

A Bioinformatics Analysis of Protein Tyrosine Phosphatases in Humans

T. K. B. GANDHI,¹ Sreenath CHANDRAN,¹ Suraj PERI,^{2,3} R. SARAVANA,¹ Ramars AMANCHY,²
T. S. Keshava PRASAD,¹ and Akhilesh PANDEY^{2,*}

Institute of Bioinformatics, Discoverer 7th Floor, International Technology Park Ltd, Bangalore 560 066, India,¹ McKusick-Nathans Institute of Genetic Medicine and Department of Biological Chemistry and Oncology, Johns Hopkins University, Baltimore, MD 21205, USA,² and Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark³

(Received 29 January 2005; revised 23 March 2005)

Abstract

Protein tyrosine phosphatases (PTPs) cooperate with protein tyrosine kinases to regulate signal transduction pathways. Genome-wide surveys cataloging protein tyrosine phosphatases in humans have recently been carried out. Here, we present a bioinformatics analysis of protein tyrosine phosphatases in the human genome to examine their domain architecture, alternative splicing and pseudogenes. We present evidence that alternative transcripts exist for 25 out of 35 PTPs analyzed. These alternative transcripts include novel exons; skipped exons as well as cryptic donor/acceptor splice sites. We discovered a novel isoform of PTPN18 based on analysis of expressed sequence tags (ESTs). The deletion of 4 exons in the catalytic domain of the novel isoform may alter the enzymatic activity toward its substrates. We were able to experimentally validate 2 of our novel isoform predictions through RT-PCR. Finally, a user-friendly web-based resource that consolidates the gene and protein annotations for all human protein tyrosine phosphatases has been developed and is freely available at <http://ptpr.ibioinformatics.org>.

Key words: Signal transduction; Genomics; Tyrosine phosphorylation; Alternative splicing; Comparative genomics

1. Introduction

Protein tyrosine phosphatases (PTPs) are responsible for the dephosphorylation of tyrosine-phosphorylated proteins and regulate a myriad of cellular processes. PTPs are characterized by the presence of an active site signature motif, (H/V)CX₅R(S/T), in the conserved catalytic PTPase domain, where X can be any amino acid.¹ They are classified on the basis of their subcellular localization into receptor and non-receptor tyrosine phosphatases. The non-receptor phosphatases contain a single catalytic PTPase domain whereas the receptor tyrosine phosphatases (RPTPs) often possess two PTPase domains. The N-terminal PTPase domain in receptor PTPs is involved in the catalytic activity while the role of C-terminal PTPase domain is not clearly understood. The ligand-induced dimerization of RPTPs may result in inhibition of phosphatase activity.²

A preliminary survey based on the initial sequencing of the human genome suggested the presence of 112 PTPs

including both the tyrosine-specific and dual-specificity phosphatases.³ Based on their unique substrate specificity, 38 genes are considered as tyrosine specific or classical PTPs⁴ which belong to the Class I Cys-based protein tyrosine phosphatases.⁵ Anderson and coworkers⁴ have also identified 12 PTP pseudogenes. Bhaduri and Sowdhamini⁶ studied the common and uncommon domain architectures of tyrosine phosphatases. Although Anderson et al., described 21 receptor tyrosine phosphatases (RPTPs) and 17 non-receptor tyrosine phosphatases in humans, we considered only 35 PTPs for our analysis as the remaining three (PTPRQ, PTPRV and PTPN20) exist only as predicted entries. However, we have included all 38 PTPs for domain analysis. Alternative splicing is a post-transcriptional process where exons from pre-mRNAs are shuffled leading to formation of diverse mature mRNAs from the same gene which ultimately could result in protein isoforms.⁷ Genome-wide analyses reported previously estimate that at least 50% of human genes undergo alternative splicing.⁸ Alternative splicing has already been shown in some members of the PTP family including PTPRA, PTPRF and PTPRC.^{9–11} In the present study, a systematic analysis

Communicated by Osamu Ohara

* To whom correspondence should be addressed. Tel. +1-410-502-6662, Fax. +1-410-502-7544, E-mail: pandey@jhmi.edu

to identify novel alternative transcript variants for PTPs was undertaken using EST evidence and comparative genomics. We have detected alternatively spliced variants for 25 (71%) of the 35 PTP genes analyzed; we experimentally validated 2 of our predictions using RT-PCR. Most of these splicing events affected coding regions. A novel isoform of PTPN18 was identified and designated as PTPN18 β . The splicing events in the exons that code for PTPase domains resulted in a variety of scenarios where the PTPase domain was altered by the addition or deletion of exons. In one instance, deletion of the linker region between the N- and C-terminal phosphatase domains in PTPRT resulted in fusion of the two domains.

