

ABSTRACT

BADHWAR, ANSHUL KUMAR. Dysregulation of the PR-3 Gene Locus in ANCA Disease: Consequences and Mechanisms. (Under the direction of Dr. Ronald Falk).

Antineutrophil cytoplasmic autoantibodies (ANCA) are commonly found in patients with an autoimmune necrotizing glomerulonephritis. Due to a defect in epigenetic silencing, circulating neutrophils from patients with ANCA disease express proteinase 3 (PR-3) and myeloperoxidase (MPO) genes which are normally expressed only in bone marrow cells. Gene expression analyses by microarray showed that patients had elevated levels of SR protein kinase 1, a master regulator of constitutive splicing reactions. We examined the processing of PR-3 transcripts in mature leukocytes of patients with PR3-ANCA and healthy individuals. Unexpectedly, naturally encoded PR-3 transcripts were also detected a subset of healthy controls, which suggests that PR-3 may have a unique function in periphery. Aberrant isoforms of PR-3 transcripts were detected in patients and not in healthy controls. We hypothesize that the formation of autoantibodies in patients may result as a loss of tolerance to these altered proteins.

Dysregulation of the PR-3 Gene Locus in ANCA Disease: Consequences and Mechanisms.

by
Anshul Badhwar

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Immunology

Raleigh, North Carolina

2011

APPROVED BY:

Dr. Sue Tonkonogy
Committee Chair

Dr. Stuart Maxwell

Dr. Ronald Falk

BIOGRAPHY

Anshul Badhwar grew up in Little Rock, Arkansas before moving to Chapel Hill, NC for his undergraduate training at the University of North Carolina at Chapel Hill. He completed his Bachelor of Science in Biology at UNC-CH after which he immediately joined the UNC School of Medicine – Nephrology. His time here was guided by Dr. Gloria Preston and Dr. Ronald Falk. In his second year conducting research at UNC, he was admitted to North Carolina State University in the Immunology Program. At UNC he has studied both epigenetics and RNA processing in Anti-neutrophil mediated cytoplasmic (ANCA) vasculitis. Upon completion of his Master's degree, he will finish up his current research at UNC-CH.

ACKNOWLEDGEMENTS

I would like to thank my lab parents Dr. Gloria Preston and Dr. Ronald Falk for all the help they have given me over the years. Gloria's passion for science naturally spread to me and has evoked a life-long connection to biology within me. I have tried my best to be a sponge and absorb what is only a fraction of the years of her experience. Dr. Falk has truly helped me to see the bigger picture as a scientist, and has kept me on track every time I have tried to veer. I know that he will always be a mentor to me; there are very few people whose advice I take as seriously as his. I hope one day I can emulate his nature as a scientist.

I would like to thank Dr. Stuart Maxwell, for I still refer to his notes from his biochemistry class, I can attribute my strong fundamentals in molecular biology to his teachings. I would also like to thank Dr. Sue Tonkonogy for taking me under her wing, to me she epitomizes the objective scientist that focuses on the broader ultimate goals, which I strive to become. I appreciate Dr. Scott Laster for giving me the immunology tool set I needed to better understand my own research.

Of course, can't forget my parents and sister, who have stayed with me through all my struggles and joys and supporting me through my development as an adult.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS.....	ix
1. INTRODUCTION.....	1
1.1. Immunological tolerance and autoimmunity	1
1.2. Neutrophil biology.....	2
1.3. Antineutrophil Cytoplasmic Autoantibody mediated-vasculitis	3
1.4. The autoantigen PR-3	5
1.5. Epigenetic dysregulation in ANCA disease	6
1.6. ANCA disease and RNA processing	7
1.7. Detection of sense versus antisense.....	10
1.8. Central hypothesis	13
2. MATERIALS AND METHODS	15
2.1. RNA Isolation from Total Leukocytes	15
2.2. RNA Quality Control.....	15
2.3. Reverse Transcription.....	16
2.4. Poly(A) RNA Purification.	16
2.5. Primer Design.	17
2.6. Tagged Strand-Specific RT-PCR.	17
2.7. Periodate treatment of RNA	18
2.8. First strand cDNA synthesis after periodate treatment.....	18
2.9. PCR after strand-specific RT	19

2.10. Generation of non-radioactive RNA probes for northern blotting.	19
2.11. Biotinylation of in vitro transcribed RNA probes	21
2.12. Denaturing agarose gel for northern blotting	21
2.13. Northern blotting	22
2.14. RNA-ligase mediated rapid amplification of cDNA ends	23
3. RESULTS.....	26
3.1. PR-3 is actively transcribed by RNA Pol II	26
3.2. Northern blotting of PR-3 mRNA from patients reveals an aberrantly spliced form of PR-3	28
3.3. Characterization of PR-3 transcripts.....	35
3.4. Myeloblastin and PR-3 are not encoded by a single mRNA	49
3.5. Detecting an antisense transcript co-expressed with sense PR-3 mRNA.....	51
3.7. A novel antisense transcript to PR-3 is detected by ligation-anchored RT- PCR	58
4. DISCUSSION	65
5. REFERENCES	74
6. APPENDIX	81
A.1. Treatment of RNA with sodium periodate before reverse transcription eliminates spurious priming of untargeted transcripts while retaining strand- specificity.	82

LIST OF TABLES

Table 1.	Probes generated for northern blotting.	21
Table 2.	Potential applications of periodate treatment of RNA before cDNA synthesis	72
Table 3.	Periodate treatments	86

LIST OF FIGURES

Figure 1. The issues with strand-specific RT-PCR.....	12
Figure 2. RNA Polymerase II-RNA immunoprecipitation (RIP) of sense PR-3.....	27
Figure 3. Dot blot analysis of psoralen-biotin labeled probes.	29
Figure 4. Northern blotting of sense PR-3.....	31
Figure 5. Northern blotting of sense PR-3.....	32
Figure 6. Northern blotting of sense PR-3 mRNA.	33
Figure 7: Semi-quantitative comparison of PR-3 blotting and PR-3 qRT-PCR.....	34
Figure 8. Schematic of 3' RACE.....	37
Figure 9. 3'RACE amplification of sense PR-3 from PR3-ANCA patient leukocyte RNA.	38
Figure 10. 3'RACE sequencing and BLAST alignment of PR-3 mRNA.....	39
Figure 11A. Sequence obtained from TOPO clone of larger RACE band.	40
Figure 11B. Genomic sequence of PRTN3 exon 5, with crucial elements highlighted.	40
Figure 12. 3' RACE amplification of 3' ends of PR-3 transcript of total leukocyte RNA.	42
Figure 13(A). 3'RACE amplification with GC optimization with multiple nested primer sets.....	44
Figure 13(B). 3'RACE amplification with GC optimization and multiple primer sets.....	45
Figure 14. Sequencing results from P3 3'RACE PCR clone.....	46
Figure 15. Schematic of template switching.....	47

Figure 16. 3' RACE clone from P5.	48
Figure 17. Comparison of PR-3 and myeloblastin protein sequences.	50
Figure 18. Analysis of strand-specificity	53
Figure 19. Issues with strand-specific RT-PCR.....	56
Figure 20. Tagged strand-specific RT-PCR after Microcon removal of unreacted 'tag' RT primer.	57
Figure 21. Pilot LA RT-PCR experiment designed to amplify antisense PR-3.....	60
Figure 22. Sample patient screen for antisense PR-3 in PR3-ANCA patients by LA RT-PCR.....	61
Figure 23. Sequencing chromatogram results of a cloned antisense PR-3 transcript. ...	62
Figure 24. RIP-ChIP of antisense PR-3.	64
Figure 25. (A) Periodate oxidation of vicinal diol at the terminal nucleotide of RNA.	83
Figure 26. Periodate oxidation before RT blocks non-specific cDNA synthesis.	85

LIST OF ABBREVIATIONS

3' RACE:	3' Rapid amplification of cDNA ends
5' RACE:	5' Rapid amplification of cDNA ends
5' RLM-RACE:	5' RNA-ligase mediated rapid amplification of cDNA ends
ANCA:	Anti-neutrophil Cytoplasmic Autoantibodies
APC:	Antigen presenting cell
AP:	Alkaline phosphatase
ASF/SF2:	Alternative splicing factor/splicing factor 2
ANRIL:	Antisense RNA in the INK4 locus
cPR3:	Complementary PR-3
cDNA:	complementary DNA
CTD:	C-terminal domain
DCs:	Dendritic cells
DMSO:	Dimethyl sulfoxide
EED:	Embryonic ectoderm development
EZH2:	Enhancer of zeste homolog 2
H3k27me3:	Tri-Methylated Histone H3 at Lysine 27
HOX:	Homeobox
HOTAIR:	HOX antisense intergenic RNA
lncRNA:	Long non-coding RNA
MPO:	Myeloperoxidase

NAT:	Naturally antisense transcripts
ncRNA:	Non-coding RNA
P15AS:	p15 antisense transcript
PMA:	Phorbol 12-myristate 13-acetate
PR-3:	proteinase 3
<i>PRTN3</i> :	PR-3 gene
PRC2:	Polycomb repressive complex 2
PMN:	Polymorphonuclear leukocytes
qRT-PCR:	Quantitative RT-PCR
RIP:	RNA-binding protein immunoprecipitation
RNA pol II:	RNA polymerase II
RRM :	RNA recognition motif
RT:	Reverse transcription
RUNX3:	Runt-related transcription factor 3
SR:	Serine/arginine-rich proteins
SRPK-1:	Serine/arginine protein kinase 1
SSC:	Saline-sodium citrate
SUZ12:	Suppressor of zeste 12 homolog
Treg:	regulatory T cells

1. INTRODUCTION

1.1. Immunological tolerance and autoimmunity

A vital component of the immune system is the ability to protect against a wide array of pathogenic microorganisms, while avoiding immune reactions to endogenous proteins [1]. In order to accomplish this, the immune system has evolved methods to prevent spurious immune responses through mechanisms of ‘tolerance’ to self-antigens. Immune tolerance is a process whereby T and B cells that react to a self or non-self antigens are either eliminated or become unresponsive. Central tolerance is maintained by the selective deletion of immature T-cells that recognize self-antigens within the thymus. As a result, these autoreactive T-cells never migrate to the periphery [2]. Similarly, immature B-cells that bind self-antigens with high affinity are prevented from completing maturation within the bone marrow [3]. Although central tolerance dictates what immature B and T cells leave the primary lymphoid organs, mechanisms of peripheral tolerance regulate the activity of T and B cells that have matured and migrated to the periphery. Peripheral tolerance can be mediated by regulatory T-cells (Treg), a subset of T-cells that can render autoreactive cells hyporeactive to self antigen. B or T cells that encounter antigen in the absence of costimulatory signals, that are normally generated through inflammatory processes, can also be rendered inactive through a process called anergy [4]. An ‘anergic’ cell is essentially maintained in a state of metabolic arrest and when challenged again with the same antigen is unresponsive. Autoimmune disease is thought to arise from a failure in one of the mechanisms described above which normally maintain self-tolerance in a healthy individual. For example, in immunodysregulation

polyendocrinopathy enteropathy X-linked syndrome (IPEX) patients possess mutations in the transcription factor FoxP3, which impairs the suppressive ability of T-regs in the periphery. Although, most autoimmune diseases are not strictly due to genetic defects, the combination of both environmental factors and genetic predispositions contribute to the common theme in the breakdown of self tolerance.

1.2. Neutrophil biology

Neutrophils are the most abundant form of leukocytes found in the blood and play an integral role in the innate immune response. Neutrophils can shape an immune response by secreting cytokines that eventually attract monocytes and dendritic cells (DCs). These signals can result in a pro-inflammatory response in which antigen-presenting cells (APC) can be recruited specifically to sites of infection [5]. By producing tumor-necrosis factor (TNF) they can drive differentiation or activation of both macrophages and DCs [6]. During the later stages of neutrophil maturation, large portions of chromatin condense causing the morphological appearance of the nucleus to change to a linear array of three or four lobes joined by thin DNA-containing filaments. The change in morphology is also thought to be part of a maturation induced silencing of most genes expressed in the earlier stages of differentiation [7]. In particular, transcription of primary granule constituents is terminally silenced once a granulocyte matures and leaves the bone marrow. For example, myeloperoxidase (MPO) and proteinase 3 (PR-3) mRNA transcripts are found almost exclusively at the early promyelocyte stage. Neutrophil elastase transcripts are found in blasts and disappear as the promyelocytes further differentiate [8].

1.3. Antineutrophil Cytoplasmic Autoantibody mediated-vasculitis

Antineutrophil Cytoplasmic Autoantibodies (ANCA) are commonly found circulating in patients with autoimmune necrotizing systemic vasculitis [9]. The small vessel vasculitides such as Wegener's granulomatosis, microscopic polyangiitis, and Churg-Strauss Syndrome are all associated with the presence of ANCA [10]. Common among these diseases is the presence of necrotizing glomerulonephritis with a paucity of immunoglobulin deposition in glomeruli. The etiology of the disease remains largely unknown and is thought to be influenced by a gamut of genetic, epigenetic and environmental factors. The two primary antigens recognized by ANCA are PR-3 and MPO, which are uniquely expressed in neutrophils and monocytes [11, 12]. During vasculitis, these autoantibodies bind to their respective antigens and activate neutrophils resulting in degranulation, and release toxic constituents of the azurophilic granules, causing injury of the blood vessel wall [13].

Mechanisms explaining the initial formation of ANCA were proposed after a serendipitous observation made while epitope mapping PR-3 specific antibodies (PR3-ANCA). Pendergraft *et al* discovered that patients with PR3-ANCA also harbored antibodies against complementary PR-3(cPR3) [16]. The procedure for epitope mapping required fragmentation of cDNA which was blunt-end ligated into a bacterial expression vector for the production of small peptides. Consequently, some of the cloned sequences had portions of PR-3 DNA inserted in an inverted orientation resulting in the expression of a protein 'complementary' to the sense PR-3 protein. Complementary proteins are considered

molecular mirror images of their sense counterparts, and a have natural affinity for their corresponding partners. The ‘specific and mutual’ interactions fundamentally of these proteins require that they originate from directly opposite complementary codons [14]. This natural interaction is driven by the inverted hydropathy of the sense codon and its cognate ‘complementary’ counterpart [15]. This is thought to be an intrinsic part of the proteomic code and a basis for codon usage throughout the genome.

Our research group concluded that patients with PR3-ANCA had additional antibodies that reacted with a “complementary” peptide and it was the formation of these antibodies that were the inciting event on the path towards the eventual pathogenic autoantibodies. The theory of autoantigen complementarity was proposed as a mechanism whereby PR-3 ANCA are produced secondarily to an immune response to the initial complementary cPR3 antigen [16]. The complementary determining region (CDR) of the anti-cPR3 antibody acts as an immunogen, resulting in the creation of an anti-idiotypic antibody. This anti-idiotypic antibody consequently cross-reacts with ‘self’ PR-3 due its ‘complementary’ nature, creating classical ANCA.

Anti-MPO antibodies are proven to be pathogenic as shown by Xiao et. al in a mouse model aimed to mimic MPO-ANCA in humans [17]. In this model, the anti-MPO antibodies were first produced in a MPO knockout mouse by immunization with mouse MPO. Purified anti-MPO IgG from these mice were then used for injection in to Rag2^{-/-} mice, which lack

functional B and T cells. Mice with passively transferred antibody developed necrotizing and crescentic glomerulonephritis similar to what it seen in humans with small vessel vasculitis [17]. Autoantibodies from patients have been shown to engage the autoantigen on the neutrophil cell surface and induce a respiratory burst *in vitro* [18]. ANCA are predominantly of the IgG4 subclass in contrast to the more commonly found autoantibody subclass of IgG1 [19].

1.4. The autoantigen PR-3

Initially when PR-3 was discovered there was a debate of whether the protein was identical to the independently discovered protein myeloblastin. Investigators concluded that both proteins were identical and encoded by the same mRNA [20]. We have significant evidence suggesting myeloblastin and PR-3 are not the same protein [21]. Bories et. al showed that specific knock down of myeloblastin mRNA with antisense oligonucleotides caused terminal differentiation of a promyelocytic cell line [22]. These investigators also showed that treatments with PMA or DMSO induced downregulation of myeloblastin mRNA expression and also induced terminal differentiation [22]. Because PR-3 and myeloblastin have been considered the same protein for almost twenty years, it will be difficult to ascertain which activities of each protein are unique or overlapping between the two proteins.

The *PRTN3* gene locus on chromosome 19 contains the coding sequence for the PR-3 mRNA and is clustered with two additional granule protein genes, azurocidin and elastase [23]. The three genes are thought to be coordinately regulated and expressed at similar stages

of differentiation. PR-3 mRNA is expressed at high levels during the myeloblast and promyelocyte stages and its transcription is eventually shut off as it matures in to a bi-lobed polymorphonuclear cell [24]. The currently annotated form of PR-3 mRNA transcript has a 23-nt 5' untranslated region and a 206-nt 3' untranslated region which also contains a consensus AAUAAA sequence for polyadenylation. The transcript contains 5 exons, for a total of 1001-nt which is spliced from a 7.19 kb pre-mRNA derived from the PRTN3 gene [25].

PR-3 protein is a serine protease with antimicrobial properties that is localized to the azurophilic granules of neutrophils which can migrate to the cell surface on an activated neutrophil [26]. PR-3 protein was originally cloned and sequenced from human leukemia cell line HL-60 [22]. PR-3 is normally expressed on the surface of neutrophils of healthy individuals, however it has been found that neutrophils from patients with PR3-ANCA have increased surface PR-3 expression [27]. At sites of neutrophil degranulation, PR-3 degrades extracellular matrix proteins, and when given to hamsters by tracheal insufflations causes emphysema [28].

1.5. Epigenetic dysregulation in ANCA disease

Expression of *PR-3* and *MPO* genes is elevated in neutrophils of patients with PR3-ANCA disease, in contrast to their normally silenced state in the mature neutrophils of healthy donors [24, 29]. Ciavatta et. al showed that trimethylation of histone H3 at lysine 27

(H3K27me3) was depleted at the *PRTN3* locus, when compared to healthy controls [30]. The H3K27me3 mark is associated with transcriptionally silent chromatin and is required for the continued maintenance of this repressive state. Without the appropriate epigenetic marks, expression of PR-3 and MPO was found to be de-repressed in patients' neutrophils, and the current hypothesis is that this could result in greater antigen availability and drive disease. Decreased levels of H3K27me3, at the *PRTN3* locus, was concomitant with an increase in levels of *JMJD3*, a H3K27 trimethylation specific demethylase. In healthy individuals, PRC2 (core complex: *EED*, *EZH2*, *SUZ12*, *RbAp46*, *RbAp46*) is a major methyltransferase responsible for maintenance of H3K27 methylation [31]. It was proposed that the Runt-related transcription factor 3 (*RUNX3*) was responsible for binding Enhancer of zeste homolog 2 (*EZH2*) in order to recruit the entire PRC2 complex. However, patients with PR3-ANCA also had low levels of *RUNX3* transcript. This suggested the mechanism for lack of recruitment of PRC2 to *PRTN3* loci in patients and account for the loss of H3K27me3 silencing marks. The resultant dysregulation of PRC2 activity would eventually facilitate the formation of a relaxed chromatin structure suitable for transcription [30].

1.6. ANCA disease and RNA processing

Transcriptome-wide gene expression studies by Affymetrix microarrays on RNA isolated from leukocytes of patients with ANCA vasculitis vs. healthy donors showed profound differences in the levels of many transcripts [32]. One of the many genes differentially expressed was the Serine/arginine-rich (SR) protein kinase 1 (*SRPK1*) which was overexpressed an average of 2.7-fold higher in total leukocytes of patients with ANCA

disease compared to healthy controls. SR proteins are rich in serine and arginine and are known to control splice site dictation in mRNA [33]. Alternative splicing is also facilitated by sequence-specific binding of SR proteins to the mRNA directly as well as snRNPs, which in turn dictate splice site usage. SRPK1 can phosphorylate up to 12 different sites of alternative splicing factor/splicing factor 2 (ASF/SF2) in a directional and processive manner by moving from the C-terminus to the N-terminus [34]. Phosphorylation of ASF/SF2 causes it to translocate from the cytoplasm to the nucleus where it alters constitutive splicing reactions [35]. The various permutations in the phosphorylation of its SR domains alter its affinity for protein partners, which in turn dictates the activity of ASF/SF2 [36]. ASF/SF2 is classified as a SR protein and is required for all pre-mRNA splicing reactions [37]. A delicate balance between phosphorylation and dephosphorylation is necessary for proper pre-mRNA splicing to occur. SRPK-1 is overexpressed in many diseases and thus we speculate it may play a role in processing PR-3 mRNA during ANCA disease.

Antisense transcription is now widely documented in humans and appears to be prevalent across the entire genome [38]. Naturally occurring antisense transcripts (NATs) have been implicated in an array of regulatory roles by altering splicing, translation, epigenetic marks, transport, mRNA processing through direct protein binding and RNA base-pairing interactions [39]. In eukaryotes, a cis-encoded transcript is transcribed from the opposite strand of its sense counterpart, comprised of regions which overlap with the sense

transcript [40]. A trans-encoded NAT is defined as a transcript with a complementary region of overlap to a gene expressed from a different genomic locus.

Microarrays and more importantly RNA-seq data indicate that a significant proportion of the transcriptome contains RNAs with limited to no coding potential [41]. These non-coding RNAs (ncRNAs) are frequently antisense to their coding counterparts and play roles in higher order gene regulation [42]. A separate subclass of ‘long’ (larger than ~200 nucleotides) non-coding RNAs have been attributed to a wide array of functionalities with variability in processing from molecule to molecule [43]. Most of the transcripts are polyadenylated and spliced, however unspliced and non-polyadenylated transcripts also exist [44]. It is also not a strict requirement that these transcripts remain in the nucleus, as some long ncRNA have regulatory roles within the cytoplasm. Many long ncRNAs are only expressed at specific times of development and have the capacity to bind specific transcription factors [45]. Long ncRNAs can also act to recruit chromatin remodeling complexes to highly specific loci within the genome. For example, in the human homeobox gene locus (HOX) two lncRNAs have been discovered that act spatially to the genes they are responsible for regulating. Specifically, *HOTAIR* recruits PRC2 to the HOXC locus, resulting in the silencing of transcription across the entire 40-kb locus [46].

