli SNPs individuati, sono stati confrontati i cromatogrammi derivanti dal primo sequenziamento (fig. 6) e da quest'ultimo (fig.14), per quei cloni che risultassero almeno parzialmente sovrapponibili, ovvero: F1-R1 e F2-R2 con F2-PRBR8R e PRBR8F-PRBR9R; F5-R5 con F2- PRBR10R; F6-R6 con PRBR10F-PRBR11R e PRBR11F-PRBR12R (fig. 2 e 3). Complessivamente, nella sequenza di PRB1, parzialmente ricostruita, relativa al soggetto 1 sono stati individuati 31 SNPs, la maggior parte dei quali in eterozigosi, in particolare tre presenti all'interno del primo esone (fig. 14).

La ricostruzione parziale della sequenza di PRB1 relativa al soggetto 2 è riportata in figura 15. Come osservato per il soggetto 1, anche in questo caso manca una porzione da 675 bp del terzo esone, oltre agli ultimi 20 nucleotidi del quarto esone. In questo

caso, relativamente alla PCR condotta in presenza dei primers PRB 2F-PRBR10R (fig. 3) sono stati sequenziati tre diversi cloni di dimensioni leggermente differenti (fig 13B, lane 2), per nessuno dei quali è stato possibile ottenere dei tracciati cromatografici leggibili, pertanto le porzioni del terzo esone riportate (fig. 15) derivano integralmente dal sequenziamento precedente (fig. 7). Relativamente ai cloni derivanti dalla PCR con primers PRB 2F-PRBR10R, valgono le medesime considerazioni fatte precedentemente per il soggetto uno. In corrispondenza della regione situata trentasette nucleotidi a monte del primo esone (fig 15), sono stati riscontrati all'estremità 5' dei cromatogrammi relativi al clone PRBR4F-PRBR6R, due picchi di fluorescenza consecutivi ai quali non è stato possibile assegnare nessun nucleotide. Si tratta di una regione localizzata subito a valle rispetto alla zona di ibridazione del primer forward, e quindi nella regione dei cromatogrammi risultata essere maggiormente problematica, per le motivazioni molecolari discusse precedentemente esaminando i risultati sperimentali riportati in figura 14. In assenza del clone PRBR13F-PRBR14R, non è stato possibile confrontare i cromatogrammi di quest'ultimo con quelli del F7-R7. Per questa ragione, i nucleotidi non rilevati nel corso del precedente sequenziamento, in corrispondenza della porzione decodificata del quarto esone e 80 bp a monte di quest'ultimo, non sono stati identificati. La validazione degli SNPs individuati è stata effettuata mediante confronto dei cromatogrammi derivanti dal primo sequenziamento (fig. 7) e dal secondo (fig.15), prendendo in esame i cloni che fossero almeno parzialmente sovrapponibili, ovvero: F1-R1 e F2-R2 con F2- PRBR8R e PRBR8F-PRBR9R; F5-R5 con F2- PRBR10R; F6-R6 con PRBR10F-PRBR11R e PRBR11F-PRBR12R (fig. 2 e 3). Nella porzione di sequenza di PRB1 ricostruita per il soggetto 2 è stato possibile individuare, nel complesso, 29 SNPs, per la maggior parte in eterozigosi, uno dei quali risulta essere presente all'interno del primo esone.

PRB1 soggetto 2 (ref. NT_187658.1)

551282
tctagaaatttcctggtggtttctgtgtctactttgcA/Gttaataaccctaaaaatactttagaatacaaaat
551206
agacttctgtgtgttacaggaccaagccttgccctcagctgagtgagatccacagaaacttgaacaatagccctatcc
551128
tgagcaggctgcggggcagctttctgtcatctatgtgtgcacagaagtgtgctgtcctcacacatctcatagatagca
551050
gaaaccatctctatcacatcaaaaattgagctaaatatgaaaatatctgtacacacacatcacaagagatggaaatag
550972
ggcataataaT/GatC/Gcagagatattgaccattttcactgctggctgtagcatcgctgtcacacacataaaca
550898
gaaataaatagcacagtggaagtggatgatagctttacctaataagattgaaactctaatagatctaaagttcagtg
550820
atgacacagtgctgatggcttgacacagttccttcaaaaccattgttcaagtcatagtttgtaccttttgaaataag
550742
acactaacaacatatacctatccaactcatctctctggttagagctctcaaagaaaaataagggatacacctgacctgc
550664
agtgagcaaagcagaagccagctctctgaggtggtgagggccaccagtggtgagctcaaaggtgccattggt (nn) tc
550586
tcctctttataaaggagttgccacgttctctccagcacagagttgggagtgactccagagcctcctgcaagatgctg
550508
ttgattctgctgtcagtgcccT/Ctgctggccctgagctcagctcagaacttaaatgaagggtaaaacagaaggggaa
550432
aagatgtggtgactctgcttggggcttaggagtgataatggtaattacggggaggagaggagaatgaaaacacagat
550354
ggggtgacagagtttcatgcctaggatcaggagacctgttgtA/Gccctcattccacactaagggtttctaatttat
550278
ttaatgtacaatgaaatccaataaagaatttgttccaggggaatgagaaggaatgatttgcatttatagatagaactg
550200
tgctgtgaaggatgcagtagagaatgcaaggcagatcctatggaagtccagctgtgaagatcctaaactgatctcagta
550122
agtacacagggatgatggtggccttgctgtacagtgcatcagcatcgatgaT/Cggcgataaacacacacagtatcag
550046
agatactgcagagacagagaattggatggaacactgtctctgtttaactagagatacagaaatatcagagccaatca
549968
ttgtcatttttctctcccttacacgcagtatctcaatgtgctgggagtggtatgggtaagattgtattgaagtgatta
549890
cttctggttacccaattgagaaagcatgtgtacataagcaatgtatttataggaA/Gatggagggcataagaacacc
549814
aaaG/AtatcacattgaagtacctggcatG/Ttgtgaactaaattagcattaagtcttgagggatgctagggagggaa
549740
aaaaggggctctctatgttgagttcatggctgtgtctgtcataaacaacctgT/Cctccccttacaccttctcc
549664
ccttccagcagcttcacagatggtggctgaCT/GagttaacttaggggatgcatggggtgtG/CatgagaagA/Tccc
549592
ttttccctgtagaacacttgtgagcttgaaggttcaagatgtaacttttccatcatcctgtgcttctcttctaga
549514
tgtcagccaggaagaatctccctccctaatagcaggtaaaG/Tcctgattcgttctcaatctgttttgactgtctttt
549438
tctqcttatgagtggtatcagttctccagtgcttcttatacaaaactttcctttcaggagttgattaatggttagtgcC
549361
/TcctaacaatataggcactcttcG/AtgcaaccttgattctgggcatcatgagcaggccaccaaattgaaC/Tggca
549288
gagatgcttggcttagatgacaacaggagtggtgtgcctcatccccC/Tggccaggaatgcctgctgggagatgaca
549212
gaaatgggcagcatcctcattctgtctcctctttataA/Ctgagagccctcaactgctttgttcttcccagctttc
549136
cactccagagttctatgtcttactgaaaatgcaaaG/Taaattaatgtctttgttccattttgtgtatttccccac
549060
tcagcttgttaccacagctgataaaaatttaG/Ctgcaactattcagtgaaatctgtgtgggcttttactctgtcttt
548984
ctcttctctttgtcctccaggaaatccacaaggaccatccccacaaggaggcaac

548256
gtcccaaagtgcccgatctcctccaggaaagccacaaggaccaccccaacaagaaggcaacaatcctcaaggtcccc
 548178
acctccagcaggaggcaatccccagcagcctcaggcacctcctgctggacagccccagggaccaccacgcctcctca
 548100
agggggcagaccttccagacctccccagtgacagcctccccagtcattctaggattcaatgacaggtatgattccagtt
 548022
 tattcttcaccaagtgtcttaattgtctacagctctccagctttattgtgccaatgaatcagctaaaagccattggca
 547944
 ttgtatagtccagatcccatttctaagatttgtattgacatatctctggaaatgggtaacaagatcctatatatttga
 547866
 acaaactctttaaggaattctgatgttgagaaacaaaattccaaataatctgtcttaagttgtgttggaacaaggaa
 547788
 gtagttaccatgttctctctggcgtctgttttctgtgcacaaactgagagacctccatttaaagttttcacctgagc
 547710
 actgtttgctcagtcctgcctcacaccagcctcttgagtccagttctctgccaagtggtcctgaactttcagcagc
 547632
 taaatgggtgtctcatttttcaaattcttactgttcaataagtacatgattaagctaacaaaaaatatctaagcaatg
 547554
 gaaaaaatgaatctaaatttaaaggcatgactcatcctacctgcctcccctccttcagaaaactgccactgttaact
 547476
 ttatggcatcttctgtttgaaatatttatgtgtacatagactactagaatatttttccccagaactaatacT/Cata
 547400
 atttatattcaggtacatatgttagtcatttataaaaaatacatttctttgaaaatttccacatacgtctatgaagctaa
 547322
 gtagatctcttcagtggttatctgtttgtttttaccattttataactactccattatgtggctgcaccgtgatttcttt
 547244
 acccaatctgtgtcactggacactgaggggtggtttcagcttctcactgtttataaaaatattgttccagttcccatctgtg
 547166
 taaatatactgtgaacaaattcagcagcaagtaataaT/Caagctaagaatgatcttctgtcttcatcacgtaagga
 547090
 acaatttgggagcacattttgtgcaagggcatccaaagagtgaacacacaaaaaattagggaggaaacacaggaggtag
 547012
 aagggatAagggggagagaggatgggctctcatgtactgta (g) tgcagtA/Tagaccagtgaggaattcgacatttcc
 546936
 T/GgccatgtcaagtctggctcatgaacttcttG/Ctttgtttgtttcaggaagtgaataa (n) aagatgagagtgta
 546862
ttc (n) aatgattcaaatG/Tccatgacattggaaaaaggtcatcatagctctaacttcaatatac

Fig. 15: sequenza del gene PRB1 relativa al soggetto 2, riferimento: NT_187658.1; le porzioni esoniche sono riportate in colore rosso sottolineate; racchiusi tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/).

La figura 16 mostra la sequenza di PRB1, parzialmente ricostruita, per il soggetto 3. Come negli altri soggetti, la ricostruzione è priva di una porzione da 20 nucleotidi del quarto esone e di una porzione del terzo esone, risultante tuttavia più estesa in questo caso rispetto a quanto osservato nei soggetti 1 e 2 (794 bp; fig. 14 e 15). Il sequenziamento dei due cloni di dimensioni differenti, inerenti la PCR condotta con primers PRB 2F-PRBR10R (fig 13B, lane 3), non ha permesso di ricostruire in modo completo la sequenza nucleotidica del terzo esone neanche per questo soggetto.

PRB1 soggetto 3 (ref. NT_187658.1)

551048
aaccatctctatcacatcaaaaattgagctaaatatgaaaatatctgtG/Acacacacatcacaagagatggaaatag
550970
ggcataaA/TaaG/TatG/CcagagatattgaccattttcactgctggC/Gtgtagcatcgctgctcacacacacata
550900
aacagaaataaatagcacagtggaagtggatgatagtcttacctaaaagattgaaactctaataatgatataaagttc
550822
agtgatgacacagtgctgatggcttgacacagttccttcaaaaccattgttcaagtcatagttgtaccttttgaaa
550744
taagacactaacaacatatcctatccaactcatctctctgggctagagtctcaaagaaaaataagggatacacctgac
550666
ctgcagtgagcaaagcagaagccagtcctctgaggtggtagggcccaccagtggtggagctc (nnnnnnnn) attggt
550588
cttctcctctttataaaggagttgccacgttctcccagcacagagttgggagtgactccagagcctcctgcaagat
550510
gctgttgattctgctgtcagtgcccttgctggcctgagctcagctcagaacttaaatgaaggtaaaacagaagagg
550432
aaaagatgtggtgactctgcttggggcttaggagtgataatgtaattacggggaA/Ggagaggagaatgaaaacac
550356
agatggggctgcagagttttcatgcctaggatcaggagacC/GtgtgtG/ACcctcattccacactaagG/Agtttc
550284
taatttatttaagtacaatgaaatccaataaagaattgttccaggggaatgagaaggtaagatttgcatttataga
550206
tagaactgtgctgtgaaggatgcagtaC/GagaG/AtgcaaggcagattcatggaagtcagctgtgaagatcctaT/
550133
AactgatctcagtaagtacacaggatgatggtggccttgctgtacagtgcatcagcatcgatgaT/Cggcgataaac
550057
acacacagtatcagagatactgcagagacagagaattggatggaacacttgtctctgtttaactagagatacagaaat
549979
atcagagccaatcattgtcatttttctctcccttacacgcagatattcaatgtgctgggagtggtatgggtaagattg
549901
tattgaaA/GtgattacttctggttaccccaattgagaaagcatgtgtacataagcaatgtatttataggaA/Gatgg
549827
agggcataagaacacccaaaA/GtatcacattgaagtacctggcatT/Gtgtgaaactaaattagcattaagtcttgagg
549753
gatgctagggagggaaaaaagggtcttctatgttgagttcatggctgttgctctgtcataacaaccctgc/tctcc
549675
ccttacaccttctccccttcagcagcttcacagatggtggctgaC/TT/Gagttaacttaggggatgcatgggggtg
549601
tG/CG/AtgagaagA/TcTC/CGttttccctgtagaacacttgtgagtcctgaaaggttcaagatgtaactGttccc
549533
atcatcctgtgcttctcttctagatgtcagccaggaagaatctccctccctaatagcaggtaaaT/GccC/TgattcA
549460
/GttctcaatctgttC/TtgactgtctttttctgcttatgaA/Gtggatcagttctccagtgcttcttatcaaaact
549387
ttcctttcG/AggaA/GttgattaatgtagtgcT/CcctaaC/TaatataggcactcttcA/Gtgcaaccttgattc
549319
tgggcatcatgagcaggccaccaaattgaaCggcagagatgcttggttagatgacaA/GcaggagtgtgttA/Gcct
549241
catccccT/CggccaggaG/AtgcctgctgggagatgacagaA/CAatgggcagcatcctcattctgtctcctctt
549169
atactgagaggcctcaactgctttgttcttcccagcT/AttccactccagagttctatgtcttactgaA/Gaatg
549095
caaaGaaattaatgtctttgttccatttttgtgtatttccccactcagcttgttacccaaA/GctgataaaaattC/T
549021
actgcaactattcagtgaaatctgtggtggcctttactctgtctttctcttctcttcttctcctccaggaaatccaca
548943
aggaccatccccacaaggaggcaac

548215
ctgctggacagccccagggaccaccacgcctcctcaagggggcagaccttccagacctccccagtgacagcctcccc
 548137
agtcatctaggattcaatgacaggtatgattccagtttattctccaccaagtgctctaattgctacagctctccagct
 548059
 ttattgtgccaatgaatcagctaaaagcccattggcattgtatagctccagatcccatttctaaagatttgtat **T/C**g
 547981
 acatattctggaaatgggtaacaagatcctatatttgaacaaactctttaaggaattctgatggtgagaaacaaaat
 547825
 tcc **A/T** **A/C**taatctgtcttaagttgtggttgcaacaaggaagtagtaccatggt **C/A**tctctggcgtctggtttc
 547753
 tgtgcacaaactgagagacctc **C/G**catttaaagttttcacctgagcactggttctcagtcctgcctcacaccagcc
 547677
 tcttgagtcagatattcctgcaa (gtggctccctgaactttca) gcagctaaatggtgtctcatttttcaaattctta
 547599
 ctggttcaataagtagcatgattaagctaacaaaaaatatctaataatgcaatggaaaaatgaatctaaatttaaggcat
 547521
 gactcatcctacctgcctccccctcctcagaaaactgccactgttaactttatggcatcttctggttgaaatatttat
 547443
 gtgtacatagac **A**actagaatatttttccccagaaactaatac **C/T**ataatttatattcaggtacatagtttagtcat
 547367
 ttaaaaaatacatttctttgaaaatttccacatacgtctatgaagctaagtagatctcttcagtggttatctggttgt
 547289
 ttttaccattttatactactccattatgtggctgcaccgtgatttctttaaccaatctgtgtcactggacactgaggg
 547211
 tggtttcagcttctcactggtataaaatagttccagttcccatctgtgtaaatatatactgtgaacaaattcagcagc
 547133
 aagtaataa **T/C**aagctaagaatgatcttctgtcttcatcacgtaaggaacaatttgagacacattttgtgcaagggc
 547057
 atccaaagagtgaacacacaaaaaattagggagggaacacaggaggtagaagggat **A/G**ggggagagaggatgggctc
 546981
 tcatgtactg **T/G**agtgacagtaagaccag (n) gaggaattc (n) acatttctctg **G/C**catgtcaagctggtctatga
 546907
acttcttctttgtttgttcaggaagtgaataagaagatgagagtgattcaaatgattcaaattccatgacattgga
 546829
aaaaggctcatcatagctctaacttcaatatac

Fig. 16: sequenza del gene PRB1 relativa al soggetto 3, riferimento: NT_187658.1; le porzioni esoniche sono riportate in colore rosso sottolineate; racchiusi tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/).