A user-friendly database containing the results from the present study has been developed as a new approach for a family of proteins with featured alternative splicing and pseudogene information pertaining to a gene. We believe that this resource will be helpful for the scientific community to quickly access the data regarding PTPs in a single resource.

2. Materials and Methods

2.1. Domain analysis of protein tyrosine phosphatases

Domain analysis of the PTPs was done using SMART¹² and Pfam¹³ tools. The PTPs were grouped on the basis of presence, position and number of domains.

2.2. Analysis of alternative splicing

Analysis of alternatively spliced transcripts was performed using sequence information from genomic DNA, mRNA and ESTs. We identified the alternative splice forms of PTPs based on EST evidence and by comparing the human genomic sequence with that of the rat and mouse.

2.2.1. Analysis based on human EST evidence

cDNAs of the longest isoforms for all known PTPs were obtained from the RefSeq database. The exonic regions of these isoforms were obtained by aligning the cDNA sequences against the genomic DNA sequence using SIM4 alignment program.¹⁴ The cDNAs were also searched against human dbESTs using BLASTN. Only alignments having an identity of 98% and above were selected for further analysis. Gaps greater than 10 bp in the alignments were analyzed for evidence of potential alternate splicing. A gap in the query sequence (representing the cDNA) signifies an exon or exon extension. Similarly, a gap in the subject sequence (representing human ESTs) signifies exon omission in the EST, which could be either known or novel. ESTs short-listed by this process were further analyzed for the presence of known splice sites at the exon-intron boundaries.

2.2.2. Analysis through comparative genomics

The gene structure of every tyrosine phosphatase was compared among mouse, rat and human genomic sequences. As exonic regions, especially coding regions, are conserved, we searched for highly conserved regions with protein coding potential. Such regions were subsequently analyzed for evidence of transcription in mouse and rat. The corresponding human genomic sequence, when present, was retrieved and analyzed for the presence of splice signals necessary to establish exon boundaries. Regions fulfilling these criteria were labeled as potential novel exons.

2.2.3. Amplification of novel exons of PTPRM and PTPRT

Amplification of the novel exons of PTPRM and PTPRT was carried out using exon-spanning primers. This was done from 2 μ g of cDNA (Invitrogen, Carlsbad, CA, USA) using Taq DNA Polymerase (Bangalore Genei, Bangalore, India) and PCR conditions of 94°C for 6 min, and 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min. A final extension was carried out at 72°C for 10 min, and the RT-PCR product was visualized by electrophoresis using a 1.2% agarose gel containing ethidium bromide. Fragments containing the novel exon of PTPRM were amplified using the forward primer GGTGTATGACAGACCCATTTGTGC and the reverse primer CTGCCCTTCAAAGAAGCTCTCGTA to detect a fragment of 307 bp. Alternatively, a forward primer GAACCCGAGAAACAGACAGACCAT and reverse primer TTGCAGTTGGCACAAATGGGTC was used to obtain a fragment of 422 bp. The forward primer was in the novel exon in the former attempt, and the reverse primer was in the novel exon of PTPRM in the latter attempt. In the case of PTPRT, the novel exonic region was amplified by using a forward primer AGCGGAACACACTGTGGGAAAT and a reverse primer CATGTCAAGCATGGTGTCAATGGC to detect a 387-bp fragment.

2.3. Identification of PTP pseudogenes

The human genomic sequence from National Center for Biotechnology Information (NCBI) build 35 was used for identification of pseudogenes. The protein and cDNA sequences from NCBI's RefSeq were used for the analysis. Genomic BLAST analysis was carried out using BLASTN by aligning PTP cDNA sequence against the human genome and default settings were used without the MegaBLAST option. Since an active functional gene should generally be localized to a unique site in the genome, one would expect a single hit. Therefore, multiple hits obtained from BLAST analysis were further analyzed for the identification of candidate pseudogenes. The sequences in the genome that aligned perfectly with their corresponding cDNA sequences, without any breaks

corresponding to introns, were regarded as ‘processed pseudogenes.’ Likewise, protein sequences were also aligned against the genome using the TBLASTN tool. Regions with significant homology ($E < 1.0e^{-5}$) were selected for further analysis. Paralogs were identified based on sequence identity and chromosomal map positions and were discarded. These alignments were further analyzed according to the following criteria that define a pseudogene: (1) A genomic sequence that is uninterrupted by introns, (2) the presence of stop codons or (3) insertion or deletion events.