Numerous publications have implicated the dysregulation or aberrant production of certain lncRNAs as a fundamental component of overall disease etiology. For example, a lncRNA antisense to *p15* tumor suppressor gene altered local heterochromatin structure and

DNA methylation status of the *p15* gene locus [47]. In many leukemias the expression profiles of the *p15* and *p15* antisense (*p15AS*) transcript are inversely correlated, indicating that this mechanism is responsible for maintaining low levels of *p15* protein in cancer [47]. It was later discovered that *p15AS* represented an unspliced form of antisense noncoding RNA in the INK4 locus (*ANRIL*). More recently, an aberrant spliced circular form (*cANRIL*) was associated with atherosclerotic risk [48]. *ANRIL* had previously been shown to regulate the expression of the INK4b/ARF/INK4a tumor suppressor locus by controlling the methylation status of histone H3 at lysine 27 (H3K27) [49]. These aberrantly produced lncRNAs may be targets for therapy and as a tool to monitor disease activity.

1.7. Detection of sense versus antisense

As our understanding of the transcriptional landscape evolved, it became evident that many genes are comprised of regions with overlapping sense and antisense transcriptional units [50]. Accurate detection and discovery of antisense transcripts by reverse transcriptase (RT)-PCR has proven to be difficult due to spurious priming of the sense RNA in cDNA synthesis reactions. Numerous molecular techniques used to study RNA frequently require a reverse transcription step in order to be suitable for downstream analyses. Since its inception, RT-PCR revolutionized our ability to detect specific transcripts and paved the way for the many manipulations of RNA. Reverse transcription has two requirements, 1) a localized double stranded structure formed by complementary base pairing between the RNA itself, a short DNA or RNA oligo and 2) a free 3'OH provided by the primer to initiate polymerization of the cDNA. In RNA preparations containing sense and antisense transcripts both were

efficiently primed in the RT step and indistinguishable once amplified by PCR [51, 52]. Once cDNA synthesis has completed strand-specificity has been lost in a traditional PCR reaction. Many modifications to RT-PCR have been proposed in order to retain strand information after PCR, however none has been considered consistently reliable.

In an *in vitro* setting, a majority of the cDNA produced by reverse transcriptase is primed from endogenous RNA templates rather than the user derived DNA oligo primers. Regions of self-complementarity in the RNA molecule form hairpin loops that facilitate the formation of priming sites for the RT. The free 3'OH formed by the hairpin structure is available for initiating cDNA synthesis [52, 53]. The short RNAs (microRNAs, degraded RNA, tRNAs) can also act as 'endogenous' primers, all of which undesirably increase the complexity of a targeted gene-specific cDNA synthesis reaction (**Figure 1**). The ability to differentiate between the sense and antisense strands is confounded by these 'falsely-primed' cDNAs. Tuiskunen et. al demonstrated that 50 – 80% of cDNAs tested were produced by endogenous template-mediated priming and not the gene-specific primer during reverse transcription [54]. Moreover, it has also been shown that the highly abundant RNAs encoding β -actin and 18S rRNA are both efficiently self-primed without exogenous primer in a standard RT reaction. These unwanted cDNAs are potentially amplified with PCR protocols using only one gene-specific primer and a universal adapter primer such as 5' and 3' RACE.

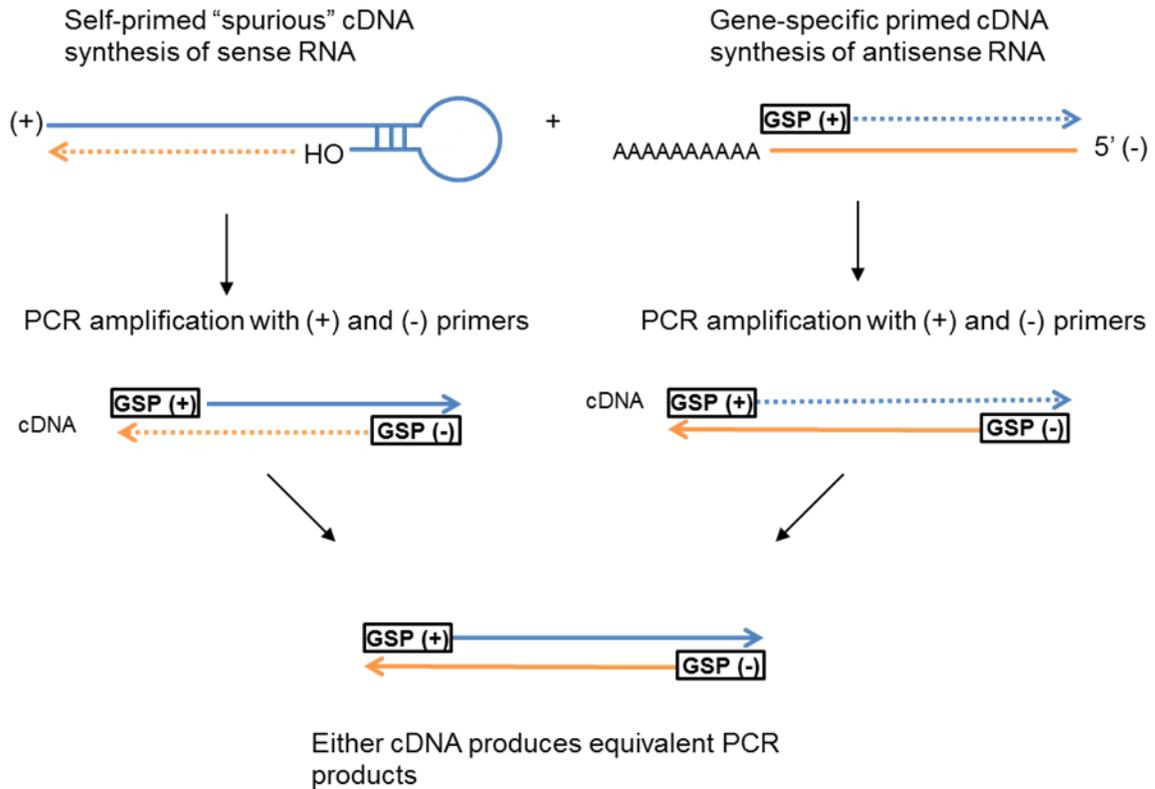


Figure 1. The issues with strand-specific RT-PCR. A transcript from either orientation produces transcripts with overlapping regions of complementarity. A localized dsRNA structure with a free 3'OH can spuriously initiate cDNA synthesis at the free 3'OH. Gene-specific cDNA synthesis is primed from the DNA oligo in the RT reaction. PCR primers for dsDNA do not differentiate between products produced from either mechanism of cDNA synthesis. Thus, regardless of the strand targeted during the RT reaction both are eventually amplified in PCR.

1.8. Central hypothesis

Several lines of evidence indicate that neutrophil gene expression is aberrant in ANCA disease. Studies of leukocyte gene expression by microarrays demonstrate a clear difference in the transcriptomic signature of ANCA patients, when compared to either healthy donors or autoimmune disease controls [24]. More importantly, mature peripheral neutrophils from patients aberrantly express PR-3 autoantigen, a transcript whose expression is traditionally repressed in neutrophils before leaving the bone marrow [7]. Upregulation of these transcripts correlates with disease activity, and is unique to ANCA disease [24]. Our research group concluded that the increased PR-3 expression was a consequence of a loss of histone methylation silencing marks normally associated with this locus. The re-expression of autoantigens in a vastly different microenvironment as compared to the bone marrow could result in unanticipated mRNA processing events. Importantly, gene expression analyses also found upregulation of SPRK-1, a highly essential kinase whose activity controls most downstream splicing reactions through ASF/SF2 [32]. Without the appropriate *trans*-acting factors, the processing and production of the autoantigens is predicted to be abnormal. In line with our work in autoantigen complementarity, the loss of histone methylation at the PRTN3 gene locus could also be conducive to antisense transcription. Most research in ANCA disease has focused on the autoantibodies and the antigen of the disease. Collectively our recent observations implicate a critical role for changes in gene expression in the actual immunopathogenesis of disease and not merely a consequence of the autoantibodies themselves. The overall objective of this work is to characterize transcripts

from the aberrantly transcribed *PRTN3* gene and its consequences in circulating leukocytes of patients with ANCA vasculitis.

We propose that the dysregulation in the epigenetic landscape of the *PRTN3* gene locus, along with the overexpression of SRPK-1, would initiate a cascade of irregular transcription and mRNA processing events in ANCA patients. We hypothesized that transcription would result in many different isoforms, some of which when translated would produce peptide fragments that would be considered ‘foreign’ by the immune system. We also hypothesized that an ‘open’ chromatin conformation would result in the production of normally or abnormally processed antisense transcripts. In this thesis I describe the discovery of alternatively spliced transcripts that are uniquely expressed in the ANCA patient population. The work required the creation of novel technical approaches to identify and characterize abnormally expressed sense and antisense transcripts.

2. MATERIALS AND METHODS

2.1. RNA Isolation from Total Leukocytes

Whole blood was collected in 6.0 mL K₂EDTA (Beckton Dickinson) tubes and processed for RNA isolation within 1 hour of blood draw. For every 5.0 mL of blood 45 mL of Red Blood Cell Lysis Buffer (0.15M NH₄Cl, 10 mM KHCO₃, 0.1 mM EDTA) was added to a 50 mL conical to selectively lyse red blood cells for 10 minutes. After RBC lysis the 50 mL conical was spun at 1000g in a fixed rotor centrifuge for 10 minutes. The supernatant was aspirated into biohazard waste and white blood cell pellet was resuspended in 10 mL of PBS for further washing. The resuspended cell pellet was spun in a 15 mL conical for 10 minutes at 1000g. The final aspiration will remove any remaining red blood cell contamination. Once the PBS was aspirated from the leukocytes the pellet was resuspended in 1 mL of TRIzol (Invitrogen) per 5 mL of initial volume of blood. RNA was stored in TRIzol for up to 2 weeks at -80°C before final RNA isolation. RNA was isolated according to manufacturer's protocols and resuspended in Nuclease Free H₂O (Ambion).

2.2. RNA Quality Control.

Isolated RNA, 1.0 µL per sample, was analyzed on a NanoDrop ND-2000 spectrophotometer (Thermo Fisher). RNA was considered acceptable for use if its OD 260/280 > 1.85, and OD 260/230 > 1.8, to ensure RNA did not contain protein, phenolic or organic contaminants. A 500 ng aliquot of the RNA was also electrophoresed in a non-denaturing 1.2% agarose gel for visible 18S and 28S bands. The gel was stained with 1:10000 5% Ethidium Bromide

Solution (Fisher Scientific) and visualized on a Gel-Doc Transilluminator (Bio-Rad). Having distinct 18S and 28S ribosomal RNA bands indicates that the RNA is intact, which otherwise cannot be inferred from a spectrophotometric reading.

2.3. Reverse Transcription.

Thermoscript RT (Invitrogen) is a thermostable reverse transcriptase, which is useful for running reactions at high temperatures in order to disrupt self-priming generally associated with reverse transcription. 1.0 μL of 10 μM RT primer, up to 5 μg of total RNA, 2.0 μL of 10 mM dNTP was brought up to 12.0 μL with nuclease free water. This pre-reaction mix was heated to 65°C for 5 minutes and cooled immediately at 4°C. The remainder of the RT reaction master mix contained 4.0 μL 5X cDNA synthesis buffer, 1.0 μL 0.1M DTT, 1.0 μL RNAsin (20U), 1.0 μL nuclease free H₂O and 1.0 μL ThermoScript RT enzyme were added to each pre-reaction mix. cDNA synthesis was carried out at 60°C for 1 hour and stored at 4°C until PCR.

2.4. Poly(A) RNA Purification.

Poly(A)Purist™ MAG Kit (Ambion) was used for Poly(A) purification because of its ability to be scaled according to starting amount of patient total leukocyte RNA. Poly(A) RNA was isolated according to manufacturer's specifications, from a total of 50 μg of total patient RNA. RNA was bound to 5.0 μL (50 μg) of Oligo(dT) Magbeads and heated to 75°C for 5 minutes and moved to room temperature for 60 minutes. Oligo(dT) beads were briefly washed with manufacturer's wash solution and eventually eluted in 200.0 μL THE RNA

storage solution (Ambion). Eluted RNA was precipitated with 0.1 volumes 5 M NH₄ Ac, 1 μL Glycogen and 2.5 volumes ethanol and stored at -20°C until ready for use in northern blotting.

2.5. Primer Design.

PCR and reverse transcription primers were generated by inputting target sequence in the Primer3 (v. 0.4.0) online tool [55]. An effort was made to select primers with similar T_m values for compatibility in PCR. Primers for gene-specific RT were also designed in Primer3.

2.6. Tagged Strand-Specific RT-PCR.

Primers spanning the complement of PRTN3 for reverse transcription were first designed in Primer3 and a unique 16-bp tag sequence was appended to the 5' end of the gene-specific portion of the tag primer. Each RT primer was thus ~20-25nt of gene-specific sequence as well as the additional tag sequence. Added 1.0 μL of 1 μM tagged gene-specific RT primer, 1.0 μg of Total RNA, 1.0 μL 10mM dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP at pH 7.0, Promega) brought up in nuclease free H₂O to 13.0 μL. Mixture was heated to 65°C for 5 minutes and quickly place on ice for 1 minute. This step facilitates the annealing of the gene-specific RT primer and relaxation of secondary structure in the RNA molecule. 4.0 μL of 5X First Strand Buffer (Invitrogen), 1.0 μL 0.100M DTT, 1.0 μL RNasin Plus (Promega), 1.0 μL Superscript III RT (200U/μL, Invitrogen) were added to the chilled pre-reaction

RNA-primer mix. The final RT reaction volume was 20.0 μL and was incubated in an Mastercycler ep (Eppendorf) at 50°C for 1 hour, followed by heat inactivation at 85°C for 10 minutes, finally stored at 4°C indefinitely. 2.0 μL of cDNA was used from the 20.0 μL RT reaction for subsequent PCR reactions. A sample schematic of strand-specific tagged RT-PCR is shown in **Figure 18a**. PCR was carried out as described earlier with Platinum Taq (Invitrogen) and 1.0 μL of “Tag” primer as a reverse primer and a gene-specific forward primer.

2.7. Periodate treatment of RNA

5 μg (5 μL) or 1 μg (1 μL) of total RNA was treated with either 2 μL 100mM sodium periodate (*Sigma-Aldrich, St. Louis, MO, USA*) or just H₂O and 3.33 μL 300 mM pH 5.3 sodium-acetate in a total 20 μL reaction. Both reactions were incubated in the dark at room temperature for 45 minutes. The oxidation reaction was quenched with 4 μL of 50% glycerol for 10 minutes. The quenched reaction was brought up to 100 μL with nuclease free water, ethanol precipitated and resuspended in 10 μL of nuclease free H₂O. RNA was further desalted by RNA Clean & Concentrator-5 columns (*Zymo-Research*) and eluted with 10 μL of nuclease free H₂O (*Ambion, Austin, TX, USA*).

2.8. First strand cDNA synthesis after periodate treatment

2 μL of each periodate treated and untreated RNA was used for reverse transcription with Superscript III (*Invitrogen, Carlsbad, California, USA*) and primed with or without a gene-

specific RT (-).627-PR3 primer for PR-3 (5'-TCCGAAGCAGATGCCGGCCTT-3'). 2 μ L of RNA and 50pM PRTN3.627-RP primer was incubated at 65°C for 5 minutes and quickly cooled on ice to facilitate primer binding and denaturation of secondary structure before reverse transcription. The RT reaction was carried out at 50°C for 1 hour in a 20 μ L reaction volume and subsequently heat inactivated at 85°C for 15 minutes.

2.9. PCR after strand-specific RT

(-).627-PR3 and (5'-GGTGTTTCTGAACAACTACGA-3') denoted (+).336-PR3 was used to amplify a 291-nt PCR product of PR-3. 2 μ L of cDNA was directly used in a 50- μ L PCR reaction containing 2.0 U of Phusion Polymerase (*Finnyzmes*) with 10 μ L of 10X Phusion HF Buffer (*Finnyzmes*), 1.0 μ L 10 mM dNTP (*Promega*,) and 0.2 μ M each primer (*IDT DNA Technologies*). Thermal cycling was carried out on a Mastercycler ep gradient (*Eppendorf, Hauppauge, NY, USA*) with the following parameters 98°C for 2 min; 35 cycles of 98°C for 30s, 62°C for 30s, and 72°C for 30s; a final 5 minute extension at 72°C and a 4°C hold. PCR products were loaded on a 1.5% TBE (Tris-borate EDTA) agarose gel, electrophoresed and visualized by UV illumination.

2.10. Generation of non-radioactive RNA probes for northern blotting.

10 μ g of pcDNA3 vector containing the middle portion of PR-3 (nt. 336-627) was linearized by the either 50U of XhoI (*NEB, Ipswich, MA*) or HindIII (*NEB*) in a 50.0 μ L reaction with 1X NEBuffer 3 (*NEB*) at 37°C for 2 hours. We purchased a pre-linearized vector for the cyclophilin probe (*Ambion*). Restriction digests were analyzed by agarose gel electrophoresis

to confirm vector was completely linearized. We chose the viral polymerase based on the polarity of the strand targeted for northern blotting. A table of the probes generated and enzymes used to generate the particular probes are listed in **Table 1**. For in vitro transcription, 1 µg of linearized vector DNA, along with 1.0 µL of each nucleotide 10mM (ATP, CTP, GTP, UTP) (GE Healthcare), 2.0 µL of 10X transcription buffer, 2.0 µL of either T7 or Sp6 phage polymerase (Ambion) was brought up to 20 µL with nuclease free water and incubated 37°C for 1 hour. T7 polymerase was used to synthesize a probe for antisense PR-3 and Sp6 was to produce a probe for sense PR-3. The in vitro transcribed probes were treated with 1.0 µL (2U/µL) Turbo DNase I (Ambion) containing 2.0 µL 10X Turbo DNase I buffer a final reaction volume of 20.0 µL. Reaction was terminated by adding 30.0 µL of nuclease free water and 50.0 µL of Acidic phenol:chloroform:isoamyl alcohol (125:25:1) pH 4.5 (Ambion) and vigorously mixed. The phenol/chloroform extraction was transferred to a pre-spun Phase-lock heavy gel (Eppendorf) and centrifuged at 12,000 g for 5 minutes for phase separation. 50.0 µL of pure chloroform (Fisher Scientific) was added to the aqueous phase containing the RNA probe and centrifuged again at 12,000g for 5 minutes. The aqueous layer was removed and ethanol precipitated by adding 1/10th volumes of 3M CH₃COONa pH 5.3, 2.5 volumes of 100% ethanol, 1.0 µL glycogen (Roche) and finally chilled overnight at -20° C. After ethanol precipitation the RNA probe was analyzed by NanoDrop-2000 (Thermo) for purity and concentration.

2.11. Biotinylation of in vitro transcribed RNA probes

Before biotinylation, 500 ng of RNA probe was heated to 100°C for 10 minutes to denature RNA secondary structures and quickly chilled in ice water. Using the BrightStar Psoralen-Biotin Nonisotopic Labeling Kit (Ambion), 500 ng of denatured probe, 1.0 µL Psoralen-biotin dissolved in dimethyl formamide was brought up with 10 µL of nuclease free H₂O. Psoralen-biotin was cross-linked to RNA with a hand-held long wave (365 nm) ultraviolet lamp for 45 minutes. The distance between the UV light source was set at 2.0 cm. Tubes were left uncapped to ensure proper irradiation. Biotinylated RNA was diluted with 89µL of TE before butanol extraction. Excess unreacted biotin was removed by adding 100 µL of water-saturated butanol to the TE-diluted RNA sample and spun at 7000g for 1 minute. The upper butanol layer containing free biotin was removed, a second butanol extraction was performed and the probe was stored at -80°C indefinitely.

Table 1. Probes generated for northern blotting.

Probe	Size (bp)	Digest	RNA Polymerase	Probe Target
AS-PR-3-m	291	XhoI	T7	Antisense
PR-3-m	291	Hind III	Sp6	Sense
pTRI cyclophilin	200	-	Sp6	Sense

2.12. Denaturing agarose gel for northern blotting

A 1.2% denaturing agarose gel was prepared by heating 1.2 g of Seakem LE agarose (Lonza), 72 mL of dH₂O, and 20 mL of 5X MOPS. 8 mL pre-warmed formaldehyde was added to the dissolved agarose solution and finally poured in to a gel holder for solidification.

2X RNA loading buffer was prepared by mixing 168 μL deionized formamide (Fisher), 54 μL of formaldehyde (Fisher), 50 μL of 5X MOPS, and 20 μL of 1% bromophenol blue (Fisher). Before loading the samples, 15 μg of RNA was heat denatured in 1X RNA loading buffer at 65°C for 15 minutes and quickly chilled on ice. 2.0 μL BrightStar Biotinylated RNA Millennium markers were also loaded for transcript size determination. Once the samples had cooled, the samples were loaded on to the gel and electrophoresed in 1X MOPS buffer at 80V for a total of 2.5 hours. The gel was transferred overnight to Hybond N+ positively charged nylon membrane (Amersham) in 20X SSC using the TurboBlotter (Whatman) capillary transfer system. After RNA had transferred it was crosslinked to the nylon membrane by the UV Stratalinker 1500 (Stratagene) using the auto-crosslink function.