Infatti, relativamente al clone di dimensioni apparenti pari a circa 1028 bp (fig 13B, lane 3, freccia rivolta in alto) i cromatogrammi sono risultati del tutto illeggibili. Per quanto riguarda il clone di dimensioni apparenti superiori (fig 13B, lane 1, freccia rivolta in basso), i tracciati cromatografici hanno consentito la ricostruzione di una sequenza di soli 39 nucleotidi a monte del primer PRBR10R (fig 3).

Come osservato nel caso degli altri due soggetti analizzati, le cause plausibili dell'insuccesso nel corretto sequenziamento di questa regione del gene PRB1 sono molteplici, dovute principalmente alle caratteristiche intrinseche del gene stesso (omologia intergenica, sequenze ripetute in tandem, presenza di omopolimeri, sequenza ricca in GC) e ai limiti oggettivi della metodica di sequenziamento. A monte del primo

esone (45 bp, fig 16) nella estremità 5' dei cromatogrammi relativi al clone PRBR4F-PRBR6R, è presente una breve porzione di sequenza caratterizzata dalla presenza di picchi non leggibili. Si tratta di una porzione nucleotidica corrispondente alla zona di ibridazione del primer forward, e quindi nella regione dei cromatogrammi che ha generato maggiori problemi di interpretazione, come già osservato nel soggetto 2 (fig. 15). In assenza del clone PRBR13F-PRBR14R (fig. 3), anche per il soggetto 3 non è stato possibile confrontare i cromatogrammi di quest'ultimo con quelli del clone F7-R7 (fig. 2). Per questa ragione, i nucleotidi non rilevati nel corso del precedente sequenziamento in corrispondenza della porzione decodificata del quarto esone, e 80 bp a monte di quest'ultimo, non sono stati identificati (fig. 16). È stato eseguito il confronto tra i cromatogrammi derivanti dal primo sequenziamento (fig. 8) e da quest'ultimo (fig. 16), ovvero:

- F1-R1 e F2-R2 *vs* F2- PRBR8R e PRBR8F-PRBR9R;
- F6-R6 *vs* PRBR10F-PRBR11R e PRBR11F-PRBR12R (fig. 2 e 3).

Questo ci ha consentito di validare gli SNPs individuati, che nella porzione della sequenza di PRB1 ricostruita per il soggetto 3 sono nel complesso 55, per la maggior parte in eterozigosi, nessuno dei quali presente a livello esonico.

Amplificazione con il terzo set di primers e purificazione dei cloni PCR

Al fine di completare la sequenza del gene PRB1 per i tre soggetti analizzati, è stato progettato un nuovo set di primers, scegliendo come riferimento la sequenza dell'intero cromosoma 12, NC_000012.12 (Sherer et al., 2006), assolutamente identica alla NT_187658.1 ma costituente l'assembly più aggiornato depositato su NCBI nella fase *in itinere* del nostro lavoro. Per le finalità della nostra ricerca, abbiamo isolato solo la porzione comprendente l'ORF (Open Reading Frame) del gene, che si estende dal nucleotide 11351823 al 11355590, del filamento antisenso. Dal momento che le porzioni nucleotidiche mancanti del gene in esame riguardano il terzo ed il quarto esone per tutti i soggetti analizzati, si è provveduto a disegnare due coppie di primers: la prima che ibridasse alle estremità del terzo esone (HOPE 1F-HOPE 2R); la seconda a monte e a valle del quarto esone (HOPE 4F-HOPE 4R; fig. 4, tab. 4). Inoltre, al fine di minimizzare l'effetto di ibridazione intergenica, quantomeno in corrispondenza del

terzo esone, abbiamo deciso di proseguire il lavoro sperimentale avvalendoci della strategia *nested-PCR*. Nello specifico, il nostro obiettivo è stato quello di amplificare tutto il gene con una coppia di primers che risultasse specifica esclusivamente per PRB1, al fine di ottenere un clone di grandi dimensioni, da utilizzare successivamente come substrato in reazioni di PCR in presenza della coppia di primers HOPE 1F-HOPE 2R. In seguito ad analisi mediante il programma bioinformatico *primerBLAST*, abbiamo individuato la coppia PRBR1F- HOPE 4R come quella ottimale in termini di specificità per l'amplificazione dell'intero gene PRB1.

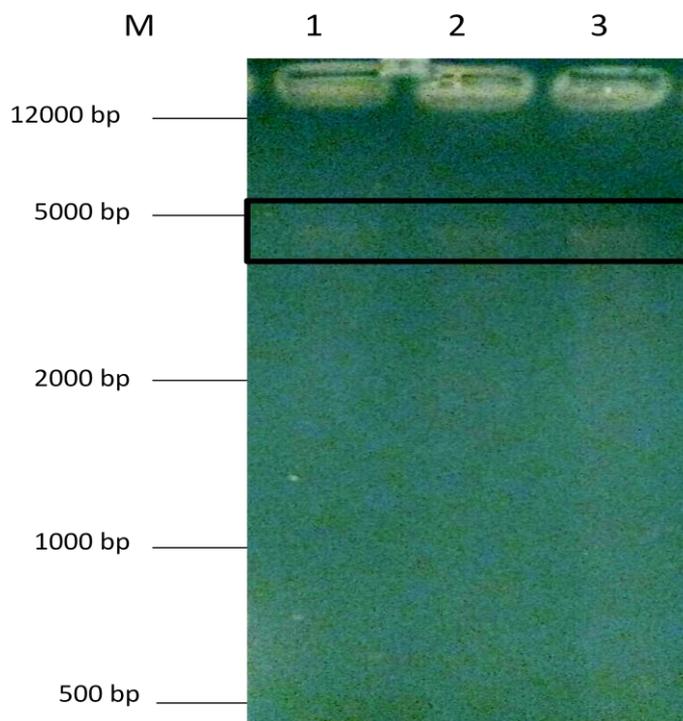


Fig. 17: elettroforesi su gel -di agarosio all'1,5% delle PCR PRBR1F-HOPE4R, nel riquadro sono evidenziate le bande relative ai frammenti di dimensioni attese (4557 bp), rispettivamente per il soggetto 1 (lane 1), 2 (lane 2) e 3 (lane 3). Il marcatore di pesi molecolari utilizzato (M) è il *marker 1kp plus* (ThermoFischer).

Nella figura 17 è riportato il profilo elettroforetico relativo ai prodotti delle PCR condotte in presenza dei primers PRBR1F-HOPE4R. In tutti i soggetti, l'amplificazione del clone atteso (4557 bp) è andata a buon fine, sebbene la bassa intensità delle bande sul gel indichi che la concentrazione del clone prodotto risulti essere relativamente bassa, probabilmente a causa della processività non ottimale della DNA polimerasi

utilizzata. Sebbene i profili elettroforetici non abbiano evidenziato frammenti di dimensioni indesiderate, i cloni sono stati purificati dal gel d'agarosio all'1,5%, al fine di ottenere per ciascuno una concentrazione idonea per un successivo impiego come substrati in reazioni di PCR con la coppia di primers HOPE 1F-HOPE 2R. Sulla base dei dati riportati in letteratura, inerenti le differenze tra le varianti di lunghezza del gene PRB1 (Lyons et al., 1988a,b), abbiamo stimato che le dimensioni dei cloni ottenuti mediante l'amplificazione del terzo esone con quest'ultima coppia di primers dovessero essere di 788 bp in presenza della variante *small*, 971 bp per la variante *medium* e infine 1154 bp per la variante *large* (tab. 4). La *nested-PCR* eseguita con la coppia di primers HOPE 1F-HOPE 2R ci ha consentito di amplificare, nel caso di tutti e tre i soggetti, cloni di dimensioni riconducibili alla variante *medium*, come da noi ipotizzato (fig. 18A). Coerentemente con i risultati ottenuti amplificando il DNA con la coppia PRB 2F-PRBR10R (fig. 13B), per tutti i soggetti è possibile osservare una banda elettroforetica di dimensioni riconducibili alla variante *small*, di intensità maggiore per il soggetto 3 rispetto agli altri due soggetti (fig. 18A lane 3), ed in linea con il genotipo atteso .

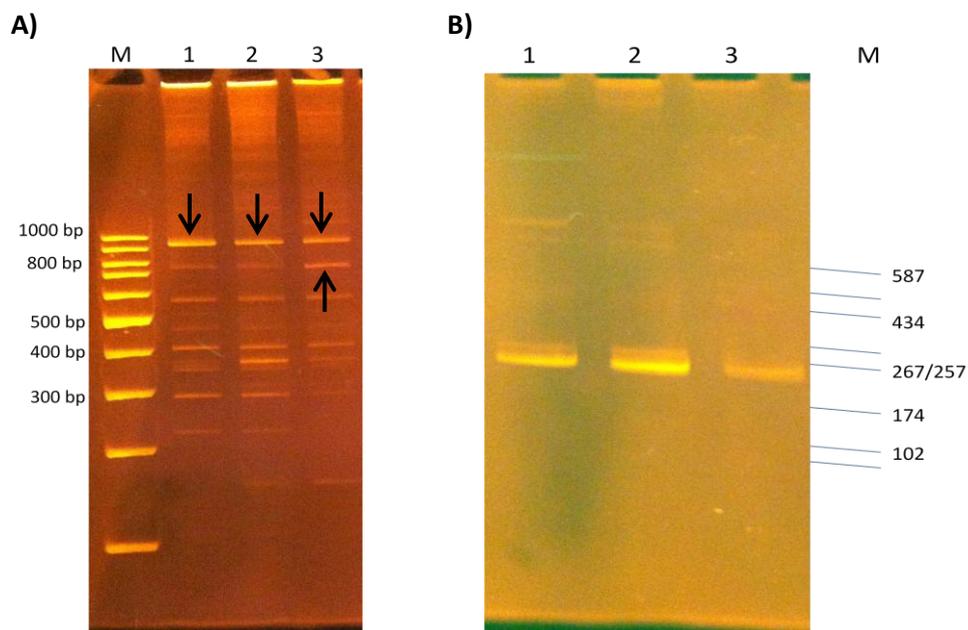


Fig. 18: **A)** elettroforesi su gel di poliacrilamide all'11% della PCR PRB HOPE1F-HOPE2R, i frammenti di taglia attesa sono indicati dalle frecce; per il soggetto 1 (lane 1), frammenti di dimensioni attese = 1154 / 971 bp; per il soggetto 2 (lane 2), frammento di taglia attesa = 971 bp; per il soggetto 3 (lane 3), frammenti di taglia attesa = 971 / 788 bp; il marcatore di pesi molecolari impiegato è il *Gene Ruler* 100bp (ThermoFischer); **B)** gel su acrilamide all'11% delle PCR PRB HOPE4F-HOPE4R, il frammento atteso ha dimensioni pari a 232 bp; lane 1: soggetto 1; lane 2: soggetto 2; lane 3: soggetto 3. Il marcatore utilizzato (M) è il puc18 *HaeIII* digested

Nel caso del soggetto 1 non è stato ottenuto alcun risultato relativamente al clone delle dimensioni di 1154 bp, come peraltro osservato anche nelle precedenti PCR con i primers PRB 2F-PRBR10R (fig. 13B). Questo risultato è in disaccordo con il genotipo ipotizzato sulla base del fenotipo precedentemente caratterizzato (tab. 1) e al momento non è possibile spiegare e commentare questo dato senza il supporto di ulteriori analisi a livello proteomico. Ad ogni modo, i cloni di dimensioni attese (quindi quelli delle dimensioni pari a 971 bp per tutti i soggetti, ed il clone delle dimensioni di 788 bp per il soggetto 3) sono stati purificati da gel di agarosio (1,8%) ed inviati ad un servizio di sequenziamento esterno. La PCR eseguita con la coppia di primers HOPE 4F-HOPE 4R ha consentito di amplificare il clone delle dimensioni attese in tutti e tre i soggetti (fig. 18A). Anche in questo caso la purificazione è stata eseguita ricorrendo all'isolamento del frammento di DNA da gel di agarosio, preparato ad una concentrazione maggiore (2%), al fine di ottenere una separazione elettroforetica più selettiva considerata la presenza di molecole di DNA aspecifiche nella miscela di reazione.

Sequenziamento dei cloni amplificati col terzo set di primers

L'analisi molecolare dei cloni ottenuti con il terzo set di primers ci ha permesso di chiudere i *gaps* presenti nelle sequenze nucleotidiche ricostruite utilizzando i risultati ottenuti nel corso dei due sequenziamenti precedenti (fig. 6-8 e 14-16). Successivamente, le intere sequenze sono state sottoposte ad analisi mediante *Blast Human*, per verificare la presenza di variazioni a singolo nucleotide ed eventualmente indicarne la posizione esatta sul cromosoma 12. I cromatogrammi relativi ai cloni HOPE 1F-HOPE 2R dei soggetti analizzati, sono stati allineati con quelli ottenuti nel corso dei sequenziamenti precedenti, ovvero: F2-R2 e F5-R5, ad eccezione del soggetto 3 (fig. 5C, lane 6), ed inoltre F2-PRBR10R e PRBR10F-PRBR11R. Analogamente, i tracciati cromatografici relativi ai cloni HOPE 4F-HOPE 4R sono stati allineati con quelli derivanti dal sequenziamento dei cloni F7-R7. Ciò ha permesso, nel caso dei soggetti 2 e 3 (fig. 20, 21 e 22), di ricostruire anche la lacuna nucleotidica relativa al sequenziamento dei cloni F7-R7 (fig. 7 e 8). Le sequenze di PRB1 ricostruite per i soggetti 1 e 2 (fig. 19 e 20) sono relative all'allele *medium*, in accordo con quanto osservato in seguito all'amplificazione dei cloni HOPE 1F-HOPE 2R e F2-PRBR10R (fig. 18A, lane 1 e 2; fig. 13B, lane 1 e 2).