2.4. Creation of a web-based PTP resource

A web-based resource was developed using an object-oriented framework, Zope Object Database (ZODB) integrated with a Zope server. Zope Page Templates (ZPT) and Document Type Markup Language (DTML) were used to create the necessary HTML pages. The Protein Tyrosine Phosphatase Resource (PTPR) database enables users to easily retrieve the details regarding the molecular information, pseudogene and alternative splicing of individual protein tyrosine phosphatases taken up in the present study. Each protein in the PTPR database is also hyperlinked to the Human Protein Reference Database (HPRD),¹⁵ which contains manual annotation of these proteins.

3. Results and Discussion

3.1. Domain architecture of protein tyrosine phosphatases

There are 38 known PTPs, which are subdivided into receptor and non-receptor PTPs. The non-receptor phosphatases contain a single catalytic PTPase domain that is found in combination with PEST, SEC14, SH2, SH3, PDZ, B41, KIND and BRO1 domains (Fig. 1A). The receptor tyrosine phosphatases (RPTPs) are localized to the plasma membrane and have one or two PTPase domains and a characteristic transmembrane domain (Fig. 1B). Out of the 21 RPTPs, 12 proteins possess two PTPase domains, whereas the rest contain only one such domain.

We categorized PTPs into 20 groups based on the presence, number and position of each domain found. For example, PTPN5, PTPN7 and PTPN20 are grouped together because they contain one PTPase domain at their C-terminus. PTPN1 and PTPN2 and also have one PTPase domain but at the N-terminus of the protein and are hence considered different from the former group (Fig. 1A). Similarly, PTPRJ and PTPRB (Fig. 1B) have one C-terminal PTPase domain preceded by a transmembrane domain, however, each protein was grouped differently as the former has 9 fibronectin type 3 (FN3) repeats whereas the latter has 17.

The modular architecture of phosphatases allows them to act as components of large protein complexes. We analyzed all PTPs for the presence of any protein domains that were not described earlier.⁶ The Sec14 domain in PTPN9 is responsible for its localization to the perinuclear region.²³ The B41 domain is found to be present in PTPN3, PTPN4, PTPN13, PTPN14 and PTPN21. B41 domains are proven to be responsible for the targeting of cytoskeletal proteins to the membrane-cytoskeleton interfaces, and the presence of B41 domains are thought to explain the role of PTPs in the regulation of cytoskeletal structure and cell adhesion.²⁴ Finally, we found a KIND domain in PTPN13, the presence of which could help in substrate selection by the enzyme.

3.2. Investigation of alternative splicing of tyrosine phosphatases

Among the 35 PTPs analyzed in the present study, novel splicing events were detected in 25 genes (Table 1). Detection of these splicing events resulted in identification of a hitherto unreported novel isoform PTP18B, 25 novel exons among 14 PTPs, 19 skipped exons among 12 PTPs and 24 cryptic splice donor and acceptor sites among 15 PTPs.

3.2.1. Identification of a novel isoform of PTPN18

PTPN18 is a non-receptor protein tyrosine phosphatase, which was isolated from a human fetal brain cDNA library.¹⁶ PTPN18 has been shown to be expressed in several tumor-derived cell lines as well.²⁵ A novel isoform of PTPN18 was found during our analysis. PTPN18 comprises 15 exons and codes for a 459-amino acid protein. An EST (BQ061600) was found that codes for a full-length isoform of PTPN18 containing an intact PTPase domain. As shown in Fig. 2, we have designated this novel transcript as PTP18 β (PTP18B); it encodes a smaller protein of 314 amino acids due to an internal deletion of 4 exons. We suggest that the longest isoform of PTPN18 (RefSeq: NP_055184) henceforth be designated as PTPN18 α . In addition, we found two other novel splice variants for this gene (BE616252 and B1858031).

3.2.2. Identification of novel exons