2.13. Northern blotting

The cross-linked nylon membranes were first blocked with pre-heated ULTRAhyb (Ambion) buffer at 68°C for 30 minutes. In a 15 mL hybridization volume, 1 pM of psoralen-biotin labeled probe was incubated with the blot at 68°C for a total of 16 hours. After hybridization, washes were performed at 68°C with 50 mL of 2X saline-sodium citrate (SSC), 0.1% SDS solution for two washes of five minutes each, and a high-stringency wash with 50 mL of 0.1 SSC, 0.1% SDS for two washes of 15 minutes each. The Brighstar Biodetect kit (Ambion) was used for non-isotopic detection of hybridized RNA probe. Following manufacturer's guidelines, the blot was washed 2 x 5 minutes with 1X Brightstar wash buffer, and washed again with 2 x 5 minutes of 1X Brightstar block buffer. The blot was then blocked with 1X Brightstar wash buffer for 30 minutes. After blocking, 0.7 μL of streptavidin (Ambion) was

diluted in 7 mL of 1X Brightstar block buffer and incubated at room temperature for 30 minutes. The streptavidin-AP solution was removed and the blot was incubated again in 1X Brightstar block buffer for 10 minutes. Membrane was washed again with 1X Brightstar wash buffer for 3 washes of 5 minutes each. Residual wash buffer was removed from the blot by further washes with 1X Brightstar assay buffer with two five minute washes. The membrane was removed from assay buffer and incubated with 6.0 mL of CDP-*star* (Ambion). Excess CDP-*star* was drained and wrapped in saran wrap before exposure on the Fluorochem Q imaging system (Cell Biosciences).

2.14. RNA-ligase mediated rapid amplification of cDNA ends

1 µg of total RNA from patient or healthy leukocytes was dephosphorylated with 10U of Calf Intestinal Phosphatase (NEB), 2.0 µL of NEBuffer 3 (NEB), 0.75 µL RNasin (Promega) which was brought to a final volume of 20.0 µL with nuclease free water. The reaction was incubated at 37°C for 1 hour and terminated by adding 80.0 µL H₂O, and 100.0 µL acidic phenol:chloroform:IAA pH 4.5 (125:25:1) to a pre-spun heavy Phase-Lock Gel tube (Eppendorf). The Phase-Lock Gel was centrifuged at 12,000g for 5 minutes and 100.0 µL of chloroform was added to the aqueous phase. After inverting the Phase-Lock gel tube five times, the extraction was centrifuged again at 12,000g for 5 minutes. The aqueous layer was removed from the Phase-Lock Gel and precipitated with sodium acetate and ethanol. Ethanol precipitated RNA was eluted in 10.0 µL of pre-warmed nuclease free water and kept at 65°C for a total of 10 minutes. The hot elution temperature was specifically used to denature secondary structure and increase overall recovery of RNA. 10.0 µL of dephosphorylated

RNA was decapped with 0.5 U of Tobacco Acid Pyrophosphatase (Epicentre), 0.5 μL RNasin, 2.0 μL 10X TAP buffer (Epicentre) and finally raised to a final reaction of 20.0 μL with nuclease free water. This reaction was incubated at 37°C for 1 hour, raised to 100.0 μL with nuclease free water and acidic phenol chloroform extracted, and finally ethanol precipitated. RNA was eluted in 11.0 μL of hot nuclease free water, incubated at 65°C for 10 minutes and kept on ice before ligation. 11.0 μL of decapped RNA was ligated with 3.0 μL 10X T4 RNA ligase buffer (NEB), 1.0 μL of 10 mM ATP, 0.33 μL 5'RLM RNA adapter: (5'-OH-GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUG-3'OH), 3.0 μL DMSO, 9.0 μL PEG8000 solution (40% w/v) and brought up to 29.5 μL with nuclease free water. After ligation, the solution was mixed thoroughly, 0.5 μL (10U) of T4 RNA ligase was added last, and the final reaction was incubated at 37°C for 1 hour and 16°C for 16 hours. After ligation, the reaction was diluted with nuclease free water to a final volume of 100.0 μL , it was desalted and purified by a RNA Clean & Concentrator-5 column (Zymo). This column removed excess unligated adapter by size exclusion, as well as removing all ligation reaction components. RNA was reverse transcribed with a gene-specific reverse primer, using protocols as described above. For the first nested PCR reaction 1.0 μL of 5'RACE Outer (5'RO) (synthesized by IDT), 10.0 μL of 5X GC Buffer (Finnyzymes), 1.0 μL 10mM dNTP mix (Promega), 1.0 μL gene-specific reverse primer, 1.0 μL cDNA synthesis reaction, 2.50 μL DMSO and 0.50 μL Phusion Polymerase (Finnyzymes) in a final reaction volume of 50.0 μL . Touchdown PCR was employed to minimize non-specific priming of closely related sequences [56]. For phase 1 of touchdown PCR for each cycle, DNA was denatured at 98°C for 20s, the annealing was started out at 70°C for 20s and

stepped down -1°C each successive cycle for 10 cycles to a final temperature of 60°C .

Regardless of cycle number, each extension was carried out at 72°C for 1 minute. In phase 2 of touchdown, the PCR annealing temperature was kept constant at 60°C for the remaining 25 cycles of PCR.

3. RESULTS

3.1. PR-3 is actively transcribed by RNA Pol II

Prior studies using quantitative RT-PCR demonstrated that ANCA patients expressed higher levels of PR-3 mRNA when compared with healthy controls. Gene expression assays do not account for sequestration of RNA in P-bodies and mRNA half-life and are only useful for absolute transcript quantitation. In order to determine if patients were actively transcribing PR-3 mRNA, we performed RNA-binding protein immunoprecipitation (RIP) to immunopurify the actively transcribing forms of RNA polymerase II from leukocytes of patients and healthy individuals. The carboxy terminal domain (CTD), specifically the YSPTSPS repeats, become progressively phosphorylated as Pol II proceeds through transcript elongation. RNA co-immunoprecipitated with RNA pol II represents a pool of transcripts being actively transcribed. PR-3 mRNA co-purified with the transcriptionally active form of RNA pol II using an antibody specific for the hyperphosphorylated form. Primers were designed for PR-3 that spanned intron 3 (336-627) to differentiate between RNA and genomic DNA (**Figure 2a**). As shown in **Figure 2b**, 4 of 6 patients with ANCA disease had the appropriate sized amplicon of roughly ~291 nt as detected by RT-PCR. No amplification products were seen from RNA immunoprecipitates from healthy donors. This was consistent with prior findings that patients with ANCA were expressing PR-3 and provided support that epigenetic silencing of PRTN3 is somehow altered in these patients.

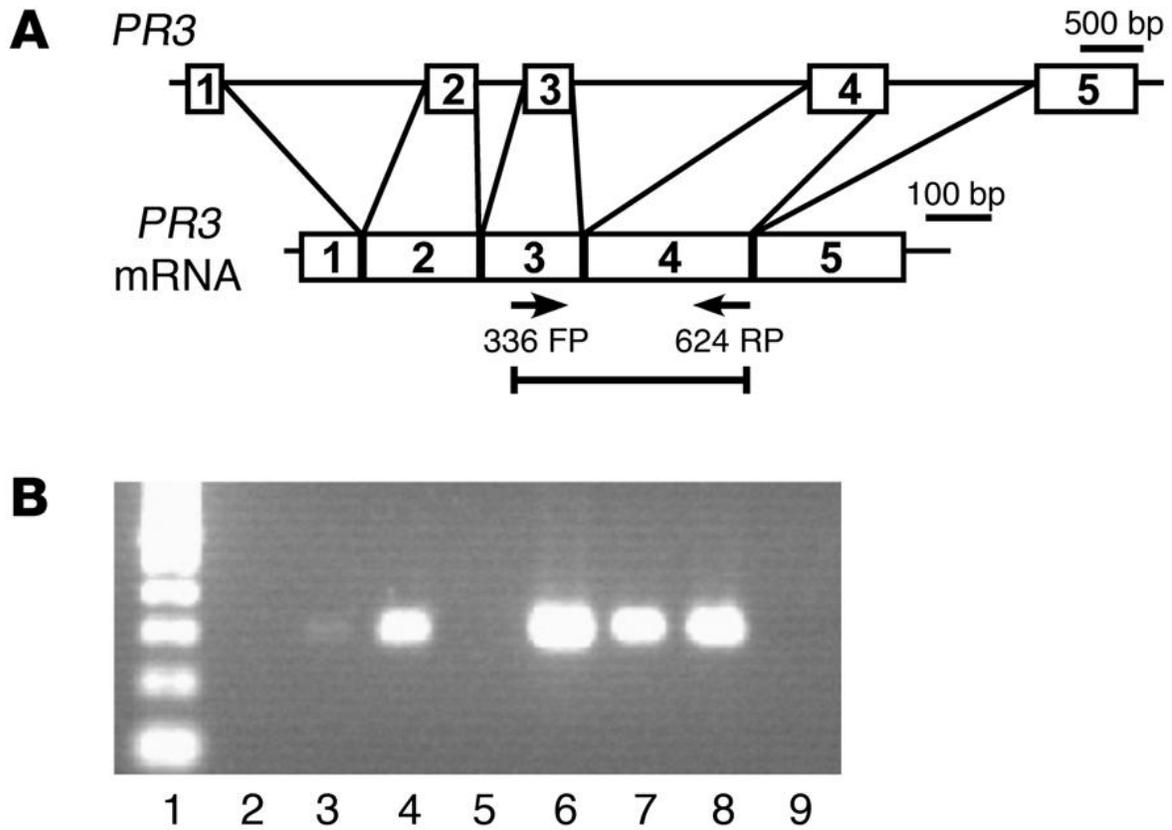


Figure 2. RNA Polymerase II-RNA immunoprecipitation (RIP) of sense PR-3. *PRTN3* gene is actively transcribed in ANCA patients. (A) Schematic of *PRTN3* gene and processed *PR-3* mRNA. Arrows mark the location of forward and reverse primers (FP and RP, respectively) used for RT-PCR analysis of RNA immunoprecipitated with anti-RNA polymerase II antibody. (B) Ethidium bromide-stained agarose gel showed RT-PCR product specific for *PR-3* mRNA present in 4 of 6 ANCA patients. Lane 1, 100-bp DNA ladder; lane 2, blank; lanes 3–8, ANCA patients; lane 9, water-only control.

3.2. Northern blotting of PR-3 mRNA from patients reveals an aberrantly spliced form of PR-3

We hypothesized that ANCA patients may express an improperly spliced form of PR-3, which in theory would contain peptides considered ‘non-self’ to leukocytes in the periphery. Five probes were designed for detection of sense PR-3 mRNA, antisense PR-3 and cyclophilin by northern blot analysis. The sense and antisense PR-3 probes were in vitro transcribed from either full length PR-3 vector (nt. 1-1001) or the middle portion (nt. 336-627). The probes were biotinylated, spotted onto a dot blot at varying dilutions of probe (1ng – 100fg) and detected by streptavidin conjugated to alkaline phosphatase (AP) using a Western Blue substrate. Probe was efficiently biotinylated similar to a standardized commercially available probe (**Figure 3**).

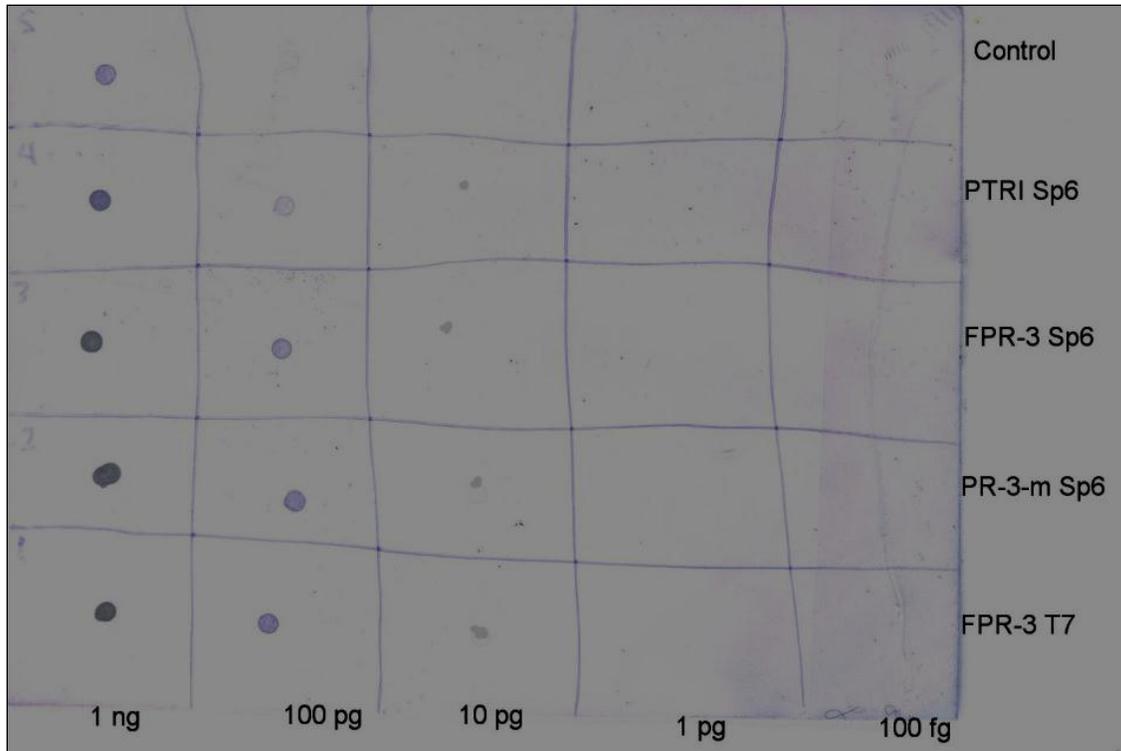


Figure 3. Dot blot analysis of psoralen-biotin labeled probes. Probes were generated for northern blotting of sense PR-3, cyclophilin and antisense PR-3. Control psoralen labeled DNA was used as a positive control and for titration of probes used for Northern blotting.

Multiple isoforms of PR-3 transcripts were observed by northern blotting of leukocyte RNA from patients with PR3-ANCA compared to healthy controls (**Figures 4-6**). The currently annotated form of PR-3 mRNA (Refseq: NM_002777) should resolve at a mobility of roughly 1.25 kb accommodating for a ~250bp poly(A) tail. Five of nine patients expressed at least one isoform of PR-3 mRNA, and of the five, three patients expressed an alternatively spliced variant larger (approx. 100 to 400 additional nucleic acids) than the currently annotated transcript (**Figure 5:Lane 1,3; Figure 5: Lane 4**). Unexpectedly, PR-3 transcripts were detected in three of nine healthy controls however all expressed the expected ~1.25 kb annotated isoform of PR-3 mRNA (**Figures 5-6**). Because whole leukocytes were collected it is impossible to determine whether PR-3 mRNA expression came from the monocytes, neutrophils or both. Levels of PR-3 transcript detected by northern blot were quantitatively similar to levels detected by standardized qRT-PCR (**Figure 7**).

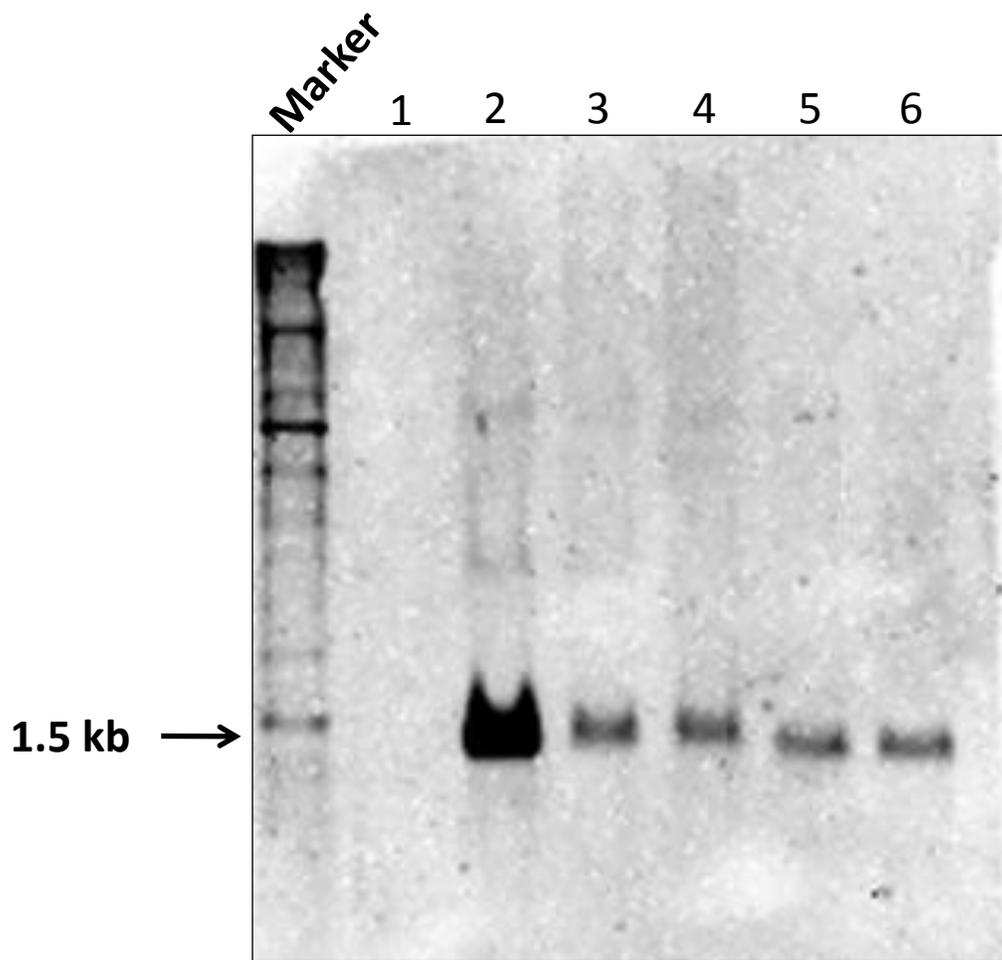


Figure 4. Northern blotting of sense PR-3. Sense PR-3 mRNA was probed with a psoralen-biotin labeled PR-3-m Sp6 probe. Lanes 1-7 represented: RNA ladder, THP-1 total RNA, Patient 1 total RNA, Patient 2 total RNA, Patient 1 poly(A) RNA, Patient 2 poly(A) RNA.

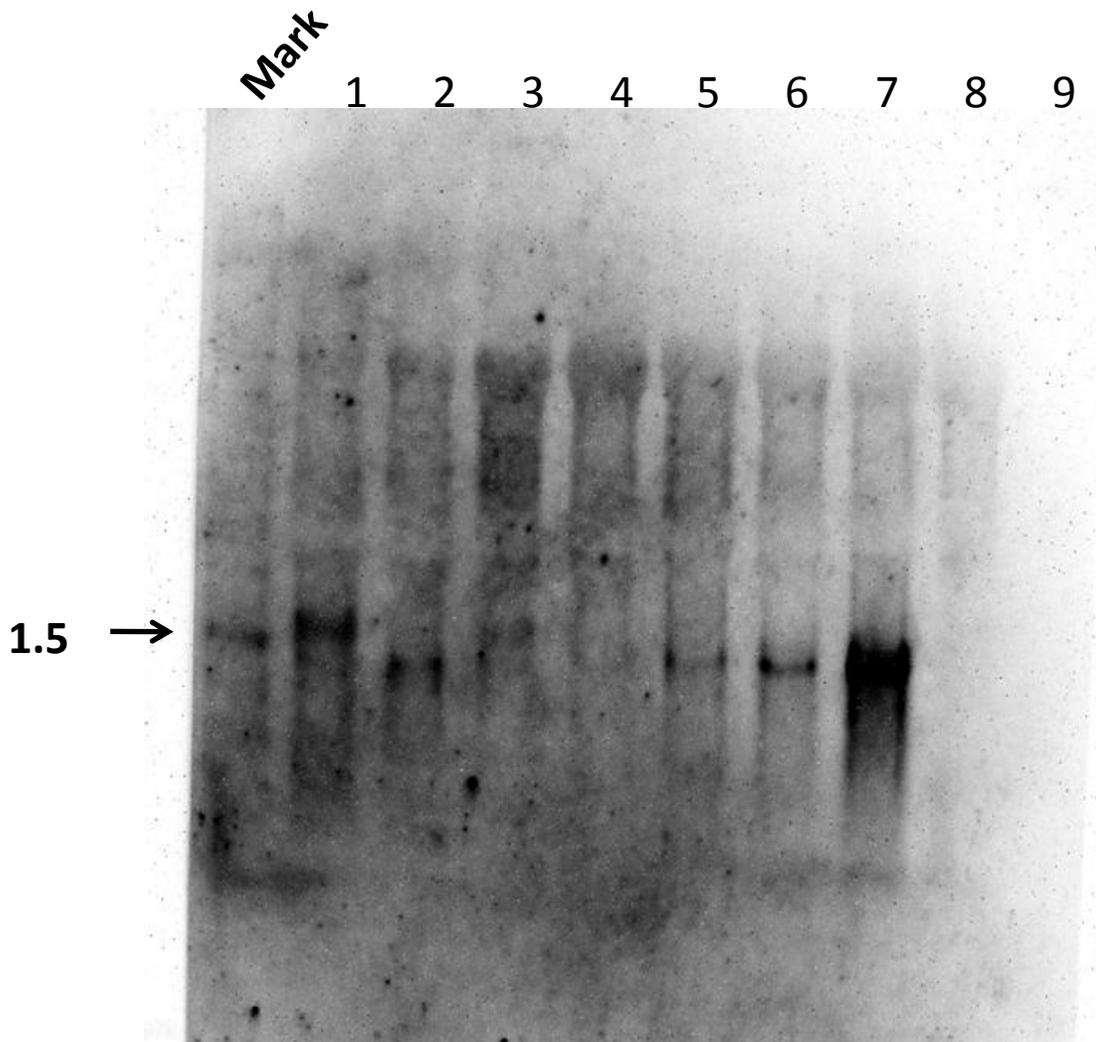


Figure 5. Northern blotting of sense PR-3. Lanes 1-4: Total RNA from PR3-ANCA patient leukocytes. Lanes 5-9: Aged matched healthy control leukocyte RNA. Patients in lanes 1 and 3 are expressing an aberrantly spliced PR-3 transcript not seen in the healthy controls. Lanes 2,5-7 are expressing varying levels of PR-3 at the expected size of ~1.3 kb.