Sequenza soggetto 1, variante Medium (ref. NC_000012.12)

11356319
tttctagaaatttcatggtggtttctgtgtctactttgcA/Gttaataaccctaaaaatactttagaatacaaaatt

11356243
ttagacttctgtgtgttacaggaccaagccttgccctcagctgagtgagatccacagaaacttgaacaatagccctat

11356165
cctgagcaggctgcggggcagctttctgtcatctatgtgtgcacagaagtgtgctgtcctcaca (nnnnnnnnnnnn

11356087
n) cagaaacctctctatcacatcaaaaattgagctaaatatgaaaatatctgtA/Gcacacacatcacaagagatgg

11356011
aaatagggcataaaaaG/TatC/GcagagatattgaccatT/CttcactgctggT/Cgtgagcatcgctgtcacaca

11355941
C/GacataaacagaaataaatagcacagtggaaagtggatG/Aatagctttacctaaaagattgaaactctaatagata

11355867
tctaaagttcagtgatgacacagtgctgatggcttgacacagttccttcaaacattggtcaagtcatagtttgtac

11355789
cttttgaaataagacactaacaacatatcctatccaactcatctctctgggctagagtctcaagaaaaataagggg

11355711
tacacctgacctgcagtgagcaaagcagaagccagtctctgaggtggtagggcccaccagtgaggctcaaag (nn

11355633
nnnnnnnnnnnn) tctctttataaagggagttgccacgttctctccagcacagagttgggagtgactccagagcct

11355555
ccT/AgcaagatgctggtgattctgctgtcagtggccT/CtgcctggccctgagctcagctcagaG/Acttaaatgaag

11355483
gtaaacagaaaggggaaaagatgtggtgactctgcttggggcttaggaggtgataatggtaattacggggaggagag

11355405
gagaatgaaaacagatggggctgcagagttttcatgcctaggatcaggagacC/Gtggtgtaccctcattccacac

11355329
taagggtttctaatttatttaattgtacaatgaaatccaataaagaattgttccaggggaatgagaaggttaagattg

11355251
catttatagatagaactgtgctgtgaaggatgcagtagagaatgcaaggcagattcatggaagtccagctgtgaagat

11355173
cctaaactgatctcagtaagtacacagggA/Ctgatgggtggccttgcgtgtacagtgcatcagcatcgatgatggcgat

11355097
aaacacacagatcagagatactgcagagacagagaattggatggaacacttgcctctgtttaactagagatacag

11355019
aaatatcagagccaatcattgtcatttttctctcccttacacgcagtatttcaatgtgctgggagtggtatgggtaag

11354941
attgtattgaagtgattacttctggttaccccaattgagaaagcatgtgtacataagcaatgtatttataggaA/Gat

11354865
ggagggcataagaacaccaaA/GtatcacattgaagtacctggcatG/Ttgtgaactaaattagcattaagtcttga

11354791
gggatgctagggagggaaaaaagggctcttctatggtgagttcatggctggtgctgtcataacaaccctgT/Cct

11354715
cccttacaccttctcccttccagcagcttcacagatgggtggctgacT/Gagttaacttaggggatgcatgggggtg

11354639
tcatgagaagA/TccC/GttttccctgtagaacacttgtgagcttgaaggttcaagatgtaactG/Tttcccatca

11354567
tctgtgcttctcttctagatgtcagccaggaagaatctccctccctaatagcaggtaaagcctgattcG/Attctca

11354491
atctgttT/Ctgactgtcttttctgcttatgagtgatcagttctccagtgcttcttatacaaaactttctttcag

11354415
gagttgattaatgtagtgcccctaacaatataggcactcttG/Atgcaaccttgattctgggcatcatgagcaggc

11354339
caccaaattgaacggcagagatgcttggcttagatgacaacaggagtggttA/GcctcatccccC/TggccaggaA

11354266
/GtgctgctgggagatgacagaaatgggcagcatcctcattctgtctctctttataA/Ctgagaggcoctcaactg

11354191
ctttgttcttcccagctttccactccagagttctatgtcttactgaaaatgcaaaGaaattaatgtctttgttcca

11354113
ttttgtgtatttcccactcagcttgttaccacagctgataaaaaattactgcaactattcagtgaaatcttgtgtgg

11354035
gcttttactctgtcttctcttctctcttctcttctcctcaggaaatccacaaggaccatccccacaaggaggcaacaagc

11353957
ccaAgggccccacctcctccaggaaagccacaaggaccacccccacaaggaggcaacaacctcaaggtccccac

11353879
ctccaggaaagccacaaggaccacccccacaaggagacaagtcccgaagtccccgatctcctccaggaaaaccacaag
 11353801
gaccacccccacaaggag↗gtaaccagccccaaggtccccacctcctccaggaaagccacaaggaccacccccaca
 11353723
ggaggcaaca↖acctcaaggtccccacctccaggaaagccacaaggaccacccccacaaggagacaagtcccaaagt
 11353645
ccccgatctcctccaggaaagccacaaggaccacccccacaaggaggtaaccaacc↖caaggtccccacctcctcca
 11353567
gaaagccacaaggaccacccccacaaggaggcaacaacctcaggtccccacctccaggaaag↖cacaaggacca
 11353489
ccccacaaggagacaagtcccgaagtcccgaatctcctccaggaaagccacaaggaccacccccacaaggaggcaac
 11353411
cagcctcaaggtccccacctcctccag↖gaaagccacaaggaccacccccacaaggaggcaacaacctcaaggtcc
 11353333
ccacctccag↖gaaagccacaaggaccacccgcacaaggaggcagcaagtccc↖aaagtgccga↖T/Gcctcctccagg
 11353257
aaagccacaaggaccacccccacaagaaggcaacaatcctcaaggtccccacctccagcaggaggcaatccccagca
 11353179
gcctcaggcacctcctgctggacagccccaggaccaccagccctcctcaagggggcagaccttccagacctcccc
 11353101
gtgacagcctcccagtcattctaggattcaatgacagg_{atgattccagtttattcttcaccaagtgctctaattgct}
 11353023
 acagctctccagctttattgtgccaatgaatcagctaaaagccattggcattgtatagtccagatcccattttctaa
 11352945
 agatttgattgacatattctggaatgggtaacaagatcctataattgtaacaaactctttaaggaattctgatgtt
 11352867
 gagaaacaaaattccaataatctgtcttaagttgtggtggcaacaaggaagtagtaccatgttctctctggcgctct
 11352789
 gttttctgtgcacaaactgagagacctccatttaaagttttcacctgagcactgtttgctcagtcctgcctcacacc
 11352711
 agcctcttgagtcaggtatttctgccaagtggtccctgaactttcagcagctaaatgggtgtctcatttttcaaattct
 11352633
 tactgttcaataagtagatgattaagctaacaaaaatataatgcaatggaaaaatgaatctaaatttaagggc
 11352555
 atgactcatcctacctgcctcccctccttcagaaaactgccactgttaactttatggcatcttctgtttgaaatattt
 11352477
 atgtgtacatagac↖Aactagaatatttttccccagaactaataccataatttatattcaggtacatatgttagtcat
 11352399
 ttaaaaaatacatttctttgaaaatttccacatacgtctatgaagtaagtagatctcttcagtggttatctgtttgt
 11352321
 ttttaccattttatactactccattatgtggctgcaccgtgatttctttaaccaatctgtgtcactggacactgaggg
 11352243
 tggtttcagcttctcactgttataaaatagttccagttcccatctgtgtaaatatctgtgaaacaaattcagcagc
 11352165
 aagtaataa↖C/Taagctaagaatgatcttctgtcttcatcacgtaaggaacaatttgagacattttgtgcaagggc
 11352089
 atccaaagagtgaacacacaaaaaattagggaggaaacacaggaggtagaaggatgggggagagaggatgggctctc
 11352011
 atgtactgtagtgcagtaagaccagtgaggaattcgacatttctgcatgtcaagctcgtgtctatgaacttcttct
 11351933
 ttgtttgtttcag↖gaagtgaataagaagatgagagtgattcaaatgattcaaatccatgacattggaaaaaggtcat
 11351855
catagctctaacttcaatataccaataaaataatcagcttg↖caatttctgattgtggtgtctgtttctcaatatt↖G/T
 11351779
 gtgaatgtgggatctgaggacc_

Fig. 19: sequenza del gene PRB1 relativa al soggetto 1, riferimento: NC_000012.12; le porzioni esoniche sono riportate in colore rosso sottolineate; racchiusi tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/); la freccia nera è subito a valle del sito donatore di splicing alternativo, la freccia verde e la azzurra sono subito a monte dei siti accettori di splicing alternativo per le varianti, rispettivamente, cp4 e cp5.

La variante *large*, che in base al fenotipo proteomico abbiamo ipotizzato di osservare per il soggetto 1, non è stata individuata né sequenziata. Infatti, sulla base dei risultati delle reazioni di PCR e dei relativi sequenziamenti, in particolar modo quelle relative al terzo esone, il soggetto 1 (analogamente al soggetto 2) risulta essere omozigote per la variante *medium*. Complessivamente, nella sequenza di PRB1 ricostruita per quest'ultimo soggetto, sono stati individuati 37 SNPs, 8 dei quali all'interno delle porzioni esoniche (fig. 19). Altrettanti SNPs sono presenti nella sequenza completa del gene PRB1 per il soggetto 2, 11 dei quali presenti all'interno degli esoni. Il sequenziamento dei due diversi frammenti relativi al clone HOPE 1F-HOPE 2R (fig. 18A, lane 3), ci ha consentito di ricostruire sia la variante *small* (fig. 22) che la variante *medium* (21) ascrivibile al soggetto 3. Le due varianti, come previsto, differiscono per 183 bp all'interno del terzo esone, e la delezione di tale porzione nucleotidica non comporta soppressione di alcuno dei siti di splicing alternativo (fig 22).

Sequenza soggetto 2, variante Medium (ref. NC_00012.12)

11356319
tctagaatcttctcgtggtggtttctgtgtctactttgcA/Gttaataacccttaaaaacttttagaatacaaaatctt
11356243
agacttcctgtgtgttacaggaccaagccttgccctcagctgagtgagatccacagaaacttgaacaatagccctatcc
11356165
tgagcaggctgcggggcagctttctgtcatctatgtgtgcacagaagtgtgctgtcctcacacatctcatagatagca
11356087
gaaacctctctatcacatcaaaaattgagctaaatgaaaatatctgtacacacacatcacaagagatggaaatag
11356009
ggcataataaT/GatC/Gcagagatattgaccattttcactgctggtgtagcatcgtgctcacacacataaaca
11355935
gaaataaatagcacagtggaagtggatgatagcttacctaaaagattgaaactctaatgatatctaaagttcagtg
11355857
atgacacagtgctgatggcttgacacagttccttcaaaaccattgttcaagtcatagttgtaccttttgaaataag
11355779
acactaacaacatctctatccaactcatctctctgggctagagctcaaagaaaaataagggatacactgacctgc
11355701
agtgagcaaagcagaagccagctctctgaggtggtgaggcccaccagtggtgagctcaaaggtgccattggt (nn) tc
11355623
tctctttataaaggaggtgccacgttccctccaGcacagagttgggagtgactccagagcctcctgcaagatgctg
11355545
ttgattctgctgtcagtgggcT/Ctctgctggcctgagctcagctcagaacttaaatgaaggtaaaacagaaggggaa
11355469
aagatgtggtgactctgcttgggcttaggagtgataatggtaattacggggagagaggagaatgaaaacacagat
11355391
gggctgcagagttttcatgcctaggatcaggagacctgttgtA/Gccctcattccacactaagggtttctaattat
11355315
ttaatgtacaatgaaatccaataaagaatgttccaggggaatgagaaggtaagatttgcatttatagatagaactg
11355237
tgctgtgaaggatgcagtagagaatgcaaggcagattcatggaagtccagctgtgaagatcctaaactgatctcagta
11355159
agtacacagggatgatggtggccttctgtgtacagtgcacatcagatcagatgaT/Cggcgataaacacacacagtatcag
11355081
agatactgcagagacagagaattggatggaacacttctctgttttaactagagatacagaaatatcagagccaatca
11355003
ttgtcatttttctctcccttacacgcagtatctcaatgtgctgggagtggtatgggtaagattgtattgaagtgatta
11354925
cttctggttaccccaattgagaaaagcatgtgtacataagcaatgtatttataggaA/Gatggagggcataagaacacc
11354849
aaaG/AtatcacattgaagtacctggcatG/Ttgtgaactaaattagcattaagtcttgagggatgctagggagggaa
11354775
aaaaggggctcttctatgttgagttcatggctgttctgtctgtcataacaaccctgT/Cctccccttacaccttctcc
11354697
ccttccagcagcttcacagatggtggctgaCT/GagtttaacttaggggatgcatggggtgtG/CatgagaagA/Tccc
11354625
ttttccctgtagaacacttgtgagcttgaaggttcaagatgtaacttttccatcatcctgtgcttctcttctag_a
11354547
tgtcagccaggaagaatctccctccctaataagcaggtataaG/Tcctgattcgttctcaatctgttttgactgtctttt
11354471
tctgcttatgaggtggatcagttctccagtgcttctcttatcaaaactttcctttcaggagttgattaatggttagtgcC
11354394
/TcctaacaatataggcactcttcG/AtgcaaccttgattctggcatcatgagcagggccaccaaattgaaC/Tggca
11354321
gagatgcttggcttagatgacaacaggagtggtgtgctcctcatccccC/Tggccaggaatgctgctgggagatgaca
11354245
gaaatgggcagcatcctcattctgtctcctctttataA/Ctgagaggccctcaactgctttgttcttcccagctttc
11354169
cactccagagttctatgtcttcaactgaaaatgcaaaG/Taaattaatgtctttgttccatttttgtgtatttcccac
11354093
tcagcttggtaccgaagctgataaaaaatttaG/Ctgcaactattcagtgaaatctgtgtgggcttttactctgtcttt
11354017
ctcttctctttgtcctccagGaaatccacaaggaccatccccacaaggaggcaacaagccccaAggccccccacct
11353939
cctccaggaagccacaaggaccacccccacaaggaggcaacaaacctcaaggtccccacctccaggaagccacaa

11353861
ggaccacccccacaaggggacaagtcccgaagtccccgatctcctccaggaaaaccacaaggaccacccccacaagga
 11353783
grgtaaccagccccaggtccccacctcctccaggaaagccacaaggaccacccccacaaggaggcaacaAacctca
 11353705
aggtccccacctccaggaaagccacaaggaccacccccacaaggagacaagtcccAaagtcccCgatctcctccagg
 11353627
aaagccacaaggaccacccccacaaggaggttaaccaGccccaggtccccacctcctccaggaaagccacaaggacc
 11353549
acccccacaaggaggcaaCaacctcaggtccccacctccaggaaagccacaaggaccacccccacaaggagacaa
 11353471
gtccgaagtCcccaatctcctccaggaaagccacaaggaccacccccacaaggaggcaaccagccTcaaggtcccc
 11353393
acctcctccagTgaaagccacaaggaccacccccacaaggagggaacaaaacctcaaggtccccacctccagTgaaa
 11353316
gccacaaggaccacccgcacaaggaggcagcaagtccaaagtgccgatctcctccaggaaagccacaaggaccacc
 11353238
ccaacaagaaggcaacaatcctcaaggtccccacctccagcaggaggcaatccccagcagcctcaggcaacctctgc
 11353160
tggacagccccaggaccaccacgccctcctcaagggggcagacctccagacctcccagtgacagcctccccagtc
 11353082
atctaggattcaatgacaggtatgattccagtttattcttcaccaagtgtctaatgtctacagctctccagctttat
 11353004
 tgtgccaatgaatcagctaaaagcccattggcattgtatagtcccagatcccatttctaaagatttgtattgacatat
 11352926
 tctggaaatgggtaacaagatcctatatatttgtaacaaactctttaaggaattctgatgttgagaaacaaaattccaaa
 11352848
 taatctgtcttaagttgtgttggaacaaggaagtagtaccatgttctctctggcgctctgttttctgtgcacaaact
 11352770
 gagagacctccatttaaagttttcacctgagcactgtttgctcagtcctgcctcacaccagcctcttgagtccagta
 11352692
 ttctgccaaagtggtccctgaactttcagcagctaaatgggtgtctcatttttcaaattcttactgttcaataagtaca
 11352614
 tgattaagctaacaaaaaatatctaataatgcaatggaaaaatataatctaaatttaaaggcatgactcatcctacctgc
 11352536
 ctcccctccttcagaaaactgccactgttaactttatggcatcttctgtttgaaatatttatgtgtacatagactact
 11352458
 agaataTTTTTCCCCagaactaatacT/Cataatttatattcaggtacatatgttagtcatttaaaaaatacatttc
 11352382
 tttgaaaatttccacatacgtctatgaagctaagtagatctcttcagtggttatctgtttgtttttaccattttatac
 11352304
 tactccattatgtggctgcaccgtgatttctttaaccaatctgtgtcactggacactgaggggtggtttcagcttctca
 11352226
 ctgttataaaatatgttccagttcccatctgtgtaaatatatactgtgaacaaattcagcagcaagtaataaT/Caagc
 11352150
 taagaatgatcttctgtcttcatcacgtaaggaacaatttgagacacattttgtgcaagggcatccaaagagtgaaca
 11352072
 cacaaaaaattaggaggaaacacaggaggtagaagggatAgggggagagaggatgggctctcatgtactgtagtga
 11351994
 gtaagaccagtgaggaattcgacatttccctgccatgtcaagtctggtctatgaacttccctG/Ctttgtttgtttcag
 11351918
gaagtgaataagaagatgagagtgattcaaatgattcaaatG/Tccatgacattggaaaaggtcatcatagctctaa
 11351842
cttcaatatacccatgacattggaaaA/TagggtcatcatagctctaaacttcaatataccaataaaataatT/Cagctt
 11351768
 gcaatttctgattgtgggtgtctgtttctcaatatttgtgaatgtgggatctgaggaccaaga

Fig. 20: sequenza del gene PRB1 relativa al soggetto 2, riferimento: NC_000012.12; le porzioni esoniche sono riportate in colore rosso sottolineate; racchiusi tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/); la freccia nera è subito a valle del sito donatore di *splicing* alternativo, la freccia verde e la azzurra sono subito a monte dei siti accettori di *splicing* alternativo per le varianti, rispettivamente, cp4 e cp5.