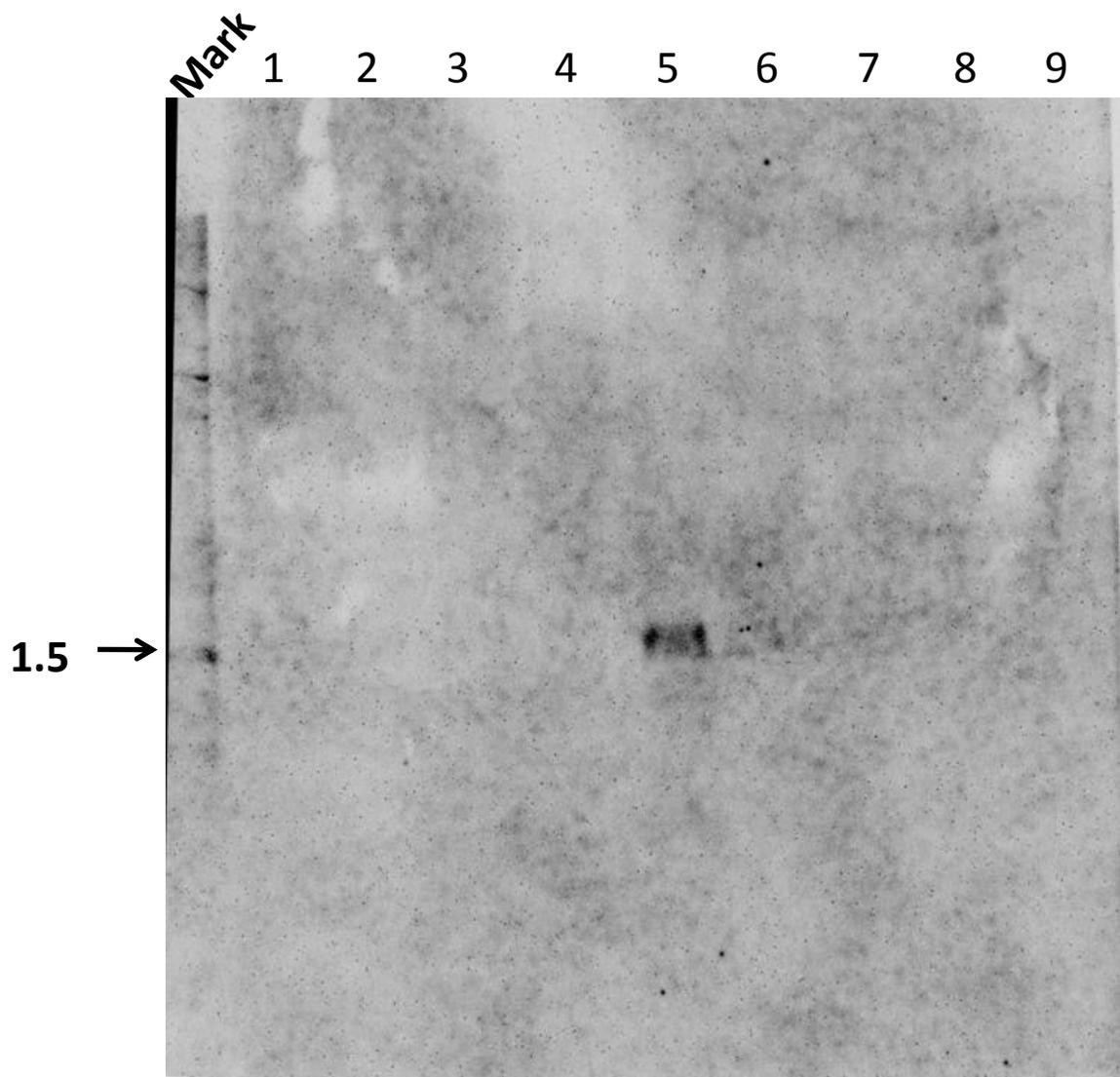


Figure 6. Northern blotting of sense PR-3 mRNA. Lanes 1-5: Total RNA from PR3-ANCA patient leukocytes. Lanes 6-9: Aged matched healthy control leukocyte RNA. One patient sample (represented in lane 5) also appears to be expressing an aberrant form of the PR-3 transcript at ~1.5 kb in size.

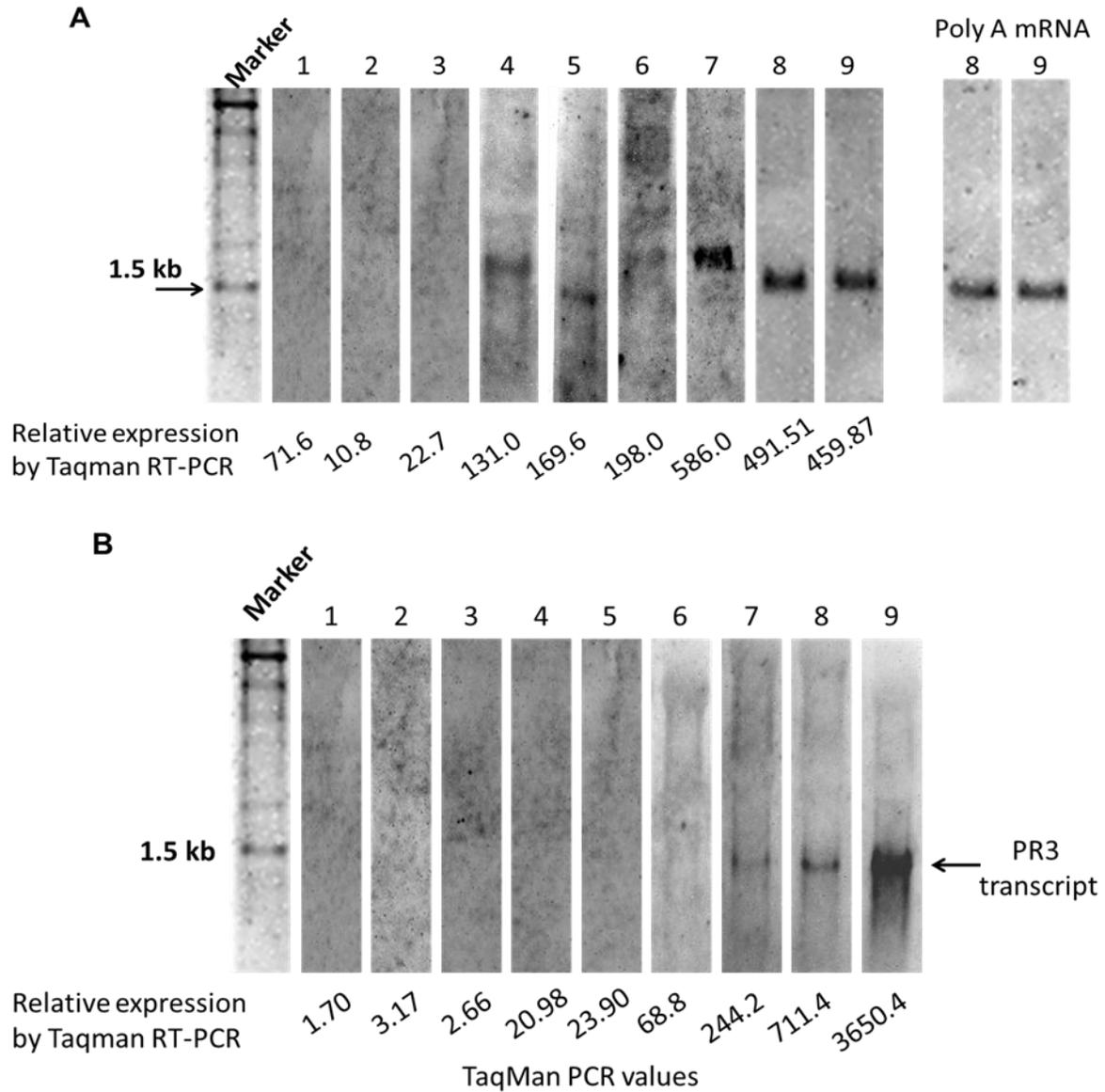


Figure 7: Semi-quantitative comparison of PR-3 blotting and PR-3 qRT-PCR. Before northern blotting a small aliquot of RNA was stored separately for gene expression analyses. Taqman data was normalized to standard curve dilutions of PR-3 transcript from high expressing THP-1 and the PR-3 null Jurkat cell lines. **(A)** Northern blotting of PR-3 in PR3-ANCA patients. **(B)** Northern blotting of PR-3 in aged matched healthy donors.

3.3. Characterization of PR-3 transcripts

We hypothesized that PR-3 mRNA expressed inappropriately may be altered in its 5' and 3' UTR usage and in splicing. The multiple isoforms found in patients by northern blotting prompted us to determine the variation in sequence between the normal and aberrant or altered forms of PR-3.

In order to characterize the 3' end of the PR-3 mRNA 3' rapid amplification of cDNA ends (RACE) was performed to determine the 3' end of PR-3 mRNA, comparing patient leukocyte RNA to that of healthy donors. RACE is a powerful technique for determining the extreme ends of a transcript requiring knowledge of enough sequence to design the forward primer for PCR as outlined in **Figure 8** [57]. The **Qt** 52-nt RACE 3' primer with adaptor sequences was used to prime RNA for reverse transcription. cDNA generated by **Qt** primer contains an additional anchor sequence to facilitate amplification of the unknown 3' end. In the first series of experiments a primer within the 5th exon of PR-3 ((+).634-PR3) was selected to minimize the potential distance between the known sequence and the appended adapter sequence. After nested PCR amplification, two isoforms of PR-3 mRNA were amplified from two individual patient samples. The lowest migrating form was found to be the currently known transcript which co-existed with a novel larger transcript (**Figure 9**). Gel excised PCR fragments were subcloned and sequenced by the UNC-CH Genome Analysis Facility. The data confirmed that PR-3 transcript was specifically amplified as shown by BLAST alignment (**Figure 10**). **Figure 11a** represents the sequence retrieved from the larger PCR amplicon. The additional sequence from the **Qt** adapter is indicative of a fully processed mRNA with a poly(A) tail. The inclusion adapter in the

sequenced amplicon gave confidence that we had the legitimate end of this particular PR-3 isoform (PRTN3-286). The corresponding alignment of PRTN3-286 with the genomic sequence of the *PRTN3* gene is depicted in **Figure 11b**. The yellow region highlights the fifth and last exon of *PRTN3*, the boxed regions represent identical 68-bp repeats with a unique 12-nt spacer within the intergenic sequence between PR-3 and elastase. The currently annotated transcript will be denoted as PRTN3-206 and the extended form as PRTN3-286; the numbering indicates the length of 3'UTR in nucleotides. The longer form, PRTN3-286, contains a repetitive terminal sequence with two 32-nt repeats, whereas PRTN-206 contains only one copy.

3' RACE

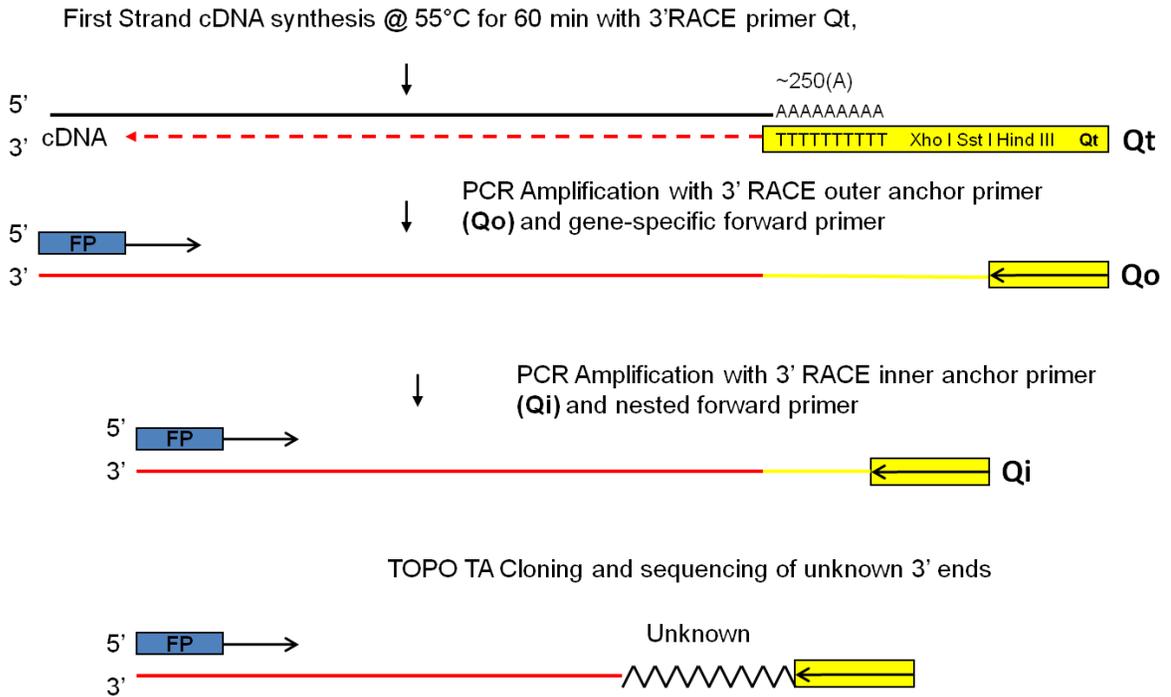


Figure 8. Schematic of 3' RACE. cDNA synthesis is carried out with specialized **Qt** primer, that contains a poly(dT) stretch for poly(A) binding which is preceded by restriction enzyme sites and a unique adapter sequence. The 'outer' race (**Qo**) primer is used in the first round of PCR amplification and a gene-specific forward primer. A second round of PCR is required to increase specificity, with a nested forward primer and the (**Qi**) 'inner' anchor primer. The many registers in which the **Qt** primer can bind the poly(A) give rise to a heterogenous population of PCR products, which requires cloning of a singular PCR product before sequencing.

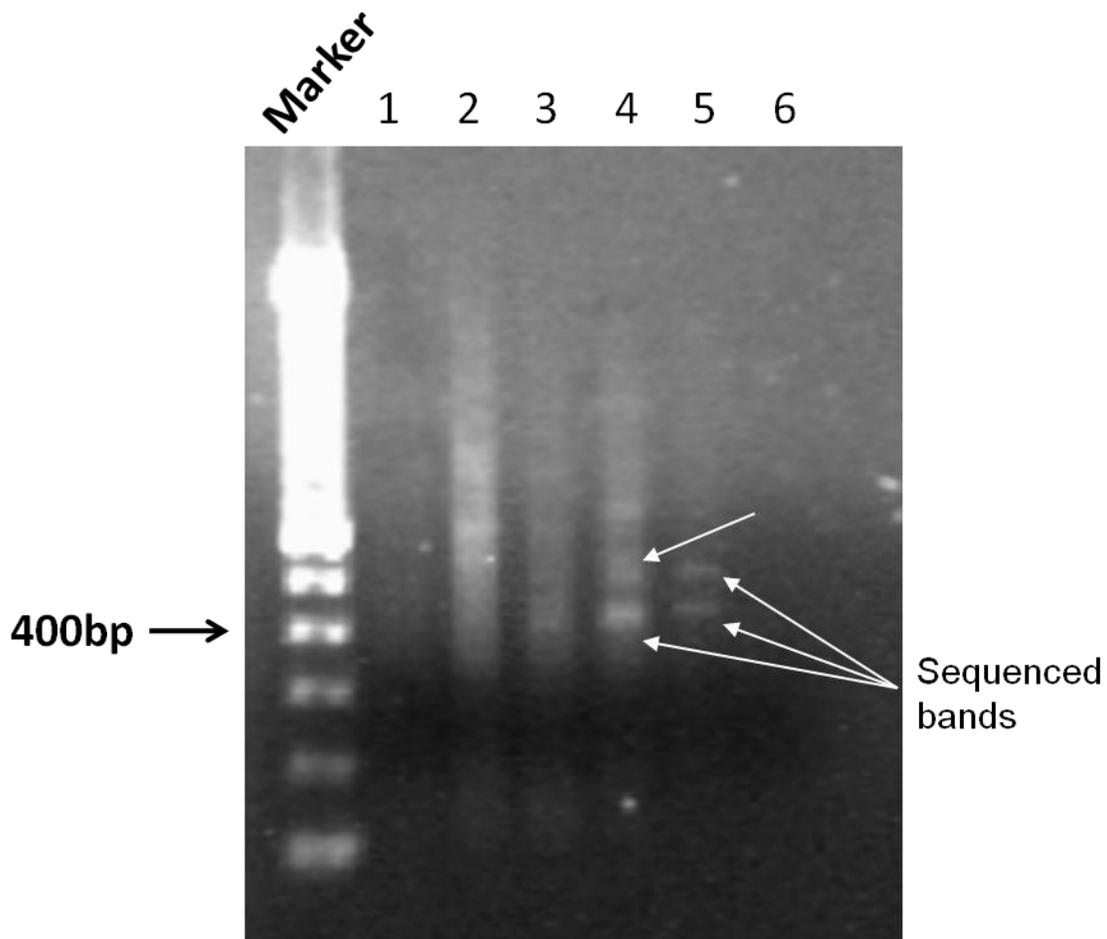


Figure 9. 3'RACE amplification of sense PR-3 from PR3-ANCA patient leukocyte RNA. Lanes 1-2: RNA from two different sample dates from PR3-ANCA patient 1. Lanes 3-4: Patient leukocytes from PR3-ANCA patient 2. Lane 5: PR3-ANCA patient 3 sample from only one sample date. 1st PCR amplification primers: (+).634-PR3 and **Qo**, 2nd round PCR primers: (+).634-PR3 and **Qi**. The smaller amplicons represent the currently annotated form of PR-3 (lanes 4 and 5) and the larger amplicons (also lanes 4 and 5) represent the larger alternatively polyadenylated PR-3 transcript.

GENE ID: 5657 PRTN3 | proteinase 3 (serine proteinase, neutrophil, Wegener granulomatosis autoantigen) [Homo sapiens]

Score = 515 bits (260), Expect = 6e-143
Identities = 307/323 (95%), Gaps = 2/323 (0%)
Strand=Plus/Plus

```
Query 14  CTTCGTGATCTGGGGATGTGNAANC-GCCTTCCCCCTGACTTCTTCACGCGGGTAGCCNT 72
          |||
Sbjct 678  CTTCGTGATCTGGGGATGTGCCACCCGCCCTTTCCCTGACTTCTTCACGCGGGTAGCCCT 737

Query 73  NTACGTGGACTGGNTCCGTTCCACGCTGCGCCGTGTGGAGGCCAAGGGCCGACCCTGAAC 132
          |||
Sbjct 738  CTACGTGGACTGGATCCGTTCCACGCTGCGCCGTGTGGAGGCCAAGGGCCGCCCTGAAC 797

Query 133  CGCCCCCTCCACAGNGCTGGCACGGGACCCCGAGCCTGGCTCCAAACCCTNGAGGGGGAT 192
          |||
Sbjct 798  CGCCCCCTCCACAGCGCTGGC-CGGGACCCCGAGCCTGGCTCCAAACCCTCGAGGGGGAT 856

Query 193  CTTTGGACAGAAGCAGCTCTTCCCCGAACACTGTGGCGTCCGGGACGGCCCCACCCGTCC 252
          |||
Sbjct 857  CTTTGGACAGAAGCAGCTCTTCCCCGAACACTGTGGCGTCCGGGACGGCCCCACCCGTCC 916

Query 253  CCCCACACTNCCTCCCACGGGGCTCCGGGAGACAGGCCGGCCCTGCACCTCACCCNACCG 312
          |||
Sbjct 917  CCCCACACTCCCTCCCACGGGGCTCCGGGAGACAGGCCGGCCCTGCACCTCACCCACCG 976

Query 313  TGACCTCAATAAACGTCGAAACT 335
          |||
Sbjct 977  TGACCTCAATAAACGTTGAAACT 999
```

Figure 10. 3'RACE sequencing and BLAST alignment of PR-3 mRNA. Sample BLAST alignment of 3'RACE PCR product derived from PR3-ANCA patient leukocyte RNA. This clone specifically aligned to the currently annotated form of PR-3 mRNA.

NATTATNCAAGNNATNNNNTCNTTNGTAATNNGGGGATGTNCTNTNNGCATT TTTGGNNTGAN
 TTTTAAANGGGGGTAGNNCTNTANGTGGATNGGATNNGTANCTNGNTGNGNNGTNTGGAGGN
 CAAGTTCNGCTCATNANNCGNCCCTGAANANCNCTGNCNGGGATNNCTAGCNTGGCTCCNAA
 CCCTTGAGNCGGATCTTTGGACANAANCAGCTCTTCCCCGAACACTGTGGAGTCCGGGACGG
 NCCCACCCGTCCCCCACACTCCCTCCCACGGGGCTCCGGGAGACAGGCCGGCCCTGCACCT
 CACCCACCGTGACCTCAATAAACGTTGAAACTCCCCCCTGGCTCCTGTCTGTCTTCCCATA
 TAGGGAGGGGAGGGGGGAAGCTGCCCACCGTGACCTCAATAAACGTTGAAACTCCCAAAAAAAAA
 AAAAAAAAAAGCTTGAGCTCGAGTCCTC

Figure 11A. Sequence obtained from TOPO clone of larger RACE band. Canonical polyadenylation of PRTN3 occurs at 12-nt downstream of first AATAAA signal. *PRTN3-286* is cleaved at the distal cytosine before polyadenylation. The red and underlined cytosines represent the two mRNA cleavage sites of PR-3. PR3-ANCA patient was found to possess an additional UTR. Text in grey represents vector sequence. Black text successfully aligned with *PRTN3*. Sequence in blue represents **Qt** adapter sequence. Red sequence represents a unique spacer element which separates the repetitive poly(A) signals in *PRTN3-286*.

TCCATATTGGAAGCATT TTTATTCTGAAACAGCTGTGGCTACCCAGCCCCCTTCCTGGCTGC
 CTGGGGCCAGCCCGGGTGACTGGCTGTCCCATCCTCCCGGGAGACTCAGGTGGCCCTGAT
 GGGTGACTGGCCGTCCCATCCTCCAGGGAGACTCAGGTGGCCCTGATGGGTGACTGGCCG
 TCCCTGTCCTCCAGGGAGACTCAGGTGGCCCCCTGATCTGTGATGGCATCATCCAAGGAATA
 GACTCCTTCGTGATCTGGGGATGTGCCACCCGCCTTTTCCCTGACTTCTTCACGCGGGTAGC
 CCTCTACGTGGACTGGATCCGTTCACGCTGCGCCGTGTGGAGGCCAAGGGCCGCCCTGAA
 CCGCCCCCTCCCACAGCGCTGGCCGGGACCCCGAGCCTGGCTCCAAACCCTCGAGGCGGATCT
 TTGGACAGAAGCAGCTCTTCCCCGAACACTGTGGCGTCCGGGACGGCCCCACCCGTCCCCC
 AACTCCCTCCCACGGGGCTCCGGGAGACAGGCCGGCCCTGCACCTCACCCCACCGTGACCT
 CAATAAACGTTGAAACTCCCCCTGGCTCCTGTCTGTCTTCCCATATAGGGAGGGGAGGGGG
 AACTGCCACCGTGACCTCAATAAACGTTGAAACTCCCCCCTGGCTCCTGTCTGTCTTCCCATATAGGGAGGGGAGGGGG
 TATAGGGAGGGGAACCAGGGAGAGGTCCCCTGGGTTCAAGTCCCCGAGCCCCAAACTCAGGAT
 GGGGGGAGTCGGAGCGGGGGATGGGGGCCAGGTTGCAGGGGTCTCTCTCCTGGATGCTCG
 TGGCTGTGGTTCTAACCCGGCCCCATCAGGGTGGCCCTGGGCAGGTTGTTATTCATCTCC

Figure 11B. Genomic sequence of PRTN3 exon 5, with crucial elements highlighted. Highlighted yellow text represents the canonical exon 5 of PR-3. The extended UTR found in *PRTN3-286* continues to align with intergenic sequence between *PRTN3* and *ELA*. Boxed regions represent exact 68-bp repeats found in the genomic sequence of *PRTN3*. Red sequence (TAG) represents the canonical stop codon of PR-3.

A larger set of RNA samples from patients and healthy donors was evaluated by 3'RACE. The data confirmed expression of both PRTN3-206, PRTN3-286 as shown in lanes 3-4 (PR3-ANCA) and 4-8 (healthy controls) (**Figure 12**). PCR amplicons were subcloned into Strataclone Blunt TOPO PCR cloning vector and sequenced. Interestingly, sequence analysis of plasmid clones from the healthy control in lane 8 had both forms of PRTN3-206 and PRTN3-286. The remaining healthy controls were also expressing PRTN3-206 (Lanes 5-7). One of the ANCA patient's (Lane 3) had PRTN3-206, and a second patient (Lane 4) was expressing PRTN3-286. Although PCR products from lanes 1 and 2 were gel-excised and cloned into the Strataclone blunt PCR cloning vector, none of the sequenced clones contained PR-3 sequence. Based on the weak amplifications from RACE, we concluded that this represented a difficult region for *Phusion* or *Taq* polymerase to accurately copy. In order to amplify this region, optimization of both reverse transcription and PCR were deemed necessary.

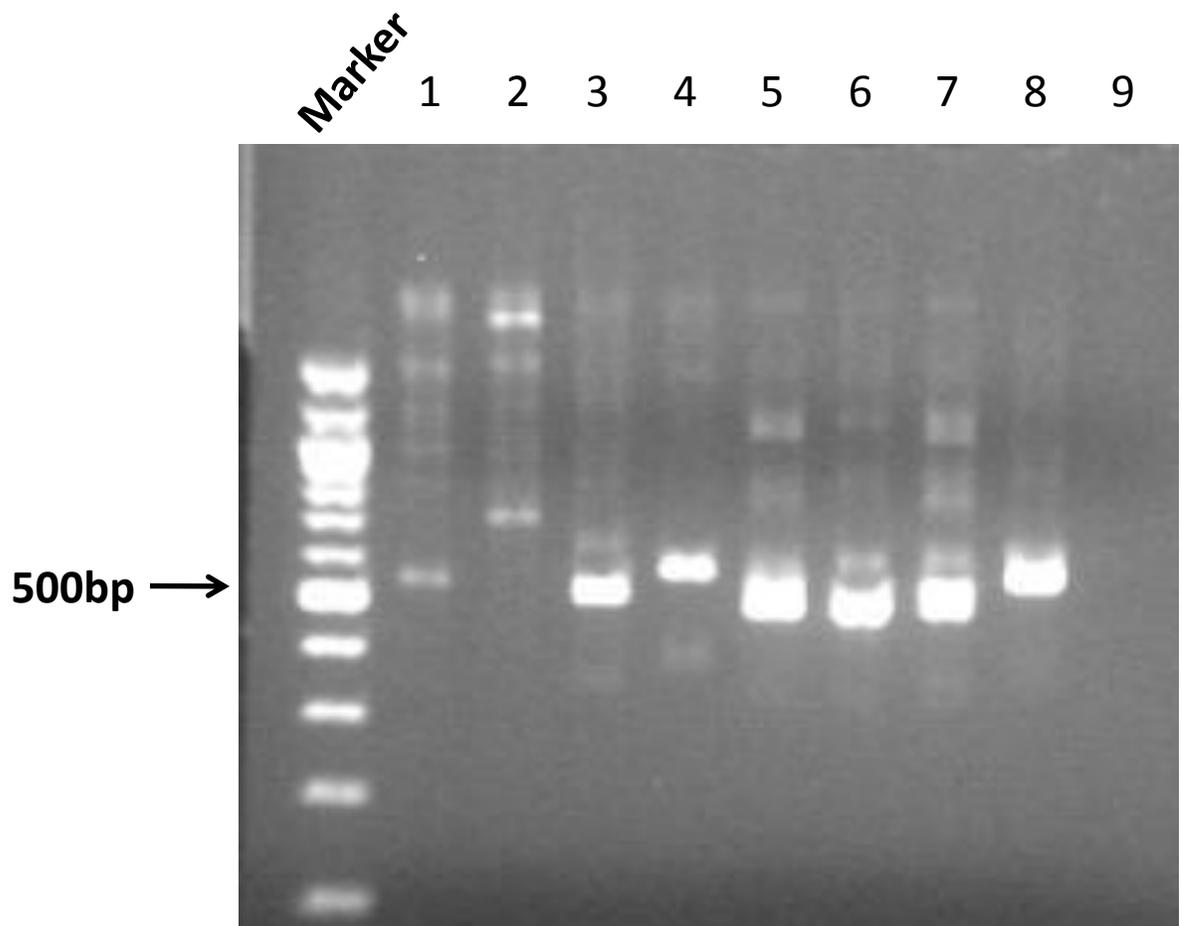


Figure 12. 3' RACE amplification of 3' ends of PR-3 transcript of total leukocyte RNA. Lanes 1-4 total RNA from patients with PR3-ANCA; lanes 5-8 age matched healthy controls patient. Lane 9, Water control. Primers used for RACE - RT Primer: Qt, PCR 1st Round: PR3.374 + Qo, PCR 2nd Round: PR3.547 +Qi.

To maximize the likelihood of detecting splice variants, we designed four different sets of nested forward primers for PCR (**Figure 13**). **Qo** was used as the reverse primer in the 1st round PCR reactions and **Qi** in all the 2nd round PCRs (**Figure 8**). The data were unique for each primer set. For example, primer set 1 produced a band only in RNA from patient 3(P3), the same patient tested by northern blot (**Figure 6**, lane 1) who expressed a larger form of PR-3. Primer set 3 was the most useful in that all four samples tested were successfully amplified by 3'RACE while primer sets 2 and 4 were unsuccessful (**Figure 13b**). An example of sequencing results of clones sampled from various bands is shown in **Figure 14a**. Sequencing results contained many 'N' nucleotides which were determined to be an issue with the extremely high GC content in the amplicon. However, this amplicon contained sequences from intron 3 and was also polyadenylated. Nucleotides from exon 4, and the majority of exon 5 were missing, although this was attributed to reverse transcriptase's template switching activity (**Figure 15**). Nevertheless, the additional intronic sequence if translated, it could code for an abnormal form of PR-3, which would be recognized as a foreign antigen to the immune system.

A

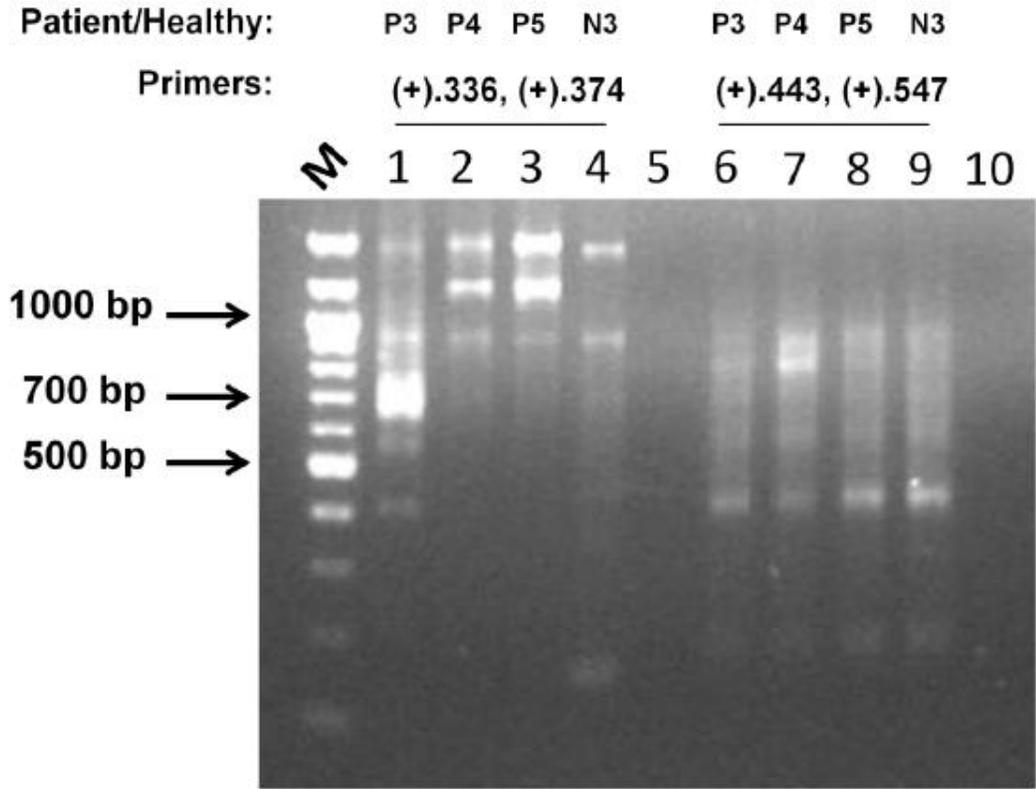


Figure 13(A). 3'RACE amplification with GC optimization with multiple nested primer sets. RNA Three PR3-ANCA patients: P3, P4, and P5 and one healthy donor: N3 was used in a modified RACE reaction optimized for high GC content. Lane 1: Only successful RACE amplicons from either set of nested primers. The smeared band from ~650 – 770 bp was used for cloning in Strataclone blunt sequencing vector.

B

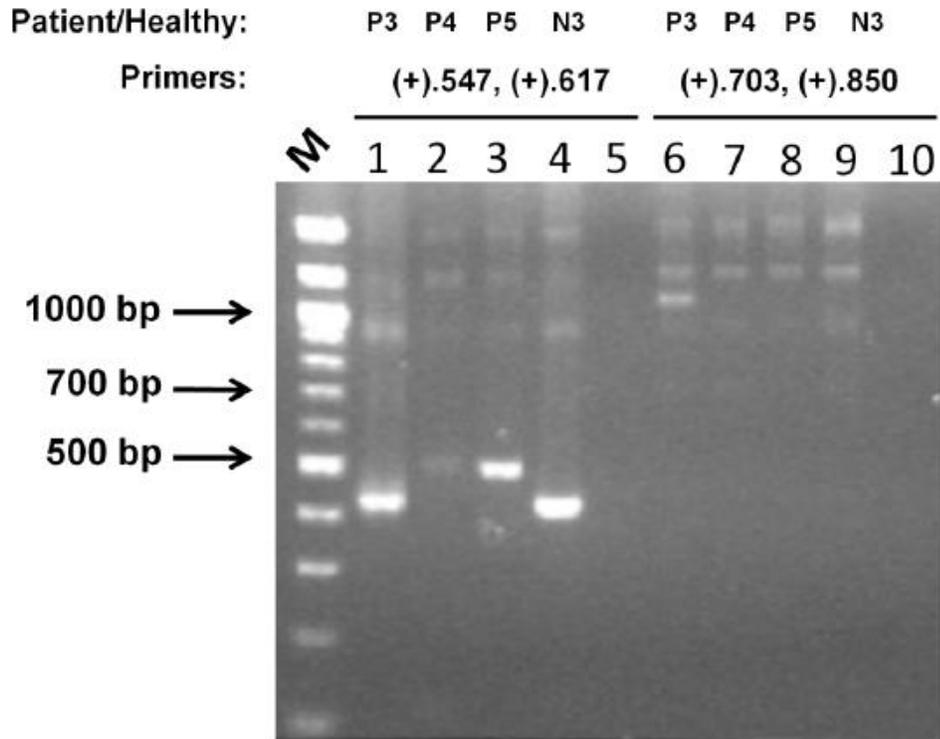


Figure 13(B). 3'RACE amplification with GC optimization and multiple primer sets. RNA Three PR3-ANCA patients: P3, P4, and P5 and one healthy donor: N3 was used in a modified RACE reaction optimized for high GC content. Lanes 1-4: All RACE reactions were successful, larger bands represent PRTN3-286, whereas smaller PR-3 bands represent PRTN3-206. Lanes 6-9: (-).703-PR3, (-).850-PR3 primer not suitable for this PR-3 3'RACE, all bands are from non-specific PCR amplification.

NTTNGCCCTTANGAGNTTNTCNTNATCCAGGTGGGNGGGCAGGGCCNCNAGGGNTNGGAGGGGCNNGNCC
AACAACTGAA CGACGTTCTCCTCATCCAGGTGGGCGGGCAGGGCCGCGAGGGCTCGGAGGGGCACGGCC

ANAGGGNTCNGGGNCCCCNNTTCNTGNAGCNAGCNNTTCNTTGNGCNCCCNTGTATNCNGGGCCNCGNCC
AGAGGGCTCCGGGACCCCCATTCTGCAGCCAGCATTTCATTGAGCACCCACTGTATACGGGGCCACGACC

NNGGCCGNTGCAGCNTGGGTCCAGTGGCNNTGGCNGGGGGAGNCCGCTCTTTGGACNCCNGGCCNNTCCT
GAGGCCGCTGCAGCCTGGGTCCAGTGGCACTGGCCGGGGGAGACCGCTCCTTGGACACCAGGCCACTCCT

CGGGCCTTGCCCCTCNCCCCCGNTGNCCTCAATAAANNTTGAANCNC*
CCGGCCCTGCACCTCACCCACCGTGACCTCAATAAACGTTGAAACTCC

Figure 14. Sequencing results from P3 3'RACE PCR clone. The sequence alignment program makes an 'N' call when the signal from the nucleotide present is ambiguous. This clone also contains a template switched cDNA product, along with sequence from intron 2. First sequence in yellow (AACAACTGAA) demarks the breakpoint in template switched sequence. Sequence in between the two yellow regions indicates sequence from intron 2. Last sequence in yellow demarks partial sequence of exon 5.

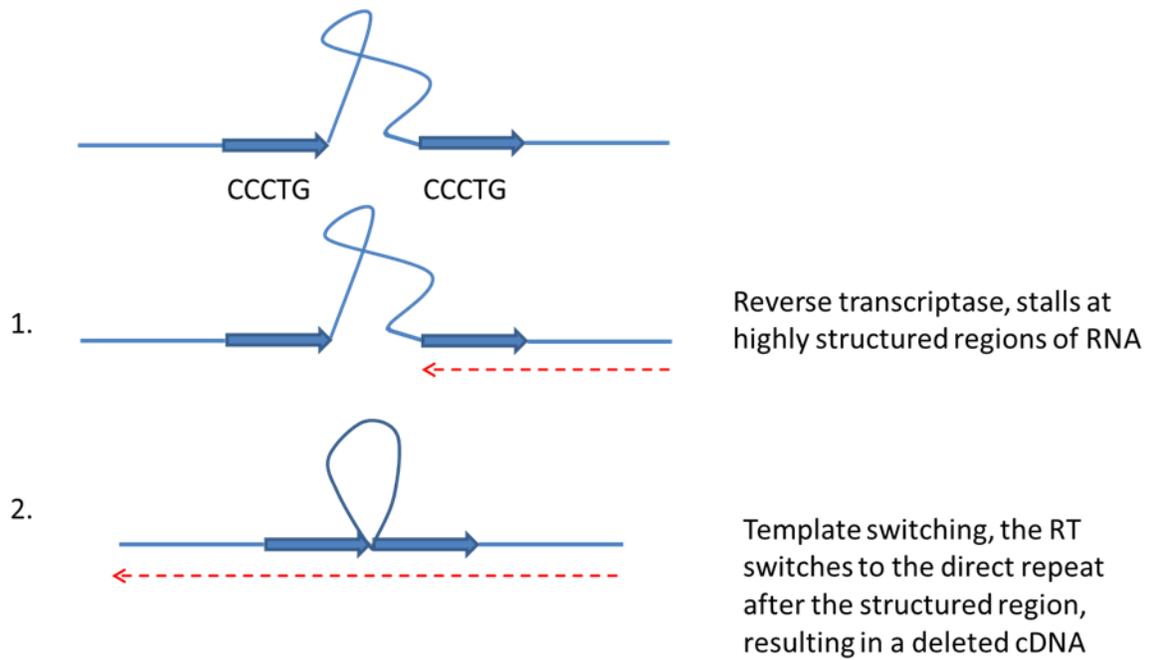


Figure 15. Schematic of template switching. Template switching is a normal phenomenon seen in viral replication processes. *In vitro* template switching can happen randomly in a cDNA synthesis reaction with a properly structure RNA possesses small direct repeats. The RT re-initiates cDNA synthesis at the second repeat effectively deleting the intervening region.

3' cDNA cloned

CNGAANCAAANGCTTAGTCGCTTGGCTTGCAATTTTCTGGTACCCGATTNGAAACGATACGC
CTGATACTAANC GTCCCCAGCGACGACNAGCGNACGANTAGTAGCAGAACGGANAGCGCCCA
TACGCAACCGCTCTCCCGGGNGTGGCGATTTCATNATGCAGCTGGCACGACAGGTTNCCGACT
GAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGG
CTTTACACTTTTATGCTCCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTAC
ACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACA
AAAGCTGGGTACCGGGCCCCCTCGAGGTGACGGTATCGATAAGCTTGATATCCACTGTG
GAATTCGCCCTTACGACGTTCTCCTCATCCAGCTGAGCAGCCCAGCCAACCTCAGTGCCTCC
GTCGCCACAGTCCAGCTGCCACAGCAGGACCAGCCAGTGCCCCACGGCACCCAGTGCCTGGC
CATGGGCTGGGGCCGCGTGGGTGCCACGACCCCCAGCCAGGTCTGCAGGAGCTCAATG
TCACCGTGGTCACCTTCTTCTGCCGGCCACATAACATTTGCACTTTCGTCCCTCGCCGCAAG
GCCGGCATCTGCTTCGGGAGACTCAGGTGGCCCCCTGATCTGTGATGGCATCATCCAAGGAAT
AGACTCCTTCGTGATCTGGGGATGTGCCACCCGCCTTTT **CCCTG** **CACCTCACCCACCGTGA**
CCTCAATAAACGTTGAAACTCCAAAAAAAAAAAAAAAAAAAAA

Exon 5

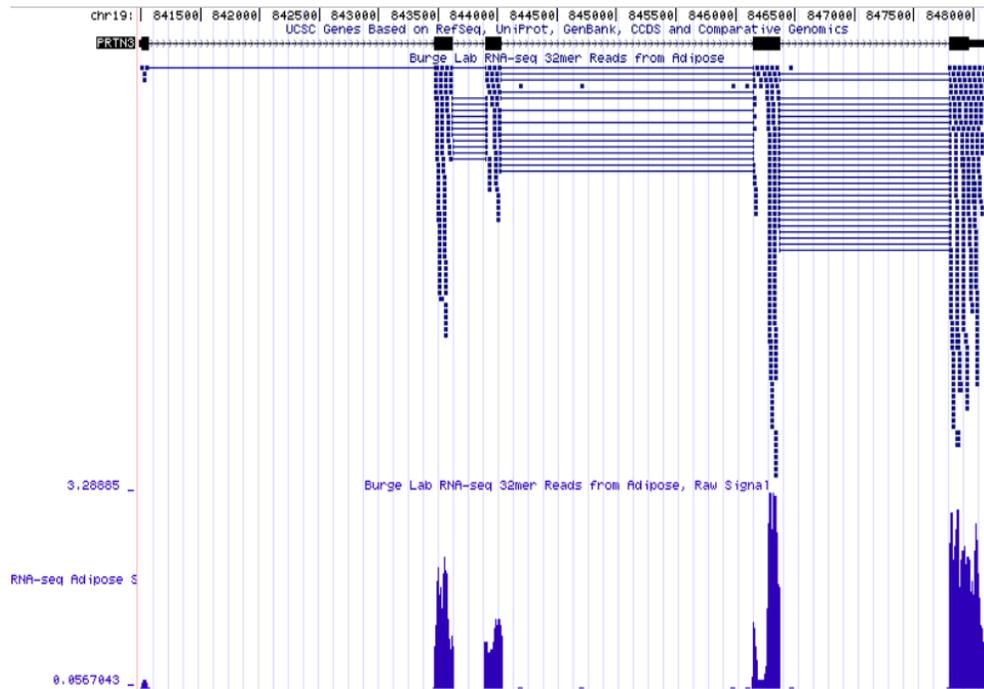
CTGAGCAGCCCAGCCAACCTCAGTGCCTCCGTGCCACAGTCCAGCTGCCACAGCAGGACCA
GCCAGTGCCCCACGGCACCCAGTGCCTGGCCATGGGCTGGGGCCGCGTGGGTGCCACGACC
CCCAGCCCAGGTCTGCAGGAGCTCAATGTCACCGTGGTCACCTTCTTCTGCCGGCCACAT
AACATTTGCACTTTCGTCCCTCGCCGCAAGGCCGGCATCTGCTTCGGGAGACTCAGGTGGCC
CCTGATCTGTGATGGCATCATCCAAGGAATAGACTCCTTCGTGATCTGGGGATGTGCCACCC
GCCTTTT **CCCTG** **ACTTCTTACGCGGGTAGCCCTCTACGTGGACTGGATCCGTTCCACGCTG**
CGCCGTGTGGAGGCCAAGGGCCGCCCTGA **ACCGCCCTCCACAGCGCTGGCCGGGACCCC**
GAGCCTGGCTCCAAACCTCGAGGCGGATCTTTGGACAGAAGCAGCTCTTCCCCGAACACTG
TGGCGTCCGGGACGGCCCCACCCGTCCCCCACACTCCCTCCCACGGGGCTCCGGGAGACAG
GCCGG **CCCTGCACCTCACCCACCGTGACCTCAATAAACGTTGAAACTCC** **CCCTGGCTCCTG**
TCTGTCCCTTCCCATATAGGGAGGGGAGGGGGAAGTGCCACCGTGACCTCAATAAACGTTGA
AACTCCCCCTGGCTCCTGTCTGTCTTCCCATATAGGGAGGG

Figure 16. 3' RACE clone from P5. (A) Sequence of template switched cDNA product missing a portion of exon5. **(B)** cDNA has entire region in red deleted from exon 5. The highlighted green repeats depict the direct repeats required for template switching of reverse transcriptase. Yellow sequence represents a truncation of exon 5 caused by RT-mediated template switching.