All'interno della variante *medium* del terzo soggetto sono state individuati 71 SNPs, di cui 16 all'interno degli esoni. Nessuno dei polimorfismi a singolo nucleotide individuati nei tre soggetti analizzati si trova in corrispondenza dei siti di splicing, sia costitutivi che alternativi. Inoltre, sebbene siano presenti dei polimorfismi a monte del sito accettore per la variante cp4, nel soggetto 2 (fig. 20) e del sito accettore per la cp5, nel soggetto 3 (fig. 21-22), appare evidente dal profilo peptidico osservato (tab 1) che questi SNPs non abbiano in alcun modo alterato la dinamica dello splicing in nessuno dei due soggetti. Alla luce di questo, è interessante constatare che nel profilo proteomico dei soggetti analizzati è presente solo il peptide Pk-o (tab. 1), codificato dalla variante di splicing cp4 (all'interno del terzo esone, sito donatore in 5' indicato dalla freccia nera e sito accettore in 3' indicato dalla freccia verde, fig. 19-22), sebbene non siano emersi ostacoli, a livello nucleotidico, per l'espressione della variante di splicing cp5 (all'interno del terzo esone, sito donatore in 5' indicato dalla freccia nera e sito accettore in 3' indicato dalla freccia azzurra, fig. 19-22). A tal proposito, è da sottolineare il fatto che finora non è stato individuato il peptide delle dimensioni previste, associato alla variante cp5 (Manconi et. al., 2015), sebbene il relativo messaggero fosse stato isolato già nel 1985 (Maeda et. al., 1985). La spiegazione possibile è che il relativo peptide possa essere suscettibile di proteolisi precocemente all'interno del cavo orale, o che la variante cp5 fosse esclusiva del soggetto analizzato dal gruppo di Maeda, mentre nella popolazione la suddetta variante sia scarsamente diffusa. Inoltre, è altamente probabile che sequenze di regolazione dello splicing *enhancer* e/o *silencer*, nell'intorno molecolare dei due siti accettori alternativi, giochino un ruolo determinante nel reclutamento specifico dei componenti dello spliceosoma, e conseguentemente nella espressione preferenziale della variante cp4 a scapito della cp5 (Smith & Valcàrcel, 2000). Ad ogni modo, è bene sottolineare che isoforme di mRNA frutto di splicing alternativo, specialmente se poco abbondanti, non danno luogo necessariamente a peptidi funzionali, ma possono essere frutto di errori dello spliceosoma (Lee & Rio, 2015).

Sequenza soggetto 3, variante Medium (ref. NC_000012.12)

11356085
aaccatctctatcacatcaaaaattgagctaaatatgaaaatatctgtG/Acacacacatcacaagagatggaaatag
11356009
ggcataaA/TaaG/TatG/CcagagatattgaccattttcactgctggC/Gtgtagcatcgctgctcacacacata
11355939
aacagaaataaatagcacagtggaagtggatgatagctttacctaagattgaaactctaataatgatatactaaagttc
11355861
agtgatgacacagtgctgatggcttgacacagttccttcaaaaccattgttcaagtcataagttgtaccttttgaaa
11355783
taagacactaacaacatctctatccaactcatctctctctgggctagagtcctcaagaaaaataagggatacactgac
11355705
ctgcagtgagcaaagcagaagccagtcctctgaggtggtagggcccaccagtgaggctc (nnnnnnnn) attggt
11355627
cttctcctctttataaaggagttgccagttcctcccagcacagagttgggagtgactccagagcctcctgcaagat
11355549
gctgttgattctgctgctcagtgcccttgctggccctgagctcagctcagaacttaaatgaaggtaaaaacagaagagg
11355471
aaaagatgtggtgactctgcttgggcttaggagtgataatggtaattacgggaA/Ggagaggagaatgaaaacac
11355395
agatgggctgcagagttttcatgcctaggatcaggagacC/GtgtgtG/AccctcattccacactaagG/Agtttc
11355323
taatttatttaagtacaatgaaatccaataaagaatttgttccaggggaatgagaaggtaagatttgcattataga
11355245
tagaactgtgctgtaaggatgcagtaC/GagaG/AtgcaaggcagattcatggaagtcagctgtgaagatcctaT/
11355173
AactgatctcagtaagtacacagggatgatggtggccttgctgtacagtgcatcagcatcgatgaT/Cggcgataaac
11355097
acacacagtatcagagatactgcagagacagagaattggatggaacacttgtctctgtttaactagagatacagaaat
11355019
atcagagccaatcattgtcattttctctccttacagcagatattcaatgtgctgggagtggtatgggtaagattg
11354941
tattgaaA/GtgattacttctggttaccccaattgagaaagcatgtgtacataagcaatgtatttataggaA/Gatgg
11354867
agggcataagaacacccaaaA/GtatcacattgaagtacctggcatT/Gtgtgaaactaaattagcattaagtcttgagg
11354793
gatgctagggagggaaaaaagggctcttctatgttgagttcatggctgttgctctgtcataacaacctgc/tctcc
11354715
ccttacaccttctccccttccagcagcttcacagatgggtggctgaC/TT/Gagttaacttaggggatgcatgggtg
11354641
tG/CG/AtgagaagA/TC/TC/GttttccctgtagaacacttgtgagtcctgaaaggttcaagatgtaactGttccc
11354573
atcatcctgtgcttctcttctagatgtcagccaggaagaatctccctccctaatagcaggtaaaT/GccC/TgattcA
11354499
/GttctcaatctggtC/TtgactgtcttttctgcttatgaA/Gtggatcagttctccagtgcttcttatcaaaact
11354427
ttcctttcG/AggaA/GttgattaatgtagtgcT/CcctaaC/TaatataggcactcttcA/Gtgcaaccttgattc
11354359
tgggcatcatgagcaggccaccaaattgaaCggcagagatgcttgcttagatgacaA/GcaggagtgtggtA/Gcct
11354285
catccccT/CggccaggaG/AtgctgctgggagatgacagaA/Caatgggagcagatcctcattctgtctcctctt
11354213
atactgagaggccctcaactgcttggcttctcccagcT/AttccactccagagttctatgtcttactgaA/Gaatg
11354139
caaaGaaattaatgtcttggcttccatttttgtgtattccccactcagcttggctacccaaA/GctgataaaaattC/T
11354065
actgcaactattcagtgaaatctgtgtgggcttttactctgtctttctctctctcttctctccaggaatccaca
11353987
aggaccatccccacaaggaggcaacaagcccaAggccccccactcctccaggaaagccacaaggaccacccccaca
11353909
aggaggcaacaaaactcaaggtccccacctccaggaaagccacaaggaccacccccacaaggAgacaagtcccgaag
11353831
tcG/Cccgatctcctccaggaaaaccacaaggaccacccccacaaggaggtaccagcccccaaggtccccacctcc

11353755
tccaggaaaA/GccacaaC/GgaccacccccacaaggaggcaacaCacctcaaG/CgtccccacctccaCgaaagcc
 11353683
acaaggaccacccccacaaggagacaagtcccgaatccccCgatctcctccaCCaaagccacaaggaccacccccaca
 11353605
aggaggtaaccaacccaaggtccccacctctccaggaaagccacaaggaccacccccacaaggaggcaagaaacc
 11353527
tcagggtccccacctccaggaaagccacaaggaccacccccacaaggagacaagtcccgaagtcccgaatctcctcc
 11353449
aggaaagccacaaggaccacccccacaaggaggcaaccagcccccaaggtccccacctctccag↵gaaagccacaag
 11353371
gaccaccccccaGaaggaggcaacaaacGtcaGgggtccccacctccag↵gaaagccacaaggaccacccgcacaagga
 11353293
ggcagcaagtcccaaagtgcccgatctcctccaggaaagccacaaggaccacccccacaagaaggcaacaatcctcaa
 11353215
ggtccccacctccagcaggaggcaatccccagcagcctcaggcacctcctgctggacagccccaggaccaccacgc
 11353137
cctcctcaagggggcagaccttccagacctccccagtgacagcctccccagtcactctaggattcaatgacaggtatga
 11353059
 tccagtttattcttcaccaagtgtcttaattgtctacagctctccagctttattgtgccaatgaatcagctaaaagcc
 11352981
 cattggcattgtatagtcccagatcccatttctaaagatttgtatT/Cgacatattctggaaatgggtaacaagatcc
 11352905
 tatatttgaacaaactctttaaggaattctgatgttgagaaacaaaattccA/TAaA/Ctaatctgtcttaagttgtg
 11352831
 ttggcaacaaggaagtagtaccatgttC/AatctctggcgctctgttttctgtgcacaaactgagagacctC/Gcatt
 11352757
 taaagtttcacctgagcactgtttgctcagtcctgcctcacaccagcctcttgagtccagttattcctgcaa (gtgg
 11352679
 tccctgaactttca) gcagctaaatgggtgtctcatttttcaaattcttactgttcaataagtagatgattaagctaac
 11352601
 aaaaaatatctaatgcaatgaaaaatatgaatctaaatttaaggcatgactcatcctacctgctccctccttca
 11352523
 gaaaactgccactgttaactttatggcatcttctgtttgaaatatttatgtgtacatagacAactagaatatttttcc
 11352445
 ccagaactaatacC/Tataatttatattcaggtacatatgttagtcatttaaaaaatacatttctttgaaaatttcc
 11352369
 acatacgtctatgaagctaagtagatctcttcagtggttatctgtttgtttttaccattttataactactccattatgt
 11352291
 ggctgcaccgtgatttctttaaccaatctgtgtcactggacactgaggggtggtttcagcttctcactgttataaaata
 11352213
 tgttccagttcccatctgtgtaaatatctgtgaaacaaattcagcagcaagtaataaT/Caagctaagaatgatctt
 11352137
 ctgtcttcatcacgtaaggaacaatttggagcacattttgtgcaagggcatccaaagagtgaacacacaaaaaattag
 11352059
 ggaggaaacacaggaggtagaaggatA/Gggggagagaggatgggctctcatgtactgtagtgcagtaagaccagtg
 11351983
 aggaattcgacatttctctgG/Ccatgtcaagctctgtctatgaacttcttctttgtttgtttcaggaagtgaataag
 11351907
aagatgagagtgattcaaatgattcaaatccatgacattggaaaaaggatcatcatagctctaacttcaA/Gtatacc
 11351831
aataaataatcagcttgcaatttctgattgtggtgtctgttttctcaatatttgtgaatgtgggatctgaggaccaa

Fig. 21: sequenza del gene PRB1 relativa al soggetto 3, riferimento: NC_000012.12; le porzioni esoniche sono riportate in colore rosso sottolineate; racchiusi tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/); la freccia nera è subito a valle del sito donatore di *splicing* alternativo, la freccia verde e la azzurra sono subito a monte dei siti accettori di *splicing* alternativo per le varianti, rispettivamente, cp4 e cp5.

Sequenza soggetto 3, variante Small (ref. NC_000012.12)

11356085

aaccatctctatcacatcaaaaattgagctaaatatgaaaatatctgtG/Acacacacatcacaagagatggaaatag

11356009

ggcataaA/TaaG/TatG/CcagagatattgaccatcttctactgctggC/Gtgtagcatcgctgctcacacacacata

11355939

aacagaaataaatagcacagtggaagtggatgatagctttacctaagattgaaactctaataatgatatctaaagttc

11355861

agtgatgacacagtgctgatggcttggacacagttccttcaaaaccattgttcaagtcataagttgtaccttttgaaa

11355783

taagacactaacaacatatcctatccaactcatctctctgtggctagagtctcaaagaaaataagggatacacctgac

11355783

ctgcagtgagcaaagcagaagccagtcctctgaggtggtgaggcccaccagtggtgagctc (nnnnnnnn) attggt

11355705

cttctcctctttataaagggagttgccagttcctcccagcacagagttgggagtgactccagagcctcctgcaagat

11355627

gctgttgattctgctgctcagtgcccttggctggccctgagctcagctcagaacttaaatgaaggtaaaacagaagagg

11355549

aaaagatgtggtgactctgcttggggcttaggagtgataatgtaattacgggaA/Ggagaggagaatgaaaacac

11355473

agatgggctgcagagtttcatgcctaggatcaggagacC/GtggtgtG/ACcctcattccacactaagG/Agtttc

11355401

taatttatttaagtacaatgaaatccaataaagaatttgttccaggggaatgagaaggtaagattgcatattataga

11355323

tagaactgtgctgtaaggatgcagtaC/GagaG/AtgcaaggcagattcatggaagtcagctgtgaagatcctaT/

11355251

AactgatctcagtaagtacacagggatgatggtggccttggctgtacagtgcatcagcatcgatgaT/Cggcgataaac

11355175

acacacagtatcagagatactgcagagacagagaattggatggaacacttgtctctgtttaactagagatacagaaat

11355097

atcagagccaatcattgtcatttttctctcccttacacgcagtatttcaatgtgctgggagtggtatgggtaagattg

11355019

tattgaaA/GtgattacttctggttaccccaattgagaaagcatgtgtacataagcaatgtatttataggaA/Gatgg

11354945

agggcataagaacacccaaaA/GtatcacattgaagtacctggcatT/Gtgtaactaaattagcattaagtcttgagg

11354871

gatgctagggagggaaaaaagggctcttctatgttgagttcatggctggtgctctgtcataacaacctgc/tctcc

11354793

ccttacaccttctccccttccagcagcttcacagatggtggctgaC/TT/Gagttaacttaggggatgcatggggtg

11354719

tG/CG/AtgagaagA/TcC/TC/GttttccctgtagaacacttgtgagcttgaaaggttcaagatgtaactGttccc

11354651

atcatcctgtgcttctcttctagatgtcagccaggaagaatctccctccctaataagcaggtaaaT/GccC/TgattcA

11354577

/GttctcaatctggtC/TtgactgtcttttctgcttatgaA/Gtgatcagttctccagtgcttcttctatcaaaact

11354505

ttcctttcG/AggaA/GttgattaatgtagtgcT/CcctaaC/TaatataggcactcttcA/Gtgcaaccttgattc

11354437

tgggcatcatgagcaggccaccaaattgaaCggcagagatgcttggcttagatgacaA/GcaggagtgtggtA/Gcct

11354363

catccccT/CggccaggaG/AtgctgtggtggagatgacagaA/Caatgggcagcatcctcattctgtctcctcttt

11354291

atactgagaggccctcaactgcttgttcttcccagcT/AttccactccagagttctatgtcttctactgaA/Gaatg

11354217

caaaGaaattaatgtcttggttccatttttgtgtatttcccactcagcttgttacccaaA/GctgataaaaattC/T

11354143

actgcaactattcagtgaaatctgtgtgggcttttactctgtctttctcttctcttcttctcttctctccaggaatccaca

11354065

aggaccatccccacaaggaggcaacaagcccaAggccccacctcctccaggaaagccacaaggaccacccccaca

11353987

aggaggcaacaaacctcaaggtccccacctccaggaaagccacaaggaccacccccacaaggAgacaagtcccgaag

11353909

tcG/Cccgatctcctccaggaaaaccacaaggaccacccccacaaggaggttaaccagccccagggtccccacctcc

11353833

tccaggaAA/GccacaaC/GgaccacccccacaaggaggcaacaCacctcaaG/CgtccccacctccaCgaaagcc

11353761
acaaggaccacccccacaaggagacaagtcccgaac**C**tccc**C**gatctcctcca**CC**aaagccacaaggaccacccccaca
 11353683
aggaggtaaccaacccaaggtccccacctcctccag**←**gaaagccacaaggaccacccccca**G**aaggaggcaacaaac
 11353605
Gtca**G**gggtccccacctccag**←**gaaagccacaaggaccacccgcacaaggaggcagcaagtcccaaagtgccgatct
 11353527
cctccaggaagccacaaggaccacccccacaagaaggcaacaatcctcaaggtccccacctccagcaggaggcaat
 11353449
cccagcagcctcaggcacctcctgctggacagcccagggaccaccaagccctcctcaagggggcagaccttcaga
 11353371
cctccccagtgacagcctccccagtcattctaggattcaatgacag**G**tatgattccagtttattcttccaccaagtgtct
 11353293
 taattgctacagctctccagctttattgtgccaatgaatcagctaaaagcccattggcattgtatagtccagatccc
 11353215
 atttctaagatttgat**T/C**gacatattctgaaatgggtaacaagatcctatatttgtaacaaactctttaaggaa
 11353139
 ttctgatgttgagaaacaaaattcc**A/T****A/C**taatctgtcttaagttgtgttggaacaaggaagtagtaccatgtt
 11353065
C/Atctctggcgctctgttttctgtgcacaaactgagagacctc**C/G**catttaaagttttcacctgagcactgtttgc
 11352991
 tcagtctgcctcacaccagcctcttgagtccagatattcctgcaa (gtggctcctgaactttca) gcagctaaatgg
 11352913
 tgtctcatttttcaaattcttactgttcaataagtacatgattaagctaacaaaaaatatctaagcaatggaaaaat
 11352835
 atgaatctaaatttaaggcatgactcatcctacctgcctcccctcctcagaaaactgccactgttaactttatggc
 11352757
 atcttctgtttgaaatatttatgtgtacatagac**A**actagaatatttttccccagaactaatac**C/T**ataatttata
 11352681
 ttcaggtacatatgttagtcatttaaaaaatacatttctttgaaaatttccacatacgtctatgaagctaagtagatc
 11352603
 tcttcagtggttatctgtttgttttaccattttatactactccattatgtggctgcaccgtgatttctttaaccaat
 11352525
 ctgtgtcactggacactgagggtggtttcagcttctcactgttataaaaatagttccagttcccatctgtgtaaatat
 11352447
 atctgtgaacaaattcagcagcaagtaataa**T/C**aagctaagaatgatcttctgtcttcatcacgtaaggaacaattt
 11352371
 ggagcacattttgtgcaagggcatccaaagagtgaacacacaaaaaattagggaggaaacacaggaggtagaagggat
 11352293
A/Ggggggagagagatgggctctcatgtactgtagtgcagtaagaccagtgagggaattcgacatttctctg**G/C**catgt
 11352219
 caagtctggctctatgaacttccttctttgtttgtttcag**gaagtgaataagaagatgagagtgattcaaatgattcaa**
 11352141
attccatgacattggaaaaaggtcatcatagctctaaacttca**A/G**tataccaataaaaataatcagcttgcaatctctg
 11352065
 atgtgtgtgtctgtttctcaatatttgtgaatgtgggatctgaggaccaa

Fig. 22: sequenza del gene PRB1 relativa alla variante *small* del soggetto 3, riferimento: NC_000012.12; le porzioni esoniche sono riportate in colore rosso sottolineate; racchiusi tra parentesi sono indicati i nucleotidi non decodificati; in carattere maiuscolo ed evidenziati in giallo sono indicati gli SNPs, i cui alleli presenti in eterozigosi sono separati da una barra diagonale (/); la freccia nera è subito a valle del sito donatore di *splicing* alternativo, la freccia verde e la azzurra sono subito a monte dei siti accettori di *splicing* alternativo per le varianti, rispettivamente, cp4 e cp5.

Analisi in silico

Al fine di valutare quali degli *SNPs* individuati nei soggetti da noi analizzati fossero già stati descritti, è stato eseguito un confronto con quelli depositati nella banca dati *dbSNP* (*The Single Nucleotide Polymorphism Database*) del portale *NCBI*. In tabella 5 sono riportati i codici identificativi di tutte le varianti a singolo nucleotide già descritte, i soggetti in cui sono state individuate e l'indicazione della porzione genica in cui sono situate. E' interessante osservare che un quarto dei 48 polimorfismi riportati ricorrono in tutti i soggetti analizzati. I polimorfismi non descritti in banca dati *dbSNP* sono riportati in tabella 6, con l'indicazione per ciascuno della posizione rispetto alla sequenza di riferimento NC_000012.12. Per alcuni dei polimorfismi rilevati nei soggetti di questo studio, abbiamo trovato un riscontro positivo per quanto concerne il locus, ma una mancata corrispondenza per quanto riguarda la sostituzione nucleotidica. La valutazione della presenza di potenziali sequenze di regolazione dello splicing enhancer o silencer, oltre che dei punti di ramificazione in prossimità dei siti accettori in 3' delle varianti cp4 e cp5, è stata eseguita mediante il software *Human Splicing Finder 3.0* (<http://www.umd.be/HSF3/HSF.htm>), selezionato tra i molti disponibili (Houdayer et al., 2008) in quanto integra algoritmi di vari software ed è continuamente aggiornato (Desmet et al., 2009).

SNP_ID	SOGGETTO 1	SOGGETTO 2	SOGGETTO 3	Esonico/Intronico
rs3983889	No	No	Si	Intronico
rs557678376	Si	Si	No	Intronico
rs3983890	Si	Si	Si	Intronico
rs3867481	Si	Si	Si	Intronico
rs3867482	Si	No	No	Intronico
rs3867483	Si	No	No	Intronico
rs539846705	No	No	Si	Intronico
rs4533125	Si	No	No	Intronico
rs200021729	Si	Si	No	Esonico
rs368963605	No	Si	Si	Intronico
rs115240812	No	No	Si	Intronico
rs369169524	No	Si	Si	Intronico
rs372010024	Si	Si	Si	Intronico
rs375954004	Si	Si	Si	Intronico
rs372159512	Si	Si	Si	Intronico
rs369585183	Si	Si	Si	Intronico
rs574591034	No	No	Si	Intronico
rs554685067	Si	Si	Si	Intronico
rs566289482	No	Si	Si	Intronico
rs774862229	Si	No	Si	Intronico
rs201400496	Si	No	Si	Intronico
rs777839431	Si	No	Si	Intronico
rs190530334	Si	Si	Si	Intronico
rs2272509	No	Si	Si	Intronico
rs114069696	Si	Si	Si	Intronico
rs552348584	Si	No	Si	Intronico
rs144282801	No	No	Si	Intronico
rs149993046	Si	Si	No	Intronico
rs10459058	Si	Si	Si	Intronico
rs112850606	No	No	Si	Intronico
rs10845341	Si	Si	Si	Esonico
rs201935302	No	No	Si	Esonico
rs188775422	No	No	Si	Esonico
rs112485268	No	Si	No	Esonico
rs111877208	No	Si	Si	Esonico
rs61930108	No	No	Si	Esonico
rs113485215	No	Si	No	Esonico
rs61930109	No	Si	No	Esonico
rs113897264	No	Si	No	Esonico
rs371841290	No	Si	No	Esonico
rs76504934	Si	Si	No	Esonico
rs4763685	Si	No	Si	Intronico
rs4081837	Si	Si	Si	Intronico
rs373510896	No	Si	No	Intronico
rs61930107	No	Si	No	Intronico
rs574095451	No	Si	No	Esonico
rs11612458	No	No	Si	Intronico
rs1863846	No	No	Si	Esonico

Tabella 5: Polimorfismi a singolo nucleotide individuati nei soggetti analizzati e già descritti in banca dati dbSNP.

Posizione SNP in cromosoma 12	SOGGETTO 1	SOGGETTO 2	SOGGETTO 3	Esonico/Intronico
11356307	Si	No	No	Intronico
11356037	No	No	Si	Intronico
11355357	No	No	Si	Intronico
11355220	No	No	Si	Intronico
11355216	No	No	Si	Intronico
11355175	No	No	Si	Intronico
11355907	Si	No	No	Intronico
11355559	Si	No	No	Esonico
11355501	Si	No	No	Esonico
11355358	Si	No	No	Intronico
11355150	Si	No	No	Intronico
11354936	No	No	Si	Intronico
11354643	No	Si	Si	Intronico
11354635	Si	No	Si	Intronico
11354632	No	No	Si	Intronico
11354638	Si	No	No	Intronico
11354513	No	Si	No	Intronico
11354497	No	No	Si	Intronico
11354500	No	No	Si	Intronico
11354468	No	No	Si	Intronico
11354465	No	Si	No	Intronico
11354422	No	No	Si	Intronico
11354418	No	No	Si	Intronico
11354401	No	Si	Si	Intronico
11354395	No	No	Si	Intronico
rs11054267 g=wt, ma è t/g	Si	No	Si	Intronico
11354180	No	No	Si	Intronico
11354150	No	No	Si	Intronico
11354069	No	Si	No	Intronico
11354072	No	No	Si	Intronico
11353747	No	No	Si	Esonico
rs199775084 ma come a/g	No	Si	Si	Esonico
11353709	No	No	Si	Esonico
11353694	No	No	Si	Esonico
11353654	No	No	Si	Esonico
11353637	No	No	Si	Esonico
rs563021727 transizione	No	No	Si	Esonico
rs776582994	No	No	Si	Esonico
rs140825288, transizione	Si	No	No	Esonico
11353577	Si	No	No	Esonico
11353367	No	No	Si	Intronico
11353385	Si	No	No	Esonico
11352946	No	No	Si	Intronico
11352861	No	No	Si	Intronico
11352859	No	No	Si	Intronico
11352770	No	No	Si	Intronico
11351888	No	Si	No	Esonico
11351767	Si	No	No	Intronico
rs3867483 transversione	No	No	Si	Intronico
rs539330997 c/t invece che c/a	No	No	Si	Intronico

Tabella 6: Polimorfismi a singolo nucleotide individuati nei soggetti analizzati e non descritti in banca dati dbSNP; sono indicate le posizioni di ciascuno SNP rispetto alla sequenza NC_000012.12, in rosso sono evidenziati i polimorfismi di cui è descritto in banca dati un diverso tipo di mutazione rispetto a quello individuato.

Tale analisi bioinformatica non ci ha permesso di formulare un'ipotesi che potesse essere plausibile per gli eventi di *splicing alternativo*, poiché nel caso di entrambe le varianti *small* e *medium* viene individuato un punto di ramificazione poco a monte dei rispettivi siti accettori in 3'. Inoltre, nell'intorno molecolare dei suddetti siti, il software individua molteplici sequenze potenzialmente riconosciute dai fattori di splicing (dato non mostrato), come ad esempio, quelli appartenenti alla famiglia delle proteine ricche in serina-arginina (SR proteins; Lareau et. al., 2015) o dalle ribonucleoproteine nucleari eterogenee (hnRNPs; Cáceres & Kornblihtt, 2002). Questa evidenza sperimentale non sorprende, in quanto risulta in linea con le caratteristiche intrinseche del gene PRB1 che, come già evidenziato in precedenza, è particolarmente ricco di ripetizioni in tandem, specialmente in corrispondenza del terzo esone. Inoltre, come riportato in letteratura, le sequenze riconosciute dai fattori di splicing sono molto corte (6-8 bp) e spesso degeneri (Smith & Valcárcel, 2000). Per chiarire il ruolo dei vari fattori in gioco, nello splicing costitutivo e alternativo di questo gene, sarà fondamentale individuare quali di questi fattori di splicing entrino in gioco specificamente all'interno delle cellule delle ghiandole parotidiche.



Fig. 23: mappa dei peptidi espressi a partire dalla variante *small* del soggetto 3, con l'indicazione dei codoni sotto ciascun aminoacido; evidenziati in giallo gli aminoacidi costituenti il peptide segnale; in celeste il peptide II-2; in verde il peptide PE; in grigio IB-6; sottolineati gli aminoacidi costituenti il peptide Pk-o, a partire dalla variante di *splicing* cp4.

Per poter valutare quali dei polimorfismi, individuati a livello esonico, potessero comportare delle sostituzioni a livello aminoacidico, abbiamo tradotto virtualmente le sequenze nucleotidiche dei tre soggetti, tramite l'apposito strumento del portale ExPASy, prendendo in considerazione solo le varianti di splicing costitutivo contenenti le variazioni rispetto alla sequenza *wild type*. Per il soggetto 3, è stata tradotta allo stesso modo anche l'allele *small*, in modo da poter ricostruire una mappa dei peptidi espressi sia da questa (fig. 23) che a partire dalla variante *medium* del soggetto 1 (fig. 24), e allinearle entrambe con le corrispettive sequenze nucleotidiche.

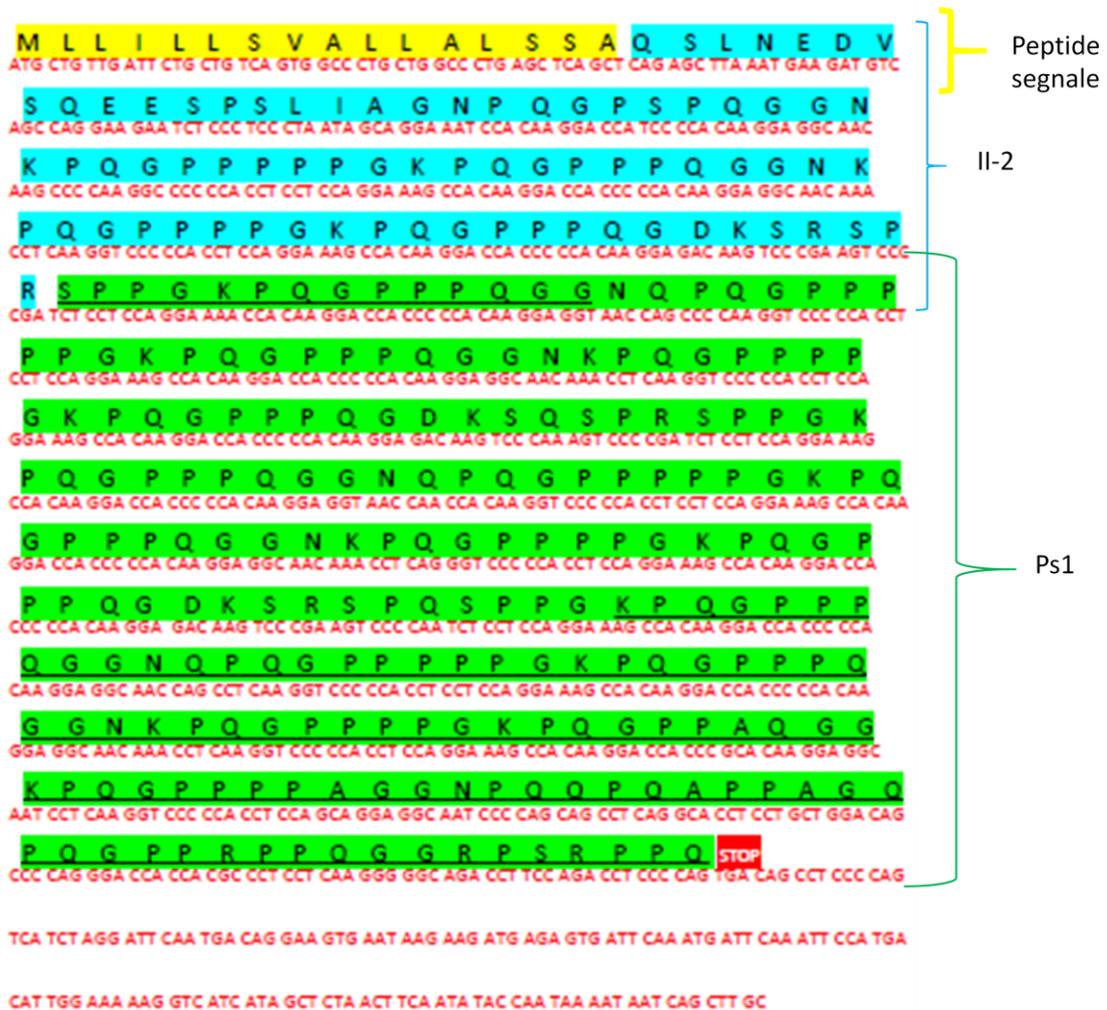


Fig. 24: mappa dei peptidi espressi a partire dalla variante *medium* del soggetto 1, con l'indicazione dei codoni sotto ciascun aminoacido; evidenziati in giallo gli aminoacidi costituenti il peptide segnale; in celeste il peptide II-2; in verde il peptide Ps1; sottolineati gli aminoacidi costituenti il peptide Pk-o, a partire dalla variante di *splicing* cp4.