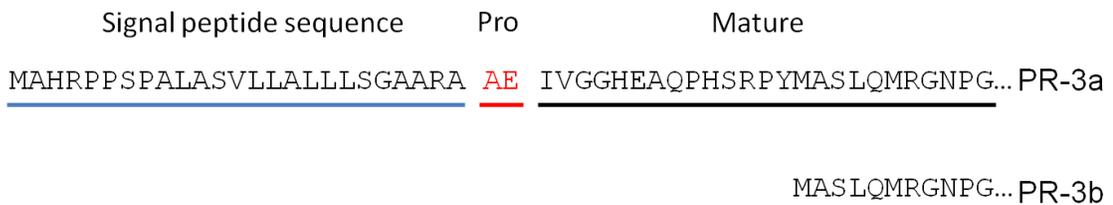
3.4. Myeloblastin and PR-3 are not encoded by a single mRNA

Labbaye et al. published that myeloblastin and PR-3 were encoded by the same mRNA based on the sequence alignments of both the protein and mRNA sequences [20]. In the 1990s there was still disagreement about the 5' ends of either mRNAs, however it was assumed that the sequence cloned for myeloblastin was truncated due to sequencing artifact and was in fact simply PR-3. We performed bioinformatic analyses of spliced expressed sequence tag (EST) from MCF-7 cell lines and human bone marrow databases and searched for sequences with homology for PR-3 (**Figure 17a**). These databases contained additional PR-3 transcripts that had skipped the canonical exon 1. These exon 1 skipped transcripts actually represented the originally discovered myeloblastin, which until now was thought to be synonymous with PR-3. Because of the high homologies between the potential protein product and the mRNA sequences, these two proteins were incorrectly assigned as 1 protein. A schematic of the two proteins is depicted in **Figure 17b**, along with the exact sequence of the alternate transcript. The exon 1 skipped protein remains in frame with the currently annotated PR-3 protein. We believe this truncated protein represents the originally reported myeloblastin, which was also verified by mining RNA-seq databases. In order delineate between PR-3 and myeloblastin, the proteins should be renamed with the currently annotated form of PR-3 protein as PR-3a and the exon 1-skip variant as PR-3b. By skipping exon 1, PR-3b lacks the signal peptide found in PR-3a, suggesting its packaging, folding, localization and function is different. This also implies that PR-3b utilizes a different transcriptional start site (TSS) and promoter sequence than PR-3a (**Figure 17c**). Evaluation of EST databases support the existence of unique TSS utilized by the two different isoforms of PR-3.

A



B



C

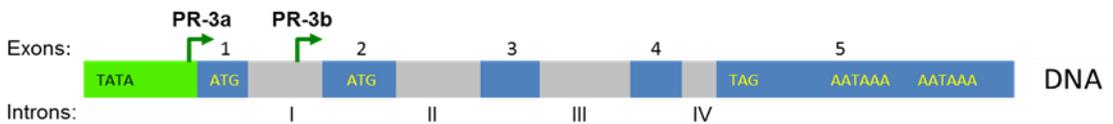


Figure 17. Comparison of PR-3 and myeloblastin protein sequences. (A) RNA-seq alignments from adipose tissue also show that most of the reads align only in exons 2-5 of PRTN3 gene. Data was retrieved from publically available Burge lab RNA-seq track from the UCSC genome browser. **(B)** PR-3b lacks the signal peptide sequence found in PR-3a, the dipeptide pro sequence and an additional 14 amino acids. PR-3 utilizes a cryptic start codon found in exon 2. **(C)** Predicted transcriptional start site of PR-3b from EST databases. The TSS for PR-3a is further upstream of PR-3b and includes transcription of exon 1.

3.5. Detecting an antisense transcript co-expressed with sense PR-3 mRNA

The histone marks identified at the *PRTN3* gene locus are characteristic of an ‘open’ chromatin conformation. This conformation facilitates aberrant sense PR-3 transcription would theoretically would also allow transcription from the non-coding strand [58]. Previous studies suggested the existence of an antisense transcript in patients with ANCA disease [16]. It is particularly difficult to study an antisense transcript that co-exists with its sense RNA counterpart in the same sample. First attempts at amplifying the antisense transcript by standard RT-PCR were confounded by ‘endogenous’ priming of the sense transcript (snap-back priming) (**Figure 1**). Reverse transcriptases are notorious for priming cDNA from simple RNA secondary structure and not the inputted DNA oligonucleotide, which creates untargeted cDNA synthesis. This creates a scenario where both the sense and antisense transcript are copied into cDNA and are effectively equivalent molecules after double strand conversion in PCR.

To study antisense transcription, the first issue we needed to resolve was snap-back priming of the sense PR-3 transcript (**Figure 1**). To test the strand-specificity of a routine RT-PCR reaction, we used three substrates: 1) HEK 293 cell line transfected with antisense PR-3 plasmid, 2) RNA from a human monocytic cell line known to express moderate levels of sense PR-3 mRNA and 3) RNA from total leukocytes from a patient also containing aberrantly expressed PR-3 transcripts. PR-3 sense mRNA is efficiently converted to cDNA without the addition of exogenous gene-specific RT primer (**Figure 18a**). Surprisingly, the antisense PR-3 transcript was also efficiently copied to cDNA without addition of exogenous

RT primer. This approach was unsuccessful in determining the strand orientation of the transcripts.

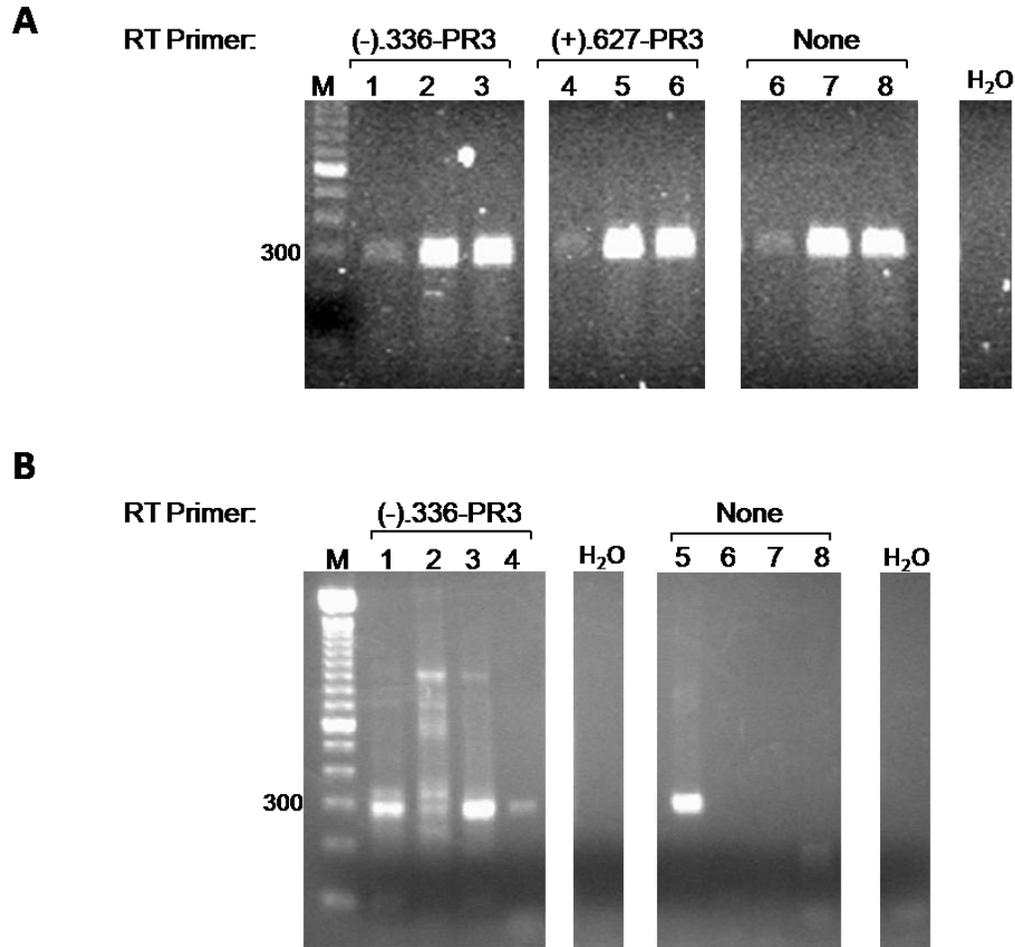


Figure 18. Analysis of strand-specificity (A) Reverse transcription of PR-3 with a primer targeted for antisense ((+).336-PR3), sense ((-).627-PR3) or without primer followed by PCR amplification of a region of PR-3 mRNA (nt 336-627). Lanes 1,4,6 Patient 1 total RNA, Lanes 2,5,7 293-freestyle transfected with plasmid vector designed to express antisense PR-3 (nt. 627-336), Lanes 3,6,7 U937 total RNA. (+).336-PR3 RT primer was used to target an antisense PR-3 transcript (Lanes 1-3), (-).627-PR3 was used to target sense PR-3 and primer was omitted to determine if PR-3 cDNA was generated from endogenous priming. In all cases an amplicon was produced regardless of the strand targeted. DNA size marker shown in Lane M. **(B)** Reverse transcription carried out at elevated temperatures (60 °C) with either (+).336-PR3 or without primer followed by PCR for PR-3 (336-627). Lanes 1,5 U937 total RNA, Lanes 2,6 PR3-ANCA patient 2, Lanes 3,7 PR3-ANCA patient 2, Lane 4,8 PR3-ANCA patient 4. In lanes 1-4 (+).336-PR3 primer was used for reverse transcription and in lanes 5-8 primer was omitted from the RT reaction. Non-specific priming of sense PR-3 mRNA was still observed in U937 total RNA even at a 60 °C reaction RT temperature.

A published approach to block snap-back priming employs higher reaction temperature is expected to reduce any secondary structure responsible for the formation of these ‘false’-primed cDNAs. To accomplish this, a thermostable reverse transcriptase is used to carry out cDNA synthesis at high temperatures. At a RT reaction temperature of 60°C self-priming of the sense PR-3 transcript was negated from RNA in three patients (Figure 3b), however self-priming was not blocked from the U937 cell line, as indicated by the minus RT primer control (**Figure 18b, lane 1**). The inability to block self-priming in the U937s may represent a difference in the 3’ ends of the PR-3 transcripts or differences in the absolute quantities. Moreover, the elevated RT reaction temperature does not sufficiently disrupt the secondary structure responsible for spurious priming, presumably due to the greater stability of RNA:RNA hybrids as compared to the corresponding RNA:Oligo hybrids.

‘Tagged strand specific’ RT-PCR is an alternative approach used by many investigators to overcome problems with ‘endogenous’-priming. This approach employs a gene-specific cDNA primer with the inclusion of a ‘tag’ sequence to specifically tag cDNA produced [59]. This ‘tag’ sequence is then used to differentiate products formed by oligo-driven rather than ‘endogenously’ primed cDNA synthesis. PCR with a gene-specific forward primer and a ‘tag’-only reverse primer is used to selectively amplify the targeted cDNAs. In theory, only the targeted strand containing the tag sequence can be amplified by PCR. Unfortunately, residual ‘tag’-RT primer can act as a primer in the subsequent PCR reaction, thus adding the ‘tag’ sequence to amplicons of either polarity, confounding the proper interpretation of strand specificity [59]. A few protocols have tried to circumvent this

problem by using size exclusion columns to selectively remove unused tagged RT primer from the cDNA synthesis reaction. Although this can improve the strand-specificity of the PCR reaction, literally 100% of the RT primer must be removed in order for the reaction to be truly strand-specific. Techniques such as tagged strand-specific RT-PCR do improve detection of the targeted strand; nonetheless these are still associated with non-specific amplification of the incorrect strand (**Figure 19**).

We employed tagged strand specific RT-PCR to attempt strand-specific cDNA amplification. We removed excess RT primer by size exclusion columns after the cDNA synthesis reactions to improve strand-specificity of the subsequent PCR reaction. With a majority of 'tagged' RT primer removed we still observed a faint band in the sense RNA specific reaction, in a sample which only contained the synthetic antisense RNA presumably a product of the snap-back priming of the antisense RNA. Although depletion of excess RT primer eliminated false positives in reactions targeted for antisense it was not sufficiently stringent enough for sense specific PCR (**Figure 20**). While levels of transcript in a transfected cell line do not represent physiologic levels, it appears that if enough template the potential for false-positives is still a legitimate concern.

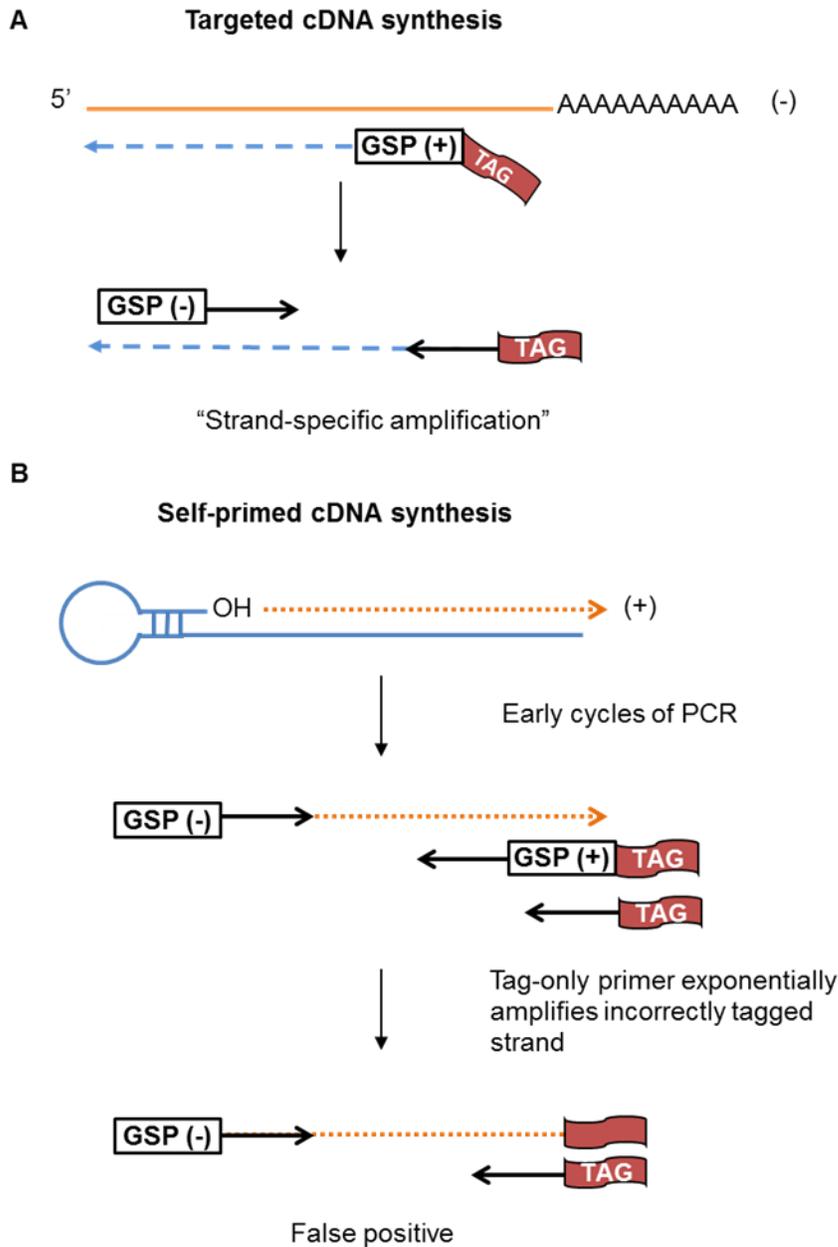


Figure 19. Issues with strand-specific RT-PCR (A) ‘tagged’ strand-specific RT-PCR. cDNA synthesis with a primer containing a gene-specific sequence along with an additional overhanging 16-bp tag. Theoretically, only cDNA generated with the ‘tag’-gene specific oligo contains the additional tag sequence required for PCR. PCR is carried out with a separate ‘tag’ only primer as well as a gene-specific primer for specificity. **(B)** In the case of a self-primed cDNA if residual ‘tagged’-RT primer is carried over it can also act as a primer during PCR and add the additional tag sequence to the untargeted cDNA during early cycles of PCR. ‘Tag’ only primer along with the gene-specific forward primer will subsequently amplify the incorrectly tagged product.

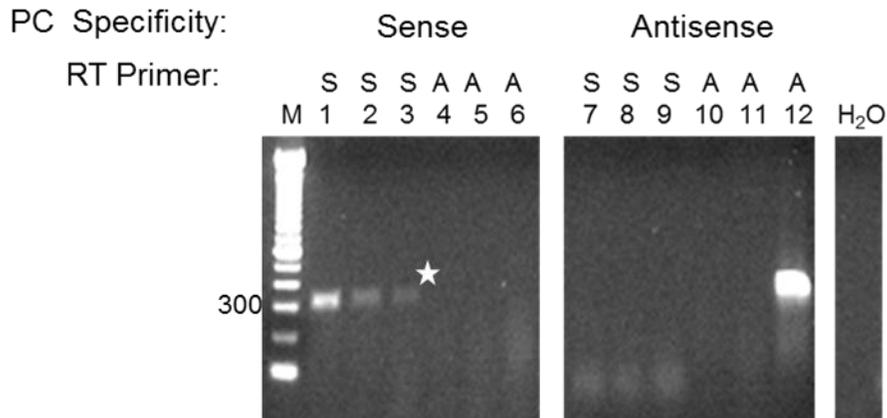


Figure 20. Tagged strand-specific RT-PCR after Microcon removal of unreacted ‘tag’ RT primer. Lanes 1,4,7,10 U937 total RNA, Lanes 2,5,8,11 U937 total RNA after PMA treatment, Lane 3,6,9,12 293 cell line transfected with an artificial antisense PR-3 transcript. RNA was reverse transcribed with a ‘tag’ gene-specific oligo designed to hybridize with either Sense (S) or Antisense (S). Orientation specificity of PCR was determined by the gene-specific forward primer used. For a sense-specific reaction ‘tag’ and (+).336-PR3 primer were used, and for antisense specific reactions ‘tag’ and (-).627-PR3 primer were used. A faint band for still appears in Lane 3, which is not expressed in the 293 cell line, indicating a false positive.

3.7. A novel antisense transcript to PR-3 is detected by ligation-anchored RT-PCR

The next approach adopted to retain strand-specificity was ligation-anchored RT-PCR (LA RT-PCR) described by Troutt et. al [60]. Although both methods described above are most widely used for strand-specific RT-PCR, this seemingly underused technique has addressed the problems associated with using tagged primers or high temperature RT reactions. LA RT-PCR, is used for both 3' end determination of unknown sequences as well as retention of strand-specificity in the PCR reaction. The strategy for ligation anchored PCR is outlined in **Figure 21a**. RNA is directly ligated at the 3' end to another short anchor deoxyribonucleotide or more recently a ribonucleotide, which is eventually used as a universal reverse transcription primer binding site. The anchor ribooligo was synthesized with a terminal $-NH_2$ group to block the 3'OH and a 5' phosphate. This ensures that ligation only occurs on the 3' ends of the mRNA and also serves as sequence to differentiate between sense and antisense transcripts.

Antisense PR-3 was detected by LA RT-PCR in 7 of 20 patients with PR3-ANCA as well as in the human monocytic leukemic THP-1 cell line (**Figures 21b, 22**). Sequencing analysis of AS-PR3 clones from patients all terminated with the same sequence at the 3' end (**Figure 23**). The cloned antisense PR-3 fragments also contained sequence from the ligated adapter, proof that this originated from an antisense RNA molecule. However, none of the amplified transcripts had a poly (A) tail, which is normally present on a traditional mRNA (**Figure 22**). The lack of poly(A) tail suggests that the currently amplified molecules may not represent the true full length transcript found in patients.. The potential for additional

sequence at the 3' end could theoretically be obtained by optimization of PCR cycling temperatures. By employing an array of PCR cycling temperatures the amplification of difficult templates that were potentially missed in the current set of experiments could be successfully amplified.