Le sequenze tradotte sono state allineate con la sequenza di PRB1 depositata nella banca dati del portale *UniProt* (*accession number* P04280), che fa riferimento alla variante *large* del gene. Quest'ultima consta di 61 aminoacidi aggiuntivi rispetto alle varianti di lunghezza *medium* dei soggetti analizzati. Inoltre, in posizione 245 e 250, la sequenza di riferimento presenta due aminoacidi differenti rispetto a quelli riscontrati nelle sequenze dei tre soggetti nelle posizioni corrispondenti (rispettivamente, 184 e 189, fig. 25). Dall'allineamento non si evidenziano sostituzioni aminoacidiche all'interno della sequenza del soggetto 2. L'analisi della sequenza aminoacidica del soggetto 1, ha messo in evidenza la sostituzione dell'asparagina in posizione 18, subito a valle del peptide segnale (ricorrente sia nella sequenza di riferimento che in quella dei soggetti 2 e 3), con una serina. Essa è frutto della transizione A→G (non depositata in

banca dati) a livello del *locus* 11355501 del cromosoma 12 (tab. 6). Si tratta di un polimorfismo presente in eterozigosi con l'allele *wild type*, che coinvolge la seconda adenina del codone AAC all'interno del primo esone (fig. 26A). Nel caso del soggetto 3, come mostra l'allineamento (fig.25), un'arginina sostituisce la glicina in posizione 120, quest'ultima presente nella sequenza di riferimento oltre che nei soggetti 1 e 2. La sostituzione aminoacidica deriva da un evento di transversione nucleotidica G→C a carico della prima guanosina del codone GGA (fig. 26B) all'interno del terzo esone. Nella banca dati *dbSNP*, è stato depositato uno SNP con codice identificativo rs776582994, relativo ad una transizione in corrispondenza dello stesso *locus* (tab. 6) Risulta arduo fare considerazioni riguardo il significato funzionale di entrambe queste sostituzioni, in virtù del fatto che non abbiamo elementi in merito al ruolo effettivo e specifico di ciascun peptide espresso a partire dalla variante presa in esame. Ipotizzando un possibile ruolo della serina in posizione 18 nel soggetto 1 come potenziale sito di fosforilazione, abbiamo esaminato i dati presenti in letteratura e relativi alla fosforilazione delle PRP salivari, per individuare quale/i chinasi siano coinvolte nel processo e quale sequenza riconoscano. Allo stato dell'arte, la chinasi che sembra essere coinvolta nel processo di fosforilazione della maggior parte del proteoma che segue la via secretoria, inclusi quindi i peptidi derivanti da PRB1, risulta essere la Fam20C, che riconosce il motivo aminoacidico S-x-E/pS (Tagliabracci et. al., 2015) o, alternativamente, il motivo SXQXXDEE (Brunati et. al., 2000). Alla luce di questo, essendo la serina nel peptide II-2 del soggetto 1 (fig. 25) un residuo aminoacidico a monte rispetto ad una lisina e un'asparagina, non può essere fosforilata. L'unico bersaglio molecolare della Fam20C è la Serina in posizione 24, quindi sempre nella porzione amino terminale di II-2, poiché situata a monte di una glutammina e di un residuo di acido glutammico. La sostituzione di un aminoacido polare (glicina) con uno a carattere basico (arginina) potrebbe avere determinare variazioni funzionali a carico del peptide Ps1 (fig. 24), ma non strutturali, dal momento che i peptidi salivari rientrano nell'ampia famiglia delle proteine intrinsecamente disordinate (IDPs; Boze et. al., 2010), dunque una sostituzione aminoacidica non è sufficiente in questo caso a determinare stravolgimenti nella conformazione del peptide Ps1. Tutti gli altri SNPs individuati a livello esonico costituiscono sostituzioni sinonime, dunque, come previsto in base al profilo peptidico, i tre soggetti non presentano modificazioni in corrispondenza dei siti di taglio convertasico (Chan & Bennick, 2001) che danno origine ai vari peptidi derivanti da PRB1 (fig. 23-24). Tuttavia, seppure silenti dal punto

di vista traduzionale, questi polimorfismi possono influenzare lo splicing alternativo, ad esempio distruggendo o creando nuovi elementi *in cis*, che possano essere riconosciuti da fattori di *splicing* (Lareau et. al., 2015). Inoltre, è possibile che possano determinare cambiamenti nella struttura secondaria del pre-mRNA o comunque modificare la struttura del gene in modo da alterare la processività della RNA Polimerasi II e, conseguentemente, influenzare la scelta dei siti di splicing (Caceres & Kornblihtt, 2002).

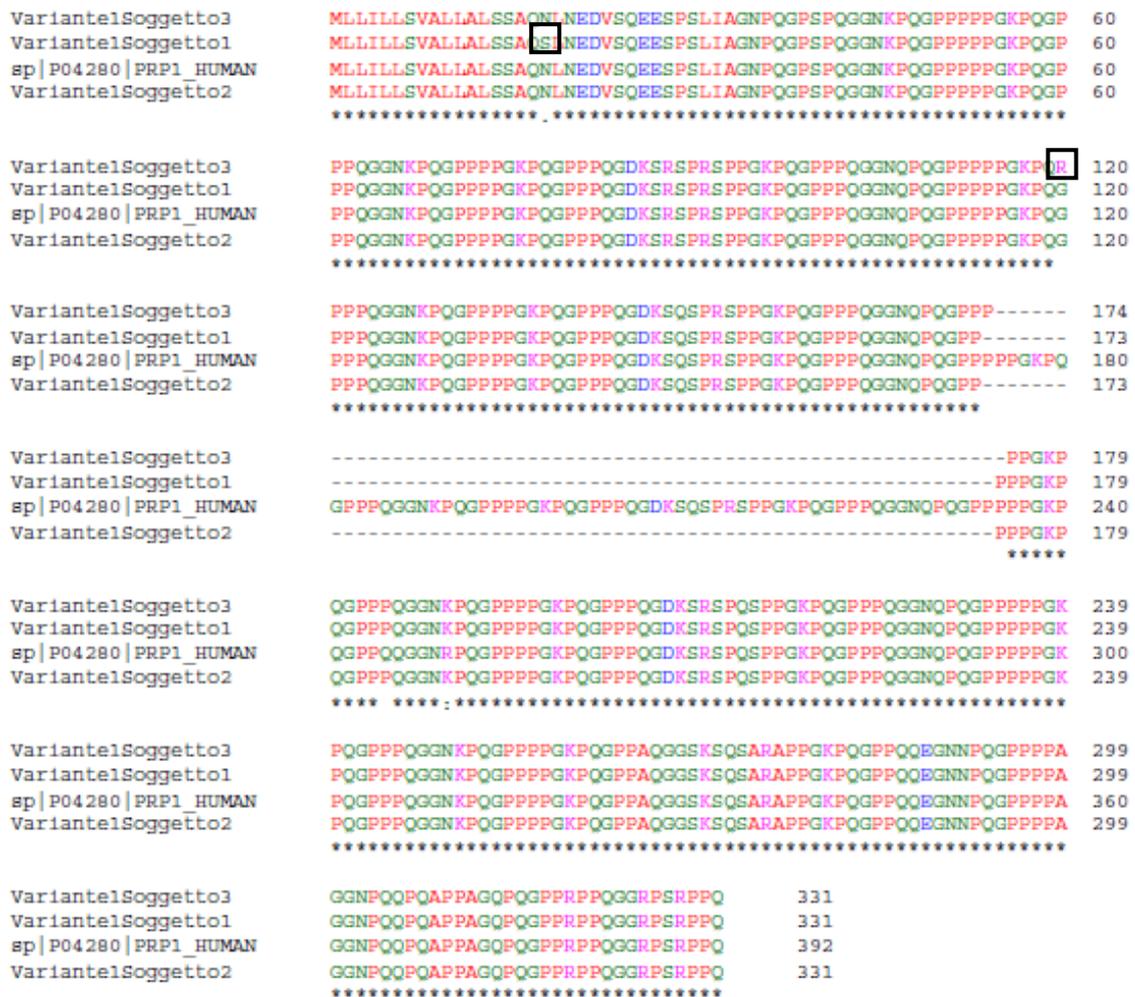


Fig. 25: Allineamento mediante *Clustal Omega* delle sequenze aminoacidiche di PRB1 dei tre soggetti analizzati con la sequenza aminoacidica di PRB1 presente in banca dati *UniProt* (accession: P04280). Tutte le sequenze sono relative alla variante di *splicing* costitutiva del gene. Nei riquadri sono evidenziati gli aminoacidi differenti tra la sequenza di riferimento e le sequenze dei tre soggetti.

La differenza tra le varianti di lunghezza di PRB1 è determinata dal numero di ripetizioni in tandem contenute all'interno del terzo esone, come mostrato dalle analisi

di Lyons e colleghi (Lyons, Stein et. al.,1988). Questo gruppo ha evidenziato, all'interno di tutto il terzo esone della variante *medium*, la presenza di 15 ripetizioni in tandem, costituite da 63 bp ciascuna. La differenza tra questa variante e quelle *small* e *large* era ascrivita alla delezione o alla inserzione di 3 delle suddette ripetizioni rispettivamente. Ciò è in contrasto con i dati da noi osservati, in quanto la differenza tra le varianti *medium* e *small* del soggetto 3 è riconducibile esattamente a 183 bp (fig. 21-22). Inoltre, tra la variante *large* presente in UniProt (P04280) e le varianti *medium* dei nostri soggetti, ci sono esattamente 61 residui aminoacidi di differenza (fig. 25), valore perfettamente coerente con la differenza osservata tra le sequenze delle due varianti di PRB1 del soggetto 3.

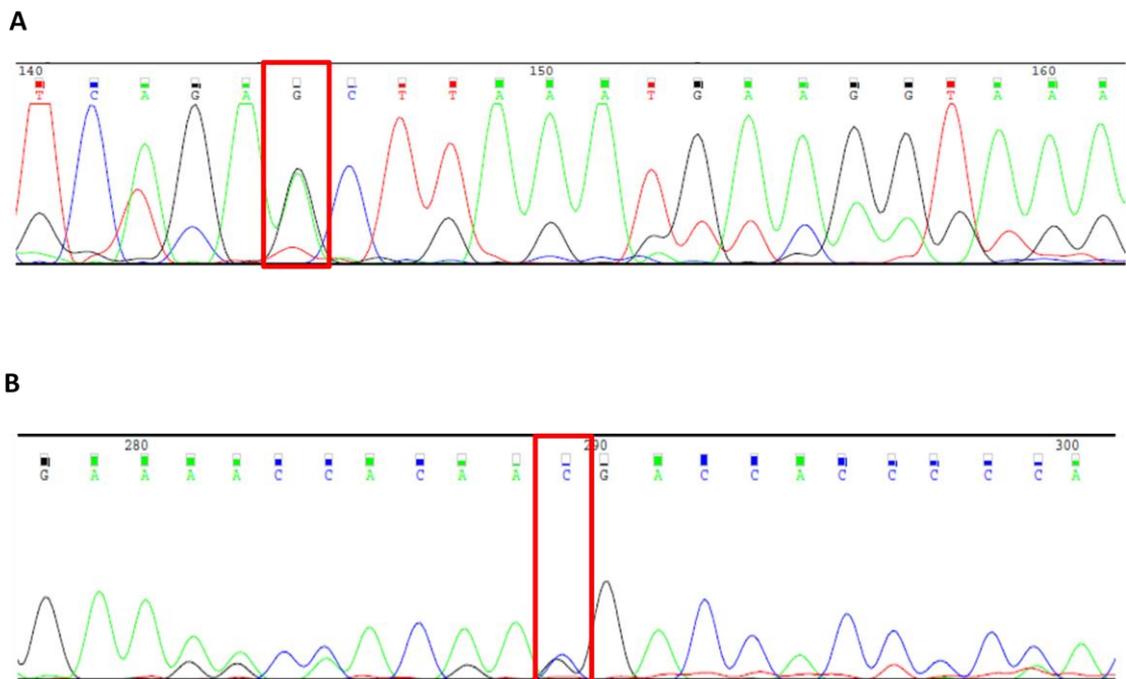


Fig. 26: **A)** porzione del cromatogramma relativa al frammento PRBR4F-PRBR6R per il soggetto 1, nel riquadro evidenziato il polimorfismo in eterozigosi che comporta transizione A→G nella seconda adenina del codone AAC, e conseguente sostituzione di una asparagina con una serina in posizione 2 della proteina II-2; **B)** porzione del cromatogramma del frammento PRB HOPE1F-HOPE2R per il soggetto 3, nel riquadro evidenziato il polimorfismo in eterozigosi che comporta transversione G→C nella prima guanosina del codone GGA, con conseguente sostituzione di una glicina con una arginina in posizione 120 della proteina Ps1

Osservando la struttura molecolare del terzo esone nei tre soggetti, ci siamo resi conto che è possibile suddividerne la sequenza (escludendo una porzione da 71 bp in 5' e un'altra da 126 in 3') in quattro ripetizioni in tandem costituite da 183 bp ciascuna, l'ultima delle quali risulta leggermente divergente dalle tre precedenti nella propria

porzione 3' (fig. 27-29). La sequenza delle prime tre ripetizioni, se tradotta virtualmente, risulta corrispondere esattamente alla porzione aminoacidica supplementare presente nella variante *large* della sequenza *con accession number* P04280. Questa osservazione ci ha portato a formulare un'ipotesi inerente l'origine delle varianti di lunghezza di PRB1, che si discosta leggermente da quella formulata da Lyons, Stein e colleghi nel 1988.



Fig. 27: rappresentazione schematica delle ripetizioni in tandem nel terzo esone (in rosso) della sequenza del soggetto 1; ciascuna ripetizione è rappresentata come un blocco, comprendente 183 bp; in giallo sono evidenziati i SNPs, gli alleli delle eterozigosi sono separati da una barra diagonale (/).



Fig. 28: rappresentazione schematica delle ripetizioni in tandem nel terzo esone (in rosso) della sequenza del soggetto 2; ciascuna ripetizione è rappresentata come un blocco, comprendente 183 bp; in giallo sono evidenziati i SNPs, gli alleli delle eterozigosi sono separati da una barra diagonale (/).

Secondo il modello proposto da questi studiosi, le varianti di lunghezza di PRB1 differiscono l'una dall'altra, passando dalla variante *small* alla variante *large*, per multipli di 3 delle ripetizioni in tandem. Ciò significa che, passando da una variante allelica all'altra, dovremmo trovare differenze nucleotidiche di 189 bp, dato numerico che si discosta di sei nucleotidi rispetto a quanto abbiamo osservato. Viene anche descritta la possibile origine della variante *small* dalla *medium*, per appaiamento sfalsato tra le ripetizioni in tandem costituite da 63 bp contenute all'interno del terzo esone (le prime tre per l'esattezza), che determinerebbe ricombinazione omologa ineguale, a livello di una delle ripetizioni stesse.



Fig. 29: rappresentazione schematica delle ripetizioni in tandem nel terzo esone (in rosso) della sequenza del soggetto 3; ciascuna ripetizione è rappresentata come un blocco, comprendente 183 bp; in giallo sono evidenziati i SNPs, gli alleli delle eterozigosi sono separati da una barra diagonale (/).

In questo modello è previsto che la variante *small* così generata contenga, oltre alla delezione di tre ripetizioni in tandem, una serie di sostituzioni nucleotidiche rispetto agli alleli ricombinanti, ma non viene però chiarita l'origine di tali sostituzioni. E' da tener presente che le sequenze da 63 bp individuate da Lyons e colleghi sono sequenze consenso, che tuttavia variano considerevolmente all'interno del terzo esone. Questo aspetto potrebbe spiegare la necessità di inserire, nel loro modello, le sei sostituzioni nucleotidiche all'interno dell'allele ricombinante. Il modello da noi proposto tiene conto delle dimensioni precise delle inserzioni/delezioni che contraddistinguono le singole varianti di lunghezza e non prevede sostituzioni nucleotidiche rispetto agli alleli

ricombinanti. Partendo dal presupposto che la variante *medium* sia quella più diffusa nella popolazione (Lyons, Stein et. al., 1988), è possibile che ci sia stato un appaiamento sfalsato tra due alleli contenenti tali varianti, in corrispondenza delle prime due ripetizioni di un allele, con la seconda e terza dell'altro allele (fig. 30). In seguito, per mezzo di *crossing-over* ineguale, hanno avuto origine: un allele contenente la variante *large*, con l'inserzione della seconda ripetizione proveniente dall'altro allele; un altro allele contenente la variante *small*, con la delezione della seconda ripetizione (fig 30). Riteniamo che l'appaiamento sfalsato e la conseguente ricombinazione, debba aver coinvolto esclusivamente le prime tre ripetizioni in tandem, aventi un maggior livello di omologia reciproca (superiore al 97%) rispetto all'ultima delle quattro (fig. 27-29).

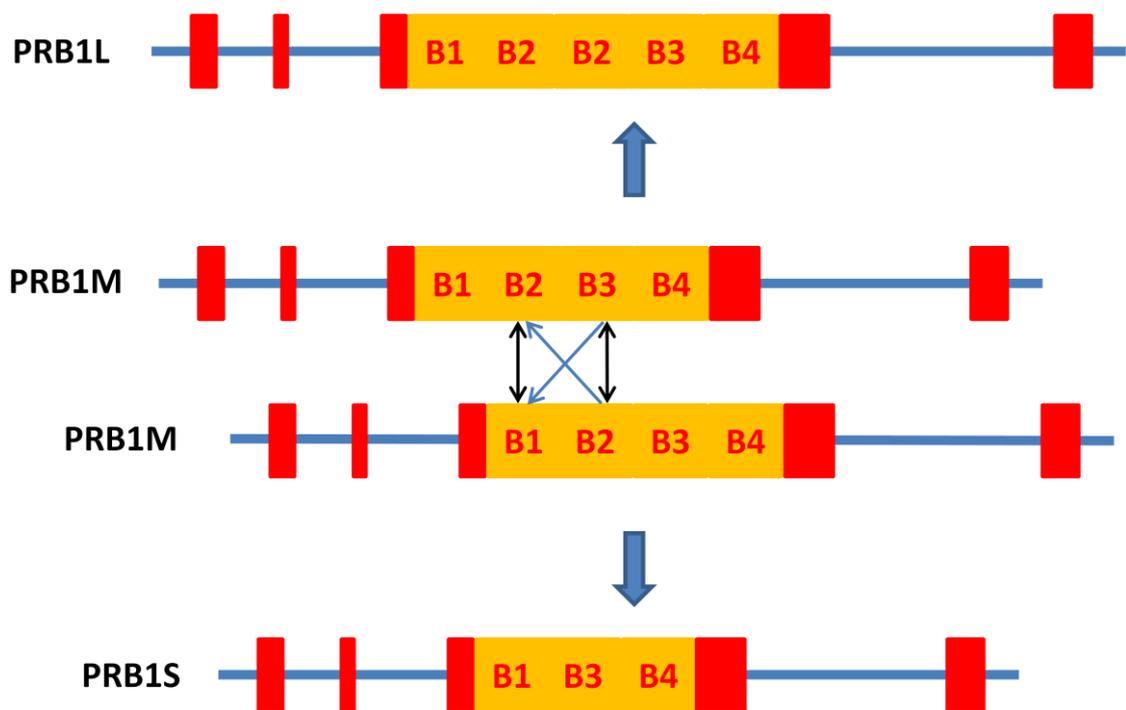


Fig. 30: modello di *crossing-over* ineguale tra due alleli contenenti l'allele *medium* di PRB1; le frecce nere indicano i blocchi che si allineano, le frecce azzurre indicano la direzione di ricombinazione; la ricombinazione produce sono un allele contenente la variante *large*, in cui si realizza l'inserzione del secondo blocco di un allele ricombinante, il quale a sua volta per delezione di questo blocco, darà origine a una variante *small*

CONCLUSIONI

Le informazioni riportate in letteratura inerenti la classe di geni PRP, e in particolar modo PRB1, oggetto di questo studio sono, dal punto di vista della caratterizzazione molecolare, frammentarie e di difficile ricostruzione. La letteratura su questa tematica abbraccia infatti un arco temporale che va dagli anni settanta (Azen & Oppenheim, 1973), in cui sono stati caratterizzati per la prima volta i peptidi associati, fino a fine anni novanta (Stubbs et al., 1998). A partire dagli anni ottanta, sono stati fatti vari tentativi di caratterizzazione delle varianti alleliche di PRB1 sulla base dei peptidi derivanti, e tale approccio ha complicato enormemente il lavoro di classificazione, sia dal punto di vista proteomico che molecolare (Maeda et al., 1985; Lyons et al., 1988a,b; Azen et al., 1993). Oggi sappiamo che a partire da PRB1, considerando le sue varianti alleliche e di splicing, sono espresse sei differenti proteine (tab. 1; Marconi et al., 2015). Tuttavia, delle tre varianti alleliche di tale gene, quattro se si considera la variante *very large* (della quale però non abbiamo individuato potenziali soggetti portatori da analizzare sotto il profilo molecolare), l'unica di cui si abbia la sequenza nucleotidica completa è la *medium*, mentre della *small* e della *large* si hanno finora solo sequenze parziali, riferite al solo terzo esone. Un esempio eclatante della confusione generata dalla letteratura in materia è costituito da un recente studio di associazione tra i livelli di espressione dei peptidi associati all'inibizione dell'infezione da HIV-1 (Burgener et al., 2012). Basandosi su evidenze precedenti (Robinovitch et al., 2001), gli autori hanno saggiato l'associazione della neutralizzazione del virus con i livelli di PRB2, partendo dal presupposto che tutti i peptidi coinvolti nel legame alla proteina capsidica gp-120 derivassero da tale gene. In realtà, i peptidi coinvolti (Robinovitch et al., 2001) hanno peso molecolare apparente compatibile con i peptidi derivanti da tutti e quattro i geni PRB ed il ruolo specifico di ciascuno di essi è ancora tutto da esplorare.

Tra gli obiettivi di questo studio c'era appunto la ricostruzione delle sequenze complete delle varianti *small* e *large*, delle quali sono stati preventivamente individuati e caratterizzati dal punto di vista proteomico due potenziali portatori, indicati in questo lavoro rispettivamente come soggetto 3 e 1. Se per il soggetto 3 ci è stato possibile individuare e sequenziare la variante allelica *small*, presente nel genoma di questo soggetto in eterozigosi con la variante *medium*, lo stesso non si può dire per il soggetto 1. Infatti, dalle nostre analisi è emerso che il soggetto 1 è portatore in omozigosi della variante *medium*, incompatibile con la presenza nella sua saliva della proteina Ps2 (tab.

1). In assenza di ulteriori analisi di spettrometria di massa, non possiamo spiegare il perché di tale discrepanza tra i dati genomici e proteomici. L'analisi delle sequenze complete non ci ha permesso di comprendere per quale ragione non sia stato individuato allo spettrometro di massa il peptide relativo alla variante di splicing classificata da Maeda con l'acronimo cP5 (Maeda et al., 1985), mentre in tutti i soggetti è stato individuato il peptide Pk-o e la relativa variante di splicing cP4, che finora è stata riportata in letteratura espressa a partire dalla variante *large*, sebbene la prima classificazione delle varianti di splicing fosse stata fatta a partire da un individuo portatore della variante *medium*. Dal momento che in nessuno dei tre soggetti sequenziati il sito accettore di splicing in 3' relativo alla variante cP5 risulta alterato, è possibile che il relativo peptide non sia stato individuato allo spettrometro di massa, o che fattori di splicing specifici riconoscano sequenze elementi regolatori dello splicing (SREs, Splicing Regulatory Elements; Hernandez-Imaz et al., 2015) nell'intorno molecolare dei due siti accettori di splicing alternativo, impedendo la trascrizione di cP5 a favore di cP4. Per poter fare luce su questo aspetto, esistono alcune possibili strategie adottabili in prospettiva. Recentemente, infatti, sono stati pubblicati studi in cui vengono accoppiati estensivi screening *in vivo* per l'individuazione di SREs, alla purificazione per affinità all'RNA e alla spettrometria di massa, che consentono di identificare anche i fattori di splicing agenti sugli SREs individuati in vivo (Wang Y & Wang Z, 2014; Wang et al., 2013). Tale strategia potrebbe essere efficacemente integrata con l'analisi di minigeni (Hernandez-Imaz et al., 2015) progettati appositamente sul modello del terzo esone di PRB1, che costituisce la sede dello splicing alternativo, e probabilmente anche degli SREs, in questo gene.

L'analisi molecolare ci ha altresì consentito di individuare una serie di SNPs, molti dei quali mai descritti finora, la maggior parte localizzati a livello intronico e alcuni anche all'interno degli esoni. In particolare, per due di questi abbiamo previsto una sostituzione non sinonima a livello aminoacidico, il cui ruolo potrà essere chiarito solo mediante appositi studi funzionali. Sebbene silenti dal punto di vista traduzionale, anche gli SNPs sinonimi possono avere una funzione importante, specialmente nello splicing alternativo. Infatti, tali sostituzioni possono determinare la genesi o la distruzione di SREs, o rafforzare siti donatori o accettori criptici. Inoltre, possono comportare l'alterazione della struttura secondaria dell'mRNA, importate sia per la definizione degli esoni che per i siti di pausa della RNA Polimerasi II, che in tal modo può andare

incontro ad alterazione della propria processività, con possibili conseguenze sulla scelta dei siti di splicing.

La ricostruzione delle sequenze complete e la comparazione con i dati presenti in letteratura (Lyons et al., 1988) ci ha permesso di formulare un modello aggiornato, che consenta di spiegare la generazione delle varianti alleliche *small* e *large* a partire dalla variante *medium*. Auspicabilmente, la ricostruzione della variante *large*, ed eventualmente della *very large*, potrà porre le basi per la validazione del nostro modello ed, eventualmente, anche la sua estensione alle varianti alleliche degli altri geni della classe PRB.

La caratterizzazione genotipica delle varianti alleliche di PRB1 potrà essere in futuro uno strumento predittivo importante riguardo la suscettibilità soggettiva nei confronti di una serie di patogeni del cavo orale. Infatti, le variazioni polimorfiche nei peptidi derivanti possono costituire la base per predire le differenze individuali al livello della microflora orale, con ricadute sulla suscettibilità a infezioni e patologie confinate in questo distretto corporeo (Newman et al., 1993 e 1996; O'Sullivan et al., 2000).

BIBLIOGRAFIA

- Amado, F., Lobo, M. J. C., Domingues, P., Duarte, J. A., & Vitorino, R. (2010). Salivary peptidomics. *Expert Review of Proteomics*, 7(5), 709–21.
<http://doi.org/10.1586/epr.10.48>
- Ann, D. K., & Lin, H. H. (1998). Transcriptional regulation of salivary proline-rich protein gene expression. *Annals of the New York Academy of Sciences*, 842, 108–114. <http://doi.org/10.1111/j.1749-6632.1998.tb09638.x>
- Azen, E. A. (1993). Genetics of salivary protein polymorphisms. *Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists*, 4(3-4), 479–85.
<http://doi.org/10.1177/10454411930040033201>
- Azen EA, Denniston CL (1974) Genetic polymorphism of human salivary proline-rich proteins: further genetic analysis. *Biochem. Genet.* 12, 109-120.
- Azen EA, Denniston C (1980). Polymorphism of Ps (parotid size variant) and detection of a protein (PmS) related to the Pm (parotid middle band) system with genetic linkage of Ps and Pm to G1, Db and Pr genetic determinants. *Biochem Genet* 18:483-501.
- Azen EA, Denniston C (1981) Genetic polymorphism of PIF (parotid isoelectric focusing variant) proteins with linkage to the PPP (parotid proline-rich protein) gene complex. *Biochem. Genet.* 19,475-485.
- Azen EA, Hurley CK, Denniston C (1979). Genetic polymorphism of the major parotid salivary glycoprotein (G1) with linkage to the genes for Pr, Db and Pa. *Biochem Genet* 17:257-279.
- Azen, E. A., Minaguchi, K., Latreille, P., & Kim, H. S. (1990). Alleles at the PRB3 locus coding for a disulfide-bonded human salivary proline-rich glycoprotein (G1 8) and a null in an Ashkenazi Jew. *American Journal of Human Genetics*, 47(4), 686–97.
- Azen, E. A., Latreille, P., & Niece, R. L. (1993). PRB1 gene variants coding for length and null polymorphisms among human salivary Ps, PmF, PmS, and Pe proline-rich proteins (PRPs). *American Journal of Human Genetics*, 53(1), 264–278.
- Azen, E. A., O'Connell, P., & Kim, H. S. (1992). PRB2/1 fusion gene: a product of unequal and homologous crossing-over between proline-rich protein (PRP) genes

PRB1 and PRB2. *Am J Hum Genet*, 50(4), 842–851. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/155012>

- Azen EA, Yu PL (1984a). Genetic determinants of Pe and Po salivary proteins with probable linkage of their genes to the salivary protein gene complex (SPC). *Biochem Genet* 22:1065-1080.
- Azen EA, Yu PL (1984b). Genetic polymorphisms of CON1 and CON2 salivary proteins detected by immunologic and concanavalin A reactions on nitrocellulose with linkage of CON1 and CON2 genes to the salivary protein gene complex (SPC). *Biochem Genet* 22:1-19.
- Beeley, J. A. (2001). Basic proline-rich proteins: multifunctional defence molecules? *Oral Dis*, 7(2), 69–70. <http://doi.org/10.1034/j.1601-0825.2001.0070201.x>
- Bennick A. (1982). Salivary proline-rich proteins. *Mol. Cell. Biochem.* 45, 83–99.
- Bennick A, Cannon M, Madapallimattam G. (1981). Factors affecting the adsorption of salivary acidic proline-rich proteins to hydroxyapatite. *Caries Res.* 15(1):9-20.
- Boze, H., Marlin, T., Durand, D., Pérez, J., Vernhet, A., Canon, F., Sarni-Manchado P., Cheynier V., and Cabane, B. (2010). Proline-rich salivary proteins have extended conformations. *Biophysical Journal*, 99(2), 656–665 .
<http://doi.org/10.1016/j.bpj.2010.04.050>
- Brunati, A. M., Marin, O., Bisinella, A., Salviati, A., & Pinna, L. A. (2000). Novel consensus sequence for the Golgi apparatus casein kinase, revealed using proline-rich protein-1 (PRP1)-derived peptide substrates. *The Biochemical Journal*, 351 Pt 3, 765–8. <http://doi.org/10.1042/0264-6021:3510765>
- Burgener, A., Mogk, K., Westmacott, G., Plummer, F., Ball, B., Broliden, K., & Hasselrot, K. (2012). Salivary basic proline-rich proteins are elevated in HIV-exposed seronegative men who have sex with men. *AIDS (London, England)*, 26(15), 1857–67. <http://doi.org/10.1097/QAD.0b013e328357f79c>
- Cabras, T., Melis, M., Castagnola, M., Padiglia, A., Tepper, B. J., Messina, I., & Barbarossa, I. T. (2012). Responsiveness to 6-n-propylthiouracil (PROP) is associated with salivary levels of two specific basic proline-rich proteins in humans. *PLoS ONE*, 7(2). <http://doi.org/10.1371/journal.pone.0030962>

- Cáceres, J. F., & Kornblihtt, A. R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends in Genetics : TIG*, 18(4), 186–193. [http://doi.org/10.1016/S0168-9525\(01\)02626-9](http://doi.org/10.1016/S0168-9525(01)02626-9)
- Canon, F., Paté, F., Cheynier, V., Sarni-Manchado, P., Giuliani, A., Pérez, J., Durand, D., Li J., and Cabane, B. (2013). Aggregation of the salivary proline-rich protein IB5 in the presence of the tannin EgCG. *Langmuir*, 29(6), 1926–1937. <http://doi.org/10.1021/la3041715>
- Chan, M., & Bennick, A. (2001). Proteolytic processing of a human salivary proline-rich protein precursor. *European Journal of Biochemistry*, 268(12), 3423–3431. Retrieved from <http://www.blackwell-synergy.com/links/doi/10.1046/j.1432-1327.2001.02241.x/full/>
- Cramer, P., Pesce, C. G., Baralle, F. E., & Kornblihtt, a R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 94(21), 11456–11460. <http://doi.org/10.1073/pnas.94.21.11456>
- Debray, H., Decout, D., Strecker, G., Spik, G., & Montreuil, J. (1981). Specificity of twelve lectins towards oligosaccharides and glycopeptides related to N-glycosylproteins. *European Journal of Biochemistry / FEBS*, 117(1), 41–55. <http://doi.org/10.1111/j.1432-1033.1981.tb06300.x>
- Dhople, A. M., & Hanks, J. H. (1977). In vitro growth of Mycobacterium lepraemurium, an obligate intracellular microbe . *Science* , 197 (0036-8075 (Print)), 379–381.
- Faculty, N. (1980). The Amino Acid and Sequence of a Salivary Its Relation to a Salivary Protein Proline-Rich The sequence of proline-rich glycoproteins , proteins , and pep tides in human salivas have been studied by several investigators However , the relationships among t, 87(4), 1071–1077.
- Halgand F., Zabrouskov V., Bassilian S., Souda P. et al. (2012). Defining intact protein primary structures from saliva: A step toward the human proteome project. *Anal Chem*. 2012 May 15; 84 (10): 4383-4395. doi: 10.1021/ ac203337s.
- Friedman, R. D., & Merritt, A. D. (1975). Partial Purification and Characterization of a Polymorphic Protein (Pa) in Human Parotid Saliva, 304–314.
- Hardt, M., Thomas, L. R., Dixon, S. E., Newport, G., Agabian, N., Prakobphol, A. et al. (2005). Toward defining the human parotid gland salivary proteome and peptidome: Identification and characterization using 2D SDS-PAGE, ultrafiltration,

HPLC, and mass spectrometry. *Biochemistry*, 44(8), 2885–2899.