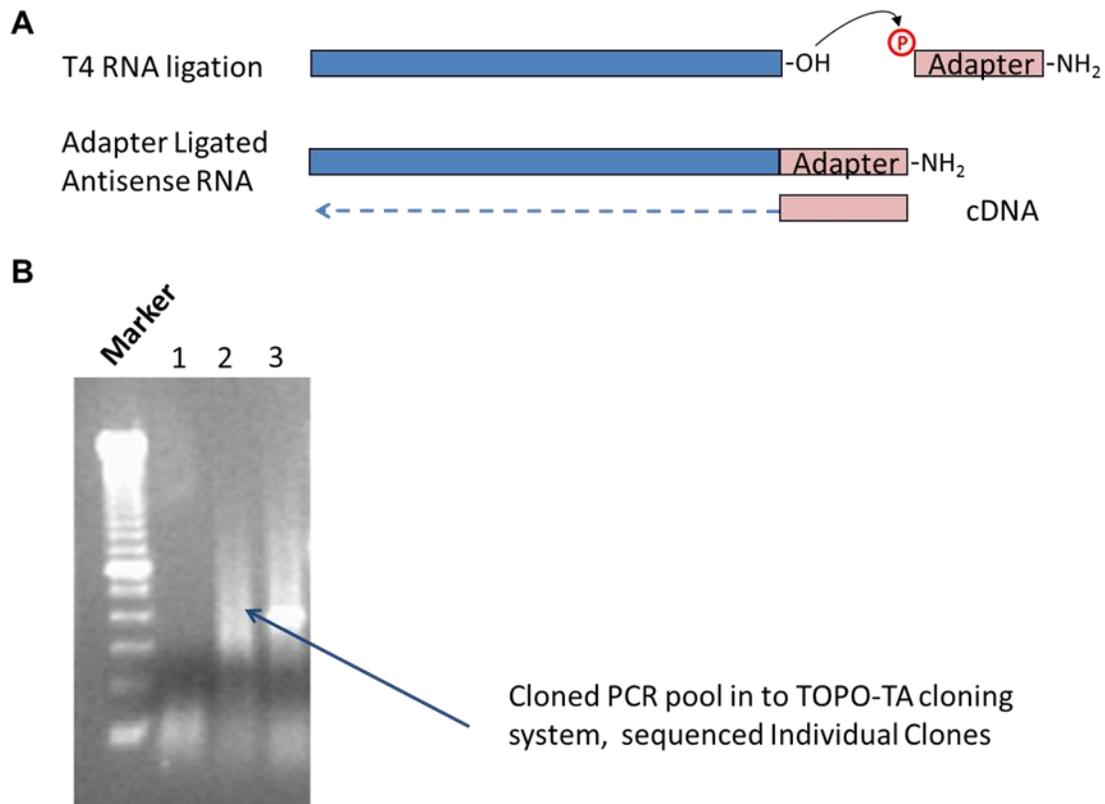


Figure 21. Pilot LA RT-PCR experiment designed to amplify antisense PR-3. (A) T4 RNA ligase can catalyze the phosphodiester bond between a free 5' phosphate donor to a 3'OH acceptor. After ligation the adapter oligo serves as universal primer binding site for reverse transcription. Because the ligation is directional, the anchor sequence can only be appended to the 3' end of the original transcript. Thus, original strand information can be inferred from the position of anchor sequence. (B) Lane 1: P15 patient leukocyte RNA sample was freeze-thawed > 5 times and used for LA RT-PCR. No amplification product must indicate low quality sample or RNA degradation. PCR Lane 2: Fresh RNA aliquot from P15 was used for LA RT-PCR. Lane 3: THP-1 cell total RNA. Reverse transcription was carried out with DT89, followed by 1st round PCR with DT89 and (-).627-PR3 primer. 2nd Round: DT89 and (-).624-PR3 primer.

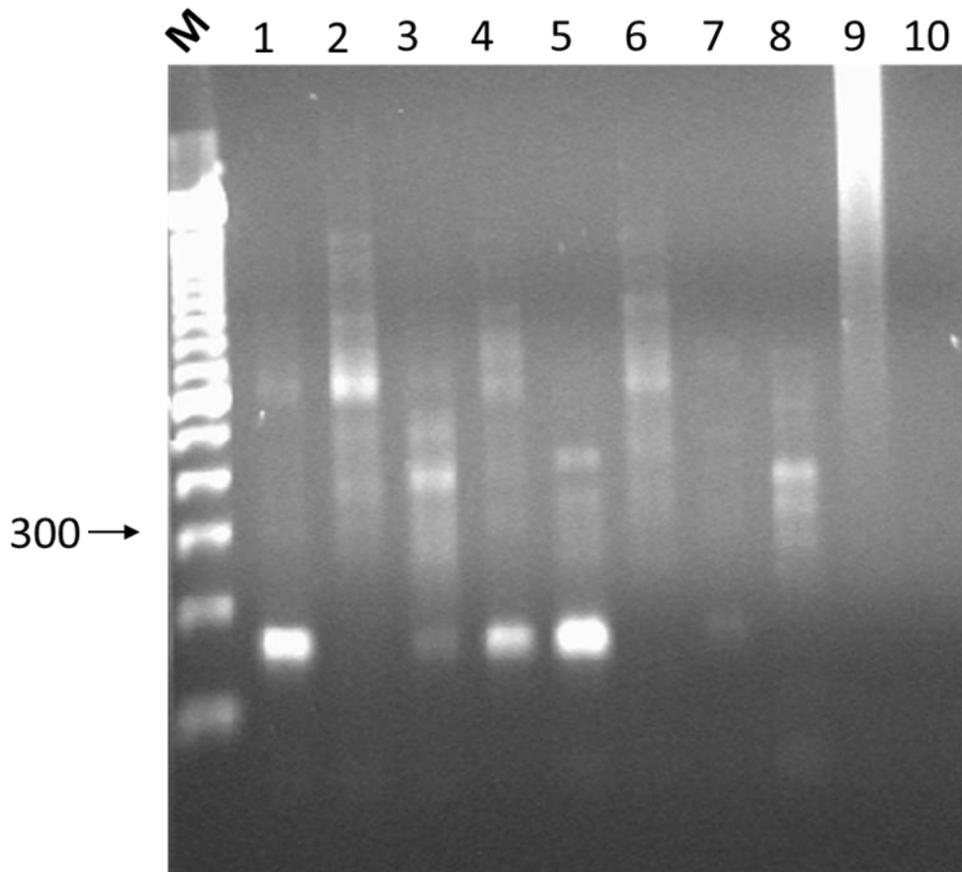


Figure 22. Sample patient screen for antisense PR-3 in PR3-ANCA patients by LA RT-PCR. A band of approximately ~180bp corresponding to AS-PR3 was detected in neutrophils from three of nine patients (Lanes: 1,4,5). Reverse transcription primer: DT89, 1st round PCR: (-).600-PR3 and DT89. 2nd Round PCR: (-).560-PR3 and DT89.

There was some evidence antisense RNA lacked a poly(A) tail. Because transcripts produced by RNA polymerase II (RNA Pol II) are generally polyadenylated, we were interested to determine whether or not antisense PR-3 could be detected in the RNA pol II complex. As described earlier for detection (RNA-pol II RIP) of the sense PR-3 molecule, RNA Pol II was cross-linked to its associated nucleic acids, and was immunopurified using an antibody specific for the transcriptionally active form of RNA Pol II. In an analysis of the PR3-ANCA patient population, we found that antisense PR-3 RNA was bound to the elongating form of RNA pol II in 4 of 10 patients (**Figure 24**). This indicates that the antisense PR-3 transcript is being actively transcribed in peripheral blood leukocytes of these four patients, bringing in to question its polyadenylation status.

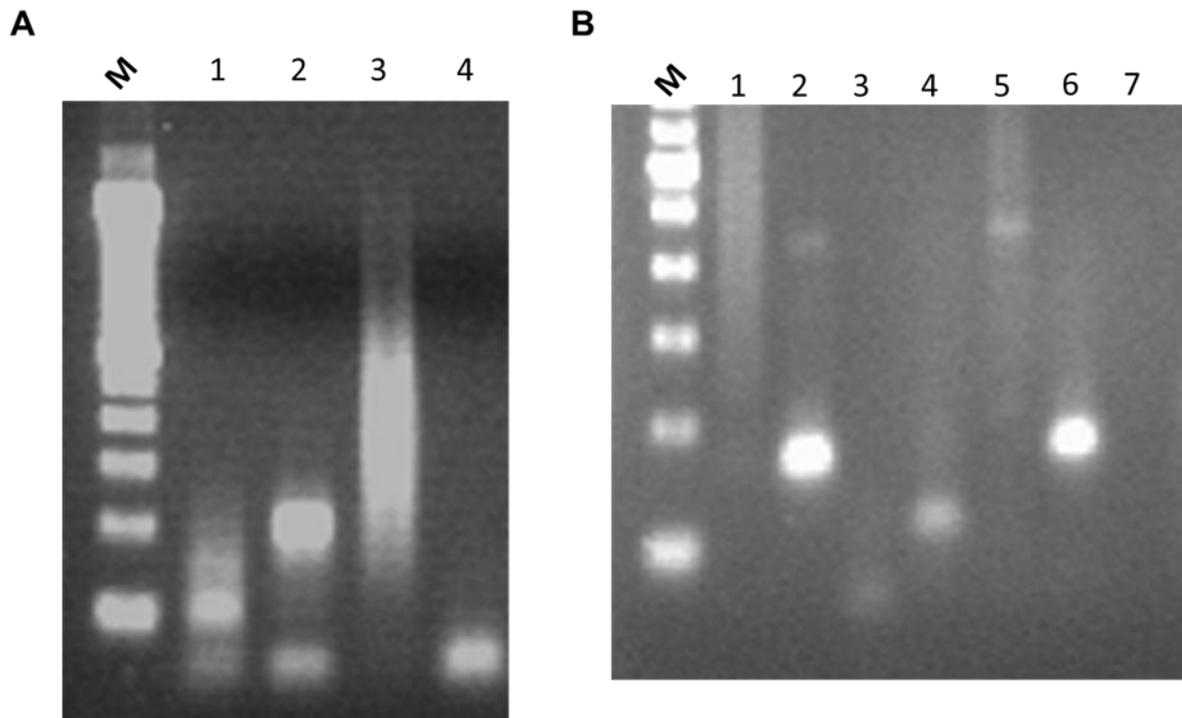


Figure 24. RIP-ChIP of antisense PR-3. (A) Lanes 1-3: RIP from PR3-ANCA patient leukocytes. Expected antisense PR-3 PCR fragment is detected only in Lane 2. Lanes 1 and 3 contained non-specific amplification of unrelated transcripts. (B) Lanes 1-4: LA RT-PCR amplification for antisense PR3 from RIP of PR3-ANCA patients. Lanes 5-6: LA RT-PCR amplification of antisense PR-3 from U937 and THP1 cells, respectively. Lane 7: H₂O control. PCR Product from lanes 2 and 6 were cloned and sequenced. Reverse transcription primer: DT89, 1st round PCR: (+).600-PR3 and DT89. 2nd Round PCR: (+).560-PR3 and DT89.

4. DISCUSSION

These studies corroborate the loss of epigenetic silencing at the *PRTN3* locus. Direct evidence is provided indicating that circulating leukocytes from ANCA patients are actively transcribing sense and antisense PR-3 transcripts. There appears to be a recurring theme in ANCA disease regarding the deleterious effects of unregulated expression of PR-3 mRNA. Our discovery of an antisense PR-3 transcript poses many interesting questions regarding gene regulation of long non-coding RNAs in disease. Unanticipated RNA processing events were detected in patients, which is presumably due to an excess or lack of appropriate *trans-acting* factors that normally process these transcripts. We were able to demonstrate that a few patients were expressing an improperly spliced form of PR-3 mRNA, which if successfully translated would contain peptide fragments that had never been displayed for negative selection in the thymus during development. Furthermore, if the PR-3 protein is expressed in mature neutrophils it is unknown whether this protein would be transported to the granules or folded properly in the absence of the appropriate cellular machinery.

Northern blotting results of sense PR-3 mRNA in patients and healthy controls also represent a shift in the initial belief that mature leukocytes do not express PR-3. A recent microarray study demonstrated that the increase in PR-3 expression in ANCA patients originated from monocytes in the patient cohort studied [29]. Patients with PR3-ANCA have been shown to have similar monocyte counts to healthy controls, thus ruling out the possibility that an increase in absolute number of monocytes is responsible for the increased PR-3 mRNA [24]. The thought that PR-3 transcription is completely repressed in the

periphery must be investigated further for its normal function in healthy individuals. It seems conceivable that during infection the neutrophils or monocytes may upregulate expression of PR-3 mRNA when stimulated with the appropriate cytokines. A series of publications have shown that increased PR-3 transcription can be induced in monocytes and PMNs from healthy donors by treatment with both TNF- α and GM-CSF [61]. We hypothesize monocytes and neutrophils in a healthy individual can express PR-3 mRNA however its expression is tightly regulated by epigenetic control.

Whether the transcriptional activity at the *PRTN3* gene locus is causal in disease is still a matter of debate. The production of aberrantly spliced transcripts may be a direct consequence of the high expression levels of SRPK-1. The imbalance of the kinase has been documented to alter ASF/SF2's activities during splicing, which in other diseases has caused the production of aberrantly spliced transcripts [62]. A direct causal link between the SR kinase and unusual PR-3 is yet to be established. Until now it was assumed that one transcript of PR-3 existed but here we document that multiple transcripts exist. These could lead to at least two different endogenously encoded proteins and moreover this could possibly lead to an array of aberrant proteins. We have detected a wide array of PR-3 mRNA isoforms in peripheral blood of patients. A larger, aberrantly spliced, transcript, which contained portions of intron 2, was found only in cells of patients with ANCA disease. A large scale screen of exon-intron usage by both targeted RNA-seq and RACE would be required to determine the extent of heterogeneity of transcripts produced in cells from patients versus healthy donors.

Determining the extent and composition of these aberrantly produced sense PR-3 transcripts expressed in patients will be challenging. After close inspection of the genomic sequences of PRTN3, we realized that large stretches of DNA contained between 67% - 80% GC content. It was empirically determined that stable RNA hairpins may have caused RT stalling or template switching during cDNA synthesis [63]. A mechanism for template switching is shown in **Figure 14**. Short direct repeats ~8 nt in length separated by structured folding can cause reverse transcriptase to temporarily disassociate from the template and re-initiate cDNA synthesis on the next repeat of the same template molecule [64]. The cDNA generated creates a deletion of the intervening region, and is thus an artifact of reverse transcription. Even if full length cDNA synthesis occurred, stable DNA hairpins formed during PCR also inhibit efficient amplification of GC-rich targets. We were able to detect what is most likely to be template switching artifacts in some of the clones sequenced. We have not ruled out the low probability that these are in fact legitimate transcripts resultant from possible deletions of fragmented pseudogenes. Although we were able to retrieve some of intron 3 sequence within this form of PR-3 mRNA, more GC-rich optimized methods which reduce secondary structure would be required to retrieve the full length transcripts.

PR-3 and myeloblastin have been considered to be the same protein since 1990. In the process of investigating disease associated PR-3 transcripts, we realized that past studies of 'normal' PR-3 transcripts were inaccurate. What is surprising is that the two forms of PR-3 transcripts had been sequenced and both were assumed to produce the same protein. We have

clear evidence that the two proteins are in fact unique isoforms of PR-3 encoded by the same gene locus. In order to limit confusion in the literature we have proposed to rename the originally discovered PR-3 as PR-3a and myeloblastin as PR-3b. This is the first documentation of both the alternative distal polyadenylation site (PRTN3-286) and exon 1 skipped forms of PR-3b. At an RNA level the 3' UTR of any mRNA contains many crucial *cis-acting* protein and microRNA binding targets for the fine tuning of both mRNA half-life and protein output. This form of PR-3 may have an entirely different biological role. The truncated form (PR-3b) would lack the signal peptide sequence, which we predict will alter its localization in specific cell compartments. PR-3b also lacks the 'PRO' peptide rendering it proteolytically active since in PR-3a these amino acids keep PR-3 catalytically inactive until it localizes the azurophilic granules. We predict that a healthy individual can express both forms of PR-3 mRNA. The implications of having two forms of PR-3 are vast and will make it difficult to tease apart which protein is responsible for the activities documented in the literature. In either case a form of PR-3 protein has been found to be overexpressed in many forms of cancer. Determining which forms are responsible for their oncogenic potential will be fruitful.

We had initially set out to find the antisense transcript responsible for producing the proposed cPR3 antigen [16]. Sequence information described here imply that the antisense transcript lacks a critical poly(A) tail required for efficient translation. We are not completely convinced of this since the potential exists that ligated fragments represent a cleavage product of the full length transcript. It was reported by Xin-zhuan Su et. al that

overwhelmingly AT rich sequences are not amplified efficiently at an extension temperature of 72°C [65]. The authors reason that at this extension temperature AT-rich duplexes are melted preventing efficient extension in PCR. Northern blotting of antisense RNA was largely inconclusive, but the results were still tantalizing. A band was visualized at roughly ~4.5 kb in the poly(A) purified RNA fractions, however a similar band was also seen in healthy controls which would not be expected if in fact the transcript coded for the hypothesized cPR3 antigen. Other than the histone mRNAs, our current understanding of translation proposes that a protein coding transcript must have a poly(A) tail to bind the *trans*-acting factors required for translation initiation.

We have not excluded the possibility that the antisense PR-3 transcript may represent a long non-coding RNA with a normal function. Azurocidin, elastase and PR-3 are organized in tandem and are coordinately regulated on chromosome 19. *HOTTIP*, a long non-coding antisense RNA in the HOX gene cluster for example, is responsible for organizing chromatin structure within the vicinity of its transcript [21]. This transcript was estimated to be expressed at extremely low levels (0.3 copies/cell) and if the antisense RNA also turns out to be a long non-coding RNA would be difficult to detect by northern blotting. If in fact the antisense PR-3 transcript is a regulatory long non coding RNA, we predict it will be responsible for the regulating expression of the PRTN3 gene locus and possibly the neighboring genes in the cluster.

Strand-specific RT-PCR remains a critical challenge even today, and is not without significant technical issues. We found that many strand-specific RT-PCR protocols still have the potential to produce results whereby the strand information would be misinterpreted due to PCR artifacts. I have developed a protocol that will side-step these potential pitfalls termed “Sodium periodate RT-PCR”. The advantages afforded in this approach lies in its ability to simultaneously reduce spurious priming of untargeted transcripts as well as maintain strand-specificity before PCR. We anticipate that periodate pre-treatment before reverse transcription will have an array of applications not limited to strand-specific RT-PCR (**Table 2**). Periodate treatment of RNA prior to RT-PCR improves current protocols by minimizing the background of unwanted cDNA synthesis (**Appendix A.1**). For example, many multiplex RT-PCR reactions are required to detect a variety of templates within the same reaction. It seems reasonable that each target could be selected prior to cDNA synthesis with gene-specific primers. cDNA templates that may have non-specifically hybridized with PCR primers would not be present in the periodate RT reaction. Other gene-specific RT-PCR reactions that would benefit from pre-treatment of RNA with sodium periodate include 5'-RLM-RACE and “classic RACE which are frequently confounded by self or spurious priming. In 5' new race one could treat RNA with sodium periodate after TAP treatment but before ligation to simultaneously prevent circularization or concatmerization of RNA as well as prevention of self-priming of other RNA in the RT reaction, which all potentially contain the adapter sequence. Oligo(dT) cDNA priming would also benefit greatly from the reduced priming of non-polyadenylated RNAs. Although transcripts with natural poly(A) tracts within their primary sequences will still be efficiently primed, these products are unavoidable

even after poly(dT) column purification or standard oligo(dT) cDNA synthesis. It is conceivable that poly(dT) purification may not be required for many cDNA libraries, or combined with periodate treatment would further increase the purity of cDNA library produced. Current methodologies for strand-specific RT-PCR could simply incorporate this method with existing protocols or simply use this method as a stand-alone technique. In fact some 'tag'-strand specific RT-PCR require a nested PCR amplification step in order to increase the specificity of PCR. We found that nested PCR amplification was not necessary for targets treated with periodate because both forward and reverse primers were unique for their corresponding templates unlike in 'tagged' strand-specific RT-PCR. Of note, it is highly probable that many researchers will find that their "gene-specific" RT-PCR reactions had in fact primarily been generated from snap-back priming and not derived from the inputted oligonucleotide. In this scenario the user may have to re-design a gene-specific RT-primer before using this approach for strand-specific RT-PCR. It should also be noted that RNA treated with sodium periodate cannot serve as a 3'OH donor for an RNA ligation. Researchers with existing strand-specific reactions could also easily incorporate a periodate pre-treatment to increase the stringency of their existing reactions.

Table 2. Potential applications of periodate treatment of RNA before cDNA synthesis

Technique	Potential advantages
Multiplex RT-PCR	Reducing the amount of spuriously primed cDNA synthesis would allow greater flexibility in primer design and compatibility of amplifying specific targets in same reaction.
5' RLM-RACE	Periodate treatment of RNA before adapter ligation to: (1) prevent mRNA circularization, mRNA:mRNA ligation (2) prevent production of unwanted cDNA with universal adapter sequence
5' and 3' classic RACE	Reduced spurious cDNA synthesis will increase likelihood of amplifying targeted cDNAs
cDNA cloning	cDNA clones will have only been generated from gene-specific primer
Strand-specific quantitative RT-PCR	(1) Completely eliminate need for high temperature RT or tagged cDNA synthesis for strand-specificity (2) Incorporate in to existing strand-specific assays to increase strand-specificity
dT ₂₀ cDNA synthesis	Eliminate priming of RNA without poly(A) tracts in RT reaction
Random hexamer cDNA synthesis	Eliminate overrepresentation of self-primed cDNA fragment
cDNA labeling with chemically modified primers	Increase efficiency of cDNA labeling
RNA-seq	Eliminate cDNA synthesis of RNA without ligated sequencing adapters

In summary this is first report of the existence of two isoforms of PR-3 protein, PR-3a and PR-3b, both encoded by the *PRTN3* gene locus. These two proteins are expected to have vastly different functions. This work documents existence of aberrant forms of PR-3 mRNA which are specifically associated with the human autoimmune disease known as ANCA disease. Initial events that incite these aberrantly spliced transcripts are yet to be defined, but they may be a result of dysregulation of normal splicing machinery. We will continue to explore the potential that an “open chromatin” conformation lends itself to transcription of antisense PR-3 transcripts. Central to the pathogenesis of autoimmunity is loss of immune tolerance to ‘self-antigens’. If in fact the aberrant PR-3 transcripts characterized in this study are translated, the proteins would contain epitopes not naturally encoded by the host. Thus these aberrant proteins have the potential to appear ‘foreign’ to the immune system and may stimulate the production of antibodies that cross-react with “self”. Lastly, the current dogma suggests that PR-3 transcription is completely null in peripheral blood leukocytes of healthy individuals; however, we have evidence suggesting that in at least some instances this is not true. Why or how these genes are expressed is an ongoing issue. The challenge moving forward is to understand how changes in the epigenetic landscape converge to shape the ANCA-disease phenotype.

5. REFERENCES

1. Van Parijs, L. and A.K. Abbas, *Homeostasis and self-tolerance in the immune system: turning lymphocytes off*. Science, 1998. **280**(5361): p. 243-8.
2. Sprent, J. and H. Kishimoto, *The thymus and central tolerance*. Transplantation, 2001. **72**(8 Suppl): p. S25-8.
3. Halverson, R., R.M. Torres, and R. Pelandra, *Receptor editing is the main mechanism of B cell tolerance toward membrane antigens*. Nat Immunol, 2004. **5**(6): p. 645-50.
4. Quill, H., *Anergy as a mechanism of peripheral T cell tolerance*. J Immunol, 1996. **156**(4): p. 1325-7.
5. Chertov, O., et al., *Identification of human neutrophil-derived cathepsin G and azurocidin/CAP37 as chemoattractants for mononuclear cells and neutrophils*. J Exp Med, 1997. **186**(5): p. 739-47.
6. Bennouna, S., et al., *Cross-talk in the innate immune system: neutrophils instruct recruitment and activation of dendritic cells during microbial infection*. J Immunol, 2003. **171**(11): p. 6052-8.
7. Cowland, J.B. and N. Borregaard, *The individual regulation of granule protein mRNA levels during neutrophil maturation explains the heterogeneity of neutrophil granules*. J Leukoc Biol, 1999. **66**(6): p. 989-95.
8. Fouret, P., et al., *Expression of the neutrophil elastase gene during human bone marrow cell differentiation*. J Exp Med, 1989. **169**(3): p. 833-45.
9. Falk, R.J. and J.C. Jennette, *ANCA small-vessel vasculitis*. J Am Soc Nephrol, 1997. **8**(2): p. 314-22.

10. Jennette, J.C. and R.J. Falk, *Small-vessel vasculitis*. N Engl J Med, 1997. **337**(21): p. 1512-23.
11. Jennette, J.C., J.R. Hoidal, and R.J. Falk, *Specificity of anti-neutrophil cytoplasmic autoantibodies for proteinase 3*. Blood, 1990. **75**(11): p. 2263-4.
12. Falk, R.J. and J.C. Jennette, *Anti-neutrophil cytoplasmic autoantibodies with specificity for myeloperoxidase in patients with systemic vasculitis and idiopathic necrotizing and crescentic glomerulonephritis*. N Engl J Med, 1988. **318**(25): p. 1651-7.
13. Falk, R.J., et al., *Anti-neutrophil cytoplasmic autoantibodies induce neutrophils to degranulate and produce oxygen radicals in vitro*. Proc Natl Acad Sci U S A, 1990. **87**(11): p. 4115-9.
14. Mekler, L.B., *On the specific mutual interaction of amino acid residues of polypeptide chains and amino acid residues with codons*. Oncology, 1973. **27**(3): p. 286-8.
15. Heal, J.R., et al., *Specific interactions between sense and complementary peptides: the basis for the proteomic code*. Chembiochem, 2002. **3**(2-3): p. 136-51.
16. Pendergraft, W.F., 3rd, et al., *Autoimmunity is triggered by cPR-3(105-201), a protein complementary to human autoantigen proteinase-3*. Nat Med, 2004. **10**(1): p. 72-9.
17. Xiao, H., et al., *Antineutrophil cytoplasmic autoantibodies specific for myeloperoxidase cause glomerulonephritis and vasculitis in mice*. J Clin Invest, 2002. **110**(7): p. 955-63.
18. Charles, L.A., et al., *Antibodies against granule proteins activate neutrophils in vitro*. J Leukoc Biol, 1991. **50**(6): p. 539-46.
19. Hussain, A., et al., *Chimeric IgG4 PR3-ANCA induces selective inflammatory responses from neutrophils through engagement of Fcgamma receptors*. Immunology, 2009. **128**(2): p. 236-44.

20. Labbaye, C., P. Musette, and Y.E. Cayre, *Wegener autoantigen and myeloblastin are encoded by a single mRNA*. Proc Natl Acad Sci U S A, 1991. **88**(20): p. 9253-6.
21. Labbaye, C., et al., *Regulation of myeloblastin messenger RNA expression in myeloid leukemia cells treated with all-trans retinoic acid*. Blood, 1993. **81**(2): p. 475-81.
22. Bories, D., et al., *Down-regulation of a serine protease, myeloblastin, causes growth arrest and differentiation of promyelocytic leukemia cells*. Cell, 1989. **59**(6): p. 959-68.
23. Jenne, D.E., *Structure of the azurocidin, proteinase 3, and neutrophil elastase genes. Implications for inflammation and vasculitis*. Am J Respir Crit Care Med, 1994. **150**(6 Pt 2): p. S147-54.
24. Yang, J.J., et al., *Circumvention of normal constraints on granule protein gene expression in peripheral blood neutrophils and monocytes of patients with antineutrophil cytoplasmic autoantibody-associated glomerulonephritis*. J Am Soc Nephrol, 2004. **15**(8): p. 2103-14.
25. Zimmerman, G.A., S.M. Prescott, and T.M. McIntyre, *Endothelial cell interactions with granulocytes: tethering and signaling molecules*. Immunol Today, 1992. **13**(3): p. 93-100.
26. Campanelli, D., et al., *Azurocidin and a homologous serine protease from neutrophils. Differential antimicrobial and proteolytic properties*. J Clin Invest, 1990. **85**(3): p. 904-15.
27. Csernok, E., et al., *Activated neutrophils express proteinase 3 on their plasma membrane in vitro and in vivo*. Clin Exp Immunol, 1994. **95**(2): p. 244-50.
28. Kao, R.C., et al., *Proteinase 3. A distinct human polymorphonuclear leukocyte proteinase that produces emphysema in hamsters*. J Clin Invest, 1988. **82**(6): p. 1963-73.
29. Ohlsson, S., et al., *Increased monocyte transcription of the proteinase 3 gene in small vessel vasculitis*. Clin Exp Immunol, 2005. **141**(1): p. 174-82.
30. Ciavatta, D.J., et al., *Epigenetic basis for aberrant upregulation of autoantigen genes in humans with ANCA vasculitis*. J Clin Invest, 2010. **120**(9): p. 3209-19.

31. Kuzmichev, A., et al., *Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein*. Genes Dev, 2002. **16**(22): p. 2893-905.
32. Alcorta, D.A., et al., *Leukocyte gene expression signatures in antineutrophil cytoplasmic autoantibody and lupus glomerulonephritis*. Kidney Int, 2007. **72**(7): p. 853-64.
33. Zahler, A.M., et al., *SR proteins: a conserved family of pre-mRNA splicing factors*. Genes Dev, 1992. **6**(5): p. 837-47.
34. Ngo, J.C., et al., *A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1*. Mol Cell, 2008. **29**(5): p. 563-76.
35. Hagopian, J.C., et al., *Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1*. J Mol Biol, 2008. **382**(4): p. 894-909.
36. Cao, W., S.F. Jamison, and M.A. Garcia-Blanco, *Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro*. RNA, 1997. **3**(12): p. 1456-67.
37. Zuo, P. and J.L. Manley, *Functional domains of the human splicing factor ASF/SF2*. EMBO J, 1993. **12**(12): p. 4727-37.
38. Yelin, R., et al., *Widespread occurrence of antisense transcription in the human genome*. Nat Biotechnol, 2003. **21**(4): p. 379-86.
39. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions*. Nat Rev Genet, 2009. **10**(3): p. 155-9.
40. Kumar, M. and G.G. Carmichael, *Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes*. Microbiol Mol Biol Rev, 1998. **62**(4): p. 1415-34.
41. Huttenhofer, A. and J. Vogel, *Experimental approaches to identify non-coding RNAs*. Nucleic Acids Res, 2006. **34**(2): p. 635-46.

42. Carninci, P., et al., *The transcriptional landscape of the mammalian genome*. Science, 2005. **309**(5740): p. 1559-63.
43. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription*. Science, 2007. **316**(5830): p. 1484-8.
44. Dinger, M.E., et al., *Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications*. Brief Funct Genomic Proteomic, 2009. **8**(6): p. 407-23.
45. Amaral, P.P. and J.S. Mattick, *Noncoding RNA in development*. Mamm Genome, 2008. **19**(7-8): p. 454-92.
46. Woo, C.J. and R.E. Kingston, *HOTAIR lifts noncoding RNAs to new levels*. Cell, 2007. **129**(7): p. 1257-9.
47. Yu, W., et al., *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA*. Nature, 2008. **451**(7175): p. 202-6.
48. Burd, C.E., et al., *Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk*. PLoS Genet, 2010. **6**(12): p. e1001233.
49. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a*. Mol Cell, 2010. **38**(5): p. 662-74.
50. Lindberg, J. and J. Lundeberg, *The plasticity of the mammalian transcriptome*. Genomics, 2010. **95**(1): p. 1-6.
51. Haddad, F., et al., *Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR*. BMC Biotechnol, 2007. **7**: p. 21.
52. Shen, M.R., J. Brosius, and P.L. Deininger, *BC1 RNA, the transcript from a master gene for ID element amplification, is able to prime its own reverse transcription*. Nucleic Acids Res, 1997. **25**(8): p. 1641-8.

53. Zhao, Y., et al., *Antisense transcripts of V(D)J rearrangements; artifacts caused by false priming?* Mol Immunol, 2009. **46**(11-12): p. 2357-62.
54. Tuiskunen, A., et al., *Self-priming of reverse transcriptase impairs strand-specific detection of dengue virus RNA.* J Gen Virol, 2010. **91**(Pt 4): p. 1019-27.
55. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers.* Methods Mol Biol, 2000. **132**: p. 365-86.
56. Don, R.H., et al., *'Touchdown' PCR to circumvent spurious priming during gene amplification.* Nucleic Acids Res, 1991. **19**(14): p. 4008.
57. Scotto-Lavino, E., G. Du, and M.A. Frohman, *3' end cDNA amplification using classic RACE.* Nat Protoc, 2006. **1**(6): p. 2742-5.
58. Ciavatta, D.J., et al., *Epigenetic basis for aberrant upregulation of autoantigen genes in humans with ANCA vasculitis.* J Clin Invest. **120**(9): p. 3209-19.
59. Craggs, J.K., et al., *Development of a strand-specific RT-PCR based assay to detect the replicative form of hepatitis C virus RNA.* J Virol Methods, 2001. **94**(1-2): p. 111-20.
60. Troutt, A.B., et al., *Ligation-anchored PCR: a simple amplification technique with single-sided specificity.* Proc Natl Acad Sci U S A, 1992. **89**(20): p. 9823-5.
61. Zhou, Z., C. Richard, and H.A. Menard, *De novo synthesis of proteinase 3 by cytokine primed circulating human polymorphonuclear neutrophils and mononuclear cells.* J Rheumatol, 2000. **27**(10): p. 2406-11.
62. Nowak, D.G., et al., *Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis.* J Biol Chem, 2010. **285**(8): p. 5532-40.
63. Cocquet, J., et al., *Reverse transcriptase template switching and false alternative transcripts.* Genomics, 2006. **88**(1): p. 127-31.

64. Roy, S.W. and M. Irimia, *When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis*. *Bioessays*, 2008. **30**(6): p. 601-5.
65. Su, X.Z., et al., *Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA*. *Nucleic Acids Res*, 1996. **24**(8): p. 1574-5.

APPENDIX

Appendix A

A.1. Treatment of RNA with sodium periodate before reverse transcription eliminates spurious priming of untargeted transcripts while retaining strand-specificity.

Non-specific cDNA synthesis produced by endogenous priming presents a frequently encountered problem in the more specialized reverse transcriptase based protocols, including, but not limited to RACE, cDNA library generation, and quantitative RT-PCR. In spite of the advances made in the techniques described above, many elaborate optimizations have focused on steps during or after cDNA synthesis. Thus a reverse transcription reaction without unwanted background cDNA synthesis would unquestionably improve upon many current protocols that require an RT step before downstream analyses.

It became apparent that current strand-specific methodologies were inadequate for the reliable distinction between sense and antisense RNA molecules. I designed a novel approach that utilized sodium periodate to block the 3'OH of the RNA molecule which is required for 'endogenous' priming. Sodium periodate oxidizes the *cis*-diol at the 2'-3' OH and the second *cis*-diol in the 7-methylguanosine leaving the remainder of the RNA molecule unaffected (**Figure 25a**). By selectively oxidizing the terminal 3'-OH only the inputted oligonucleotides drive cDNA synthesis and not priming from endogenous RNAs (**Figure 25b**). For cDNA synthesis to occur the user-derived oligo must hybridize to its RNA target for proper recognition by reverse transcriptase. Based upon the oligo chosen, either polarity RNA can be specifically targeted without self-priming of the opposite strand. Favorably, only cDNA generated from the targeted orientation is amplified in the subsequent PCR reaction.

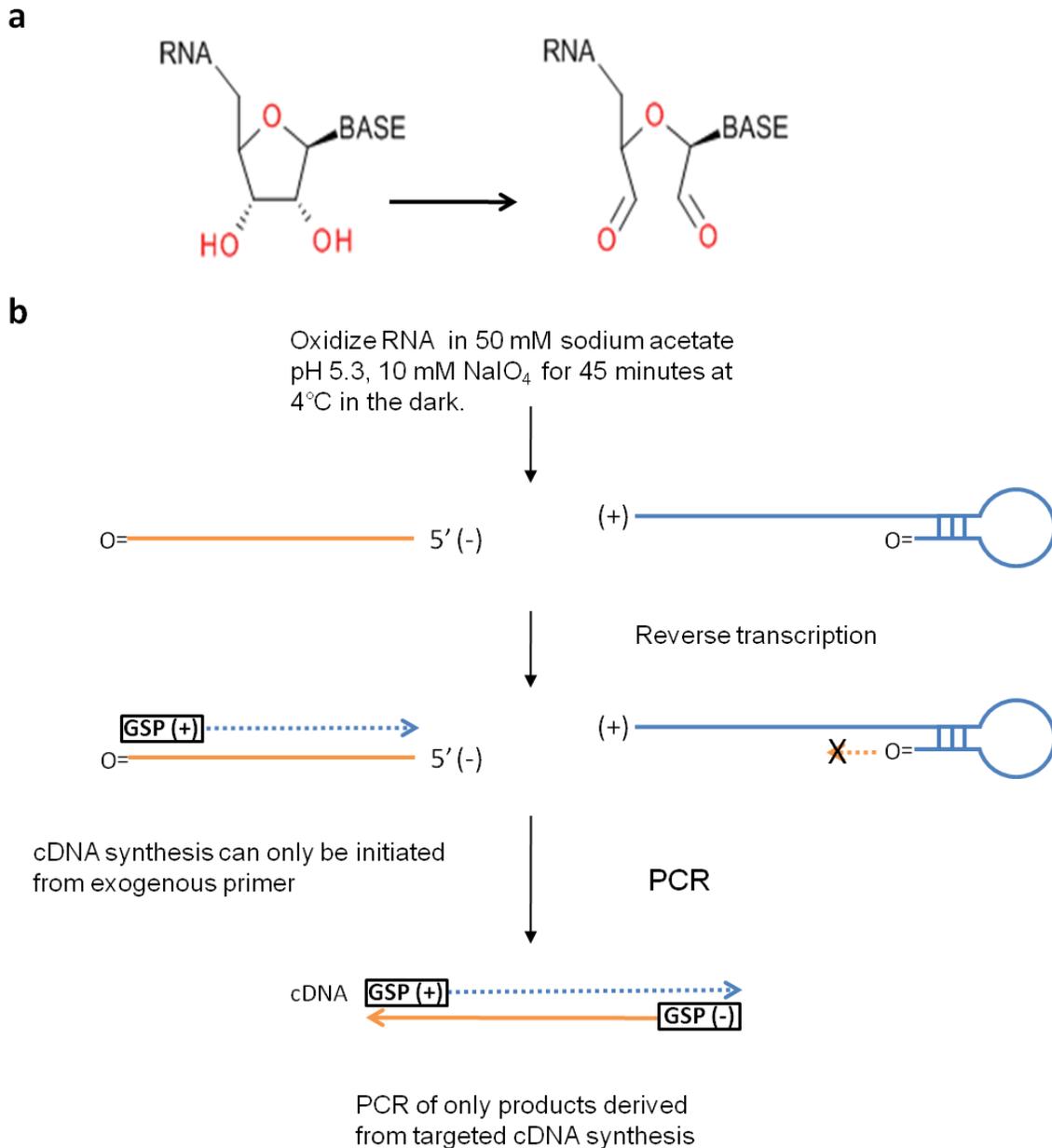


Figure 25. (A) Periodate oxidation of vicinal diol at the terminal nucleotide of RNA. The 7-methyl guanosine also contains a *cis*-diol which is also oxidized, that will not interfere with cDNA synthesis. **(B)** All RNA with compatible 3'-OH groups is oxidized before reverse transcription to prevent spurious initiation of cDNA synthesis. cDNA from untargeted orientation is not produced and thus PCR can be carried out with primers for the intended strand.

The following experiments document that pre-treatment of RNA with sodium periodate benefits many molecular biology techniques (**Table 3**). Briefly, total RNA isolated from the THP-1 human monocytic leukemic cell line (high expression for PR-3) was treated with 10 mM freshly prepared NaIO₄ in the dark on ice for 45 minutes, quenched with glycerol solution and desalted before RT-PCR. Without periodate treatment sense PR-3 is efficiently primed in the absence of inputted primers in the RT reaction (**Figure 26**, Lanes 6 and 7) or with a non-specific RT primer (**Figure 26**, Lanes 10 and 11). Periodate treatment effectively blocked spurious snap-back priming of the sense transcript (**Figure 26**, Lanes 3 and 4). Conversely, when a gene-specific primer targeted for sense PR-3 mRNA was used, a specific band corresponding to PR-3 was observed (**Figure 26**, Lanes 1 and 2). An important point is that this was accomplished without a reduction in sensitivity.

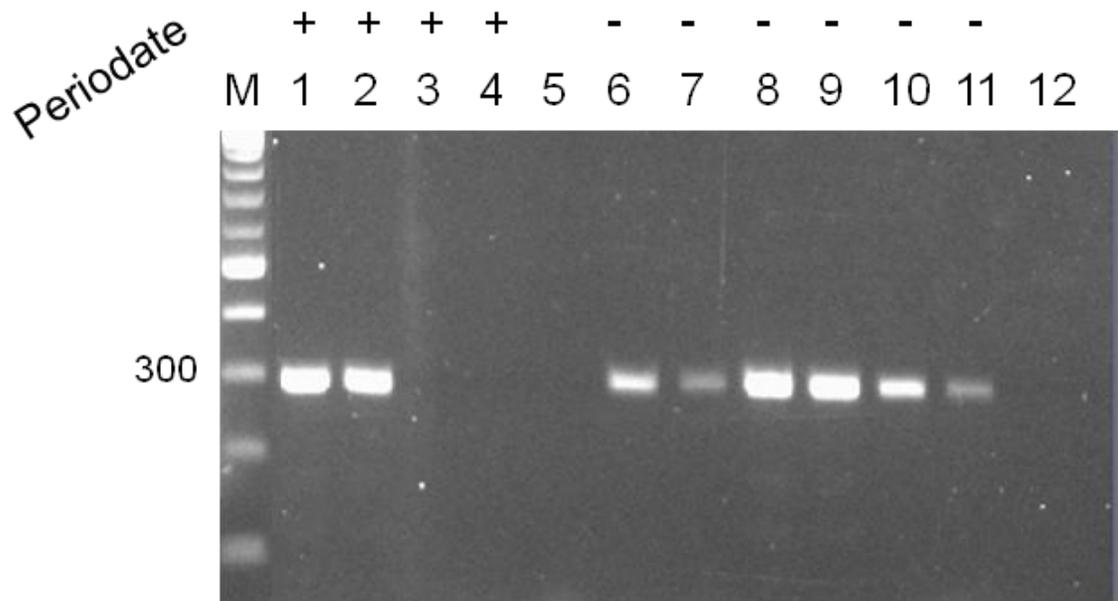


Figure 26. Periodate oxidation before RT blocks non-specific cDNA synthesis.

Treatments and primers used for experiments are outlined in Table 1. Total RNA from THP-1 cell line was isolated and treated with and without sodium periodate. Lanes 1-2, gene-specific priming of sense PR-3 followed by PCR. Lanes 3-4, when RT primer is omitted and RNA is oxidized sense PR-3 cannot be amplified by PCR. Lane 5, water control. Lanes 6-7, without periodate treatment sense PR-3 self-primers its cDNA synthesis and can be amplified by PCR without addition of RT primer. Lane 8-9, gene-specific cDNA synthesis of sense PR-3 without periodate treatment, yields a specific band for sense PR-3. Lanes 10-11 a gene-specific primer for an antisense PR-3 Lane 12, Water control.

Table 3. Periodate treatments of samples shown in Figure 26.

Lane	RNA (μg)	Treatment	cDNA synthesis primer	Spurious cDNA synthesis
1	1.0 μg	10 mM NaIO_4	(+)627.PR3	N/A
2	0.2 μg	10 mM NaIO_4	(+)627.PR3	N/A
3	1.0 μg	10 mM NaIO_4	None	Undetected
4	0.2 μg	10 mM NaIO_4	None	Undetected
5	H_2O		(+)627.PR3	N/A
6	1.0 μg	Mock	None	Detected
7	0.2 μg	Mock	None	Detected
8	1.0 μg	Mock	(+)627.PR3	N/A
9	0.2 μg	Mock	(+)627.PR3	N/A
10	1.0 μg	Mock	Non-specific primer	Detected
11	0.2 μg	Mock	Non-specific primer	Detected
12	H_2O			