<http://doi.org/10.1021/bi048176r>

- Hart, E. J., Schmidt, K.H., Vasudevan, K. N. (1973). Genetic Polymorphism of Proline-Rich Human Salivary Proteins.1067–1069.
- Hatton, M. N., Loomis, R. E., Levine, M. J., & Tabak, L. a. (1985). Masticatory lubrication. The role of carbohydrate in the lubricating property of a salivary glycoprotein-albumin complex. *The Biochemical Journal*, 230(3), 817–820.
- Hattori, M. (2005). Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso. Protein, Nucleic Acid, Enzyme*, 50(2), 162–168. <http://doi.org/10.1038/nature03001>
- Helmerhorst, E. J., Sun, X., Salih, E., & Oppenheim, F. G. (2008). Identification of Lys-Pro-Gln as a novel cleavage site specificity of saliva-associated proteases. *Journal of Biological Chemistry*, 283(29), 19957–19966. <http://doi.org/10.1074/jbc.M708282200>
- Houdayer, C., Dehainault, C., Mattler, C., Michaux, D., Caux-Moncoutier, V. et al.(2008). Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Human Mutation*, 29(7), 975–982. <http://doi.org/10.1002/humu.20765>
- Isemura, S., & Saitoh, E. (1997). Nucleotide sequence of gene PBI encoding a protein homologous to salivary proline-rich protein P-B, *J21*(6), 1025–1030.
- Isemura, S., Saitoh, E., & Sanada, K. (1982). Fractionation and Characterization of Basic Proline-Rich Peptides of Human Parotid Saliva and the Amino Acid Sequence of Proline-Rich Peptide P-E, *2075*, 2067–2075.
- Juusola, J., & Ballantyne, J. (2003). Messenger RNA profiling: A prototype method to supplant conventional methods for body fluid identification. *Forensic Science International*, 135(2), 85–96. [http://doi.org/10.1016/S0379-0738\(03\)00197-X](http://doi.org/10.1016/S0379-0738(03)00197-X)
- Karn RC, Goodman PA, and Yu PL (1985) Description of a genetic polymorphism of a human proline-rich salivary protein, Pc, and its relationship to other proteins in the salivary protein complex (SPC). *Biochem. Genet.* 23:37-51.
- Kauffman, D. L., Bennick, a, Blum, M., & Keller, P. J. (1991). Basic proline-rich proteins from human parotid saliva: relationships of the covalent structures of ten proteins from a single individual. *Biochemistry*, 30(14), 3351–6. <http://doi.org/10.1021/bi00228a001>

- Kauffman, D. L., & Keller, P. J. (1979). The basic proline-rich proteins in human parotid saliva from a single subject. *Archives of Oral Biology*, 24(4), 249–256. [http://doi.org/10.1016/0003-9969\(79\)90085-2](http://doi.org/10.1016/0003-9969(79)90085-2)
- Kauffman, D. L., Keller, P. J., Bennick, A., & Blum, M. (1993). Alignment of Amino Acid and DNA Sequences of Human Proline-rich Proteins. *Critical Reviews in Oral Biology & Medicine*, 4(3/4), 287–292. <http://doi.org/10.1177/10454411930040030501>
- Kauffman, D., Wong, R., Bennick, a, & Keller, P. (1982). Basic proline-rich proteins from human parotid saliva: complete covalent structure of protein IB-9 and partial structure of protein IB-6, members of a polymorphic pair. *Biochemistry*, 21(25), 6558–62. <http://doi.org/Doi 10.1021/Bi00268a036>
- Kim, H. S., Lyons KM et al. (1992). The structure and evolution of the human salivary proline-rich protein gene family. *Mammalian Genome* 4: 3-14.
- Kim, H. S., & Maeda, N. (1986). Structures of two HaeIII-type genes in the human salivary proline-rich protein multigene family. *Journal of Biological Chemistry*, 261(15), 6712–6718.
- Kim, H.-S., Smithies, O., & Maeda, N. (1990). A physical map of the human salivary proline-rich protein gene cluster covers aver 700kbp of DNA. *Genomics*, 6(1984), 260–267.
- Kupfer, D. M., White, V. L., Jenkins, M. C., & Burian, D. (2010). Examining smoking-induced differential gene expression changes in buccal mucosa. *BMC Medical Genomics*, 3, 24. <http://doi.org/10.1186/1755-8794-3-24>
- Lareau, L. F., & Brenner, S. E. (2015). Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Molecular Biology and Evolution*, 32(4), 1072–1079. <http://doi.org/10.1093/molbev/msv002>
- Levine, M. J., & Ellison, S. A. (1973). THE ISOLATION FROM HUMAN PAROTID SALIVA AND PARTIAL CHARACTERIZATION OF THE PROTEIN CORE OF A MAJOR Departments of Oral Biology and Biochemistry , State University, 18, 827–837.
- Li, Y., John, M. A. R. S., Zhou, X., Kim, Y., Sinha, U., Jordan, R. C. K. et al. (2004). Salivary Transcriptome Diagnostics for Oral Cancer Detection Salivary Transcriptome Diagnostics for Oral Cancer Detection. *Clin Cancer Res*, 10(310), 8442–8450.

- Lu, Y., & Bennick, A. (1998). Interaction of tannin with human salivary proline-rich proteins. *Archives of Oral Biology*, 43(9), 717–728. [http://doi.org/10.1016/S0003-9969\(98\)00040-5](http://doi.org/10.1016/S0003-9969(98)00040-5)
- Lyons, K. M., Azen, E. A., Goodman, P. A., & Smithies, O. (1988). Many protein products from a few loci: Assignment of human salivary proline-rich proteins to specific loci. *Genetics*, 120(1), 255–265.
- Lyons, K. M., Stein, J. H., & Smithies, O. (1988). Length polymorphisms in human proline-rich protein genes generated by intragenic unequal crossing over. *Genetics*, 120(1), 267–278.
- Maeda, N., Kim, H. S., Azen, E. A., & Smithies, O. (1985). Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system. *Journal of Biological Chemistry*, 260(20), 11123–11130.
- Manconi, B., Castagnola, M., Cabras, T., Olianias, A., Vitali, A., Desiderio, C., Sanna, M.T., Messana, I. (2015). The intriguing heterogeneity of human salivary proline-rich proteins. Short title: Salivary proline-rich protein species. *Journal of Proteomics*. <http://doi.org/10.1016/j.jprot.2015.09.009>
- Messana, I., Cabras, T., Pisano, E., Sanna, M. T., Olianias, A., Manconi, B., Pellegrini M. et al. (2008). Trafficking and postsecretory events responsible for the formation of secreted human salivary peptides: a proteomics approach. *Molecular & Cellular Proteomics : MCP*, 7(5), 911–26. <http://doi.org/10.1074/mcp.M700501-MCP200>
- Minaguchi, K., Shirotani, T., & Suzuki, K. (1988). New variants of Ps salivary polymorphic proteins, 11–16.
- Minaguchi, K., Suzuki, K., & Pierce, L. E. E. S. (1981). FREQUENCIES OF SALIVARY GENETIC MARKER SYSTEMS IN CAUCASIANS WITH AN EMPHASIS ON Pm AND Ph . *Forensic Science International*, 17, 5–7.
- Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Research*, 10(2), 459–72. <http://doi.org/10.1093/nar/10.2.459>
- O’Sullivan, J. M., Cannon, R. D., Sullivan, P. A., & Jenkinson, H. F. (1997). Identification of salivary basic proline-rich proteins as receptors for *Candida albicans* adhesion. *Microbiology*, 143(2), 341–348.
- Oppenheim, F. G., Salih, E., Siqueira, W. L., Zhang, W., & Helmerhorst, E. J. (2007). Salivary proteome and its genetic polymorphisms. *Annals of the New York Academy of Sciences*, 1098, 22–50. <http://doi.org/10.1196/annals.1384.030>

- Polymorphisms, P. R. B. C. (1996). Coded Polymorphisms Concanavalin-A Binding, 11-1, 143–153.
- Robinovitch, M. R., Ashley, R. L., Iversen, J. M., Vigoren, E. M., Oppenheim, F. G., & Lamkin, M. (2001). Parotid salivary basic proline-rich proteins inhibit HIV-1 infectivity. *Oral Dis*, 7(2), 86–93. <http://doi.org/10.1034/j.1601-0825.2001.0070204.x>
- Saitoh, E., Isemura, S., & Sanada, K. (1999). Further Fractionation of Basic Proline-Rich Peptides from Human Parotid Saliva and Complete Amino Acid Sequence of Basic Proline-Rich Peptide P-H 1, 1999(1983), 1991–1999.
- Saliva, P. (1983). Complete Amino Acid Sequence of a Basic Proline-Rich, 888, 883–888.
- Santa, T., Al-Dirbashi, O. Y., & Fukushima, T. (2007). Derivatization reagents in liquid chromatography/electrospray ionization tandem mass spectrometry for biomedical analysis. *Drug Discov Ther*, 1(2), 108–118. <http://doi.org/10.1002/bmc>
- Scherer, S. E., Muzny, D. M., Buhay, C. J., Chen, R., Cree, A., Ding, Y. et al.(2006). The finished DNA sequence of human chromosome 12. *Nature*, 440(7082), 346–351. <http://doi.org/10.1038/nature04569>
- Shintani M, Minaguchi K, Lim KA, Hashimoto M, Suzuki K. Salivary proline-rich protein polymorphisms in Chinese, Malays and Indians in Singapore. *Hum Hered.* 1990;40(2):89-98.
- Scully, C., Bagan, J. V, Hopper, C., & Epstein, J. B. (2008). Oral cancer: Current and future diagnostic techniques. *American Journal of Dentistry*, 21(4), 199–209. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-52049115657&partnerID=40&md5=de536d78c2d133647266d4b30b0407c3>
- Siqueira, W.L. and Oppenheim, F.G. (2009). Small Molecular Weight Proteins/Peptides Present in the in vivo formed Human Acquired Enamel Pellicle. *Arch Oral Biol.*54(5):437-444. doi:10.1016/j.archoralbio.2009.01.011.
- Siqueira, W.L., Salih, E., Wan, D.L. et al. (2008). Proteome of Human Minor Salivary Gland Secretion. *J Dent Res.* 87(5): 445-450.
- Smith, C. W. J., & Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in Biochemical Sciences*, 25(8), 381–388. [http://doi.org/10.1016/S0968-0004\(00\)01604-2](http://doi.org/10.1016/S0968-0004(00)01604-2)
- Stubbs, M., Chan, J., Kwan, A., So, J., Barchynsky, U., Rassouli-Rahsti, M. et al. (1998). Encoding of human basic and glycosylated proline-rich proteins by the PRB

gene complex and proteolytic processing of their precursor proteins. *Archives of Oral Biology*, 43(10), 753–770. [http://doi.org/10.1016/S0003-9969\(98\)00068-5](http://doi.org/10.1016/S0003-9969(98)00068-5)

- Tagliabracci, V. S., Wiley, S. E., Guo, X., Kinch, L. N., Durrant, E., Wen, J. et al. (2015). A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell*, 161(7), 1619–1632. <http://doi.org/10.1016/j.cell.2015.05.028>
- Vitorino, R., Alves, R., Barros, A., Caseiro, A., Ferreira, R., Lobo, M. C. et al. (2010). Finding new posttranslational modifications in salivary proline-rich proteins. *Proteomics*, 10(20), 3732–3742. <http://doi.org/10.1002/pmic.201000261>

ELENCO DELLE PUBBLICAZIONI

- Mirko Zaffagnini, Simona Fermani, Matteo Calvaresi, Roberto Orrù, Luisa Iommarini, Francesca Sparla, Giuseppe Falini, Andrea Bottoni, Paolo Trost (2015). Tuning Cysteine Reactivity and Sulfenic Acid Stability by Protein Microenvironment in Glyceraldehyde-3-Phosphate Dehydrogenases of *Arabidopsis Thaliana*. *Antioxidants & Redox Signaling*, December 2015; DOI: 10.1089/ars.2015.6417
- Roberto Orrù, Elena Atzori, Alessandra Padiglia (2015). Development of a molecular method for the rapid screening and identification of the three functionally relevant polymorphisms in the human TAS2R38 receptor gene in studies of sensitivity to the bitter taste of PROP. *SpringerPlus* 4(1):246, December 2015; DOI: 10.1186/s40064-015-1045-0
- Roberto Orrù, Paolo Zucca, Matteo Falzoi, Elena Atzori, Antonio Rescigno, Alessandra Padiglia. First step towards the biomolecular characterization of *Pompia*, an endemic Citrus-like fruit from Sardinia (Italy). *Plant Biosystems*, manuscript accepted January 2016

RINGRAZIAMENTI

Notoriamente non sono molto portato per le smancerie, quindi sarò breve, ma chi legge sa che quanto scrivo viene dall'anima. Voglio ringraziare tutte le persone che mi hanno supportato e, soprattutto, sopportato nell'arco di questi tre anni di dottorato, a partire dalla mia famiglia. Un ringraziamento speciale va alla mia tutor, la Professoressa Alessandra Padiglia, dalla quale ho imparato tanto dal punto di vista professionale e, soprattutto, umano. Grazie anche a tutti i colleghi e tirocinanti, che mi sento ora di poter chiamare amici dopo aver condiviso con me spazio e tempo, gioie e frustrazioni, sia fuori che dentro i laboratori; in ordine sparso, i Dottori e le Dottoresse: Elisa Vacca, Monica Sanna, Elena Atzori, Alfredo D'Alessandro, Valentina Piras, Alessia Mascia. Ringrazio anche i Professori Massimo Castagnola e Irene Messana (i vostri incoraggiamenti sono stati davvero gratificanti), la Prof.ssa Antonella Fais, la Prof.ssa Alessandra Olianias, la Prof.ssa Rosaria Medda e le Dottoresse Francesca Pintus e Benedetta Era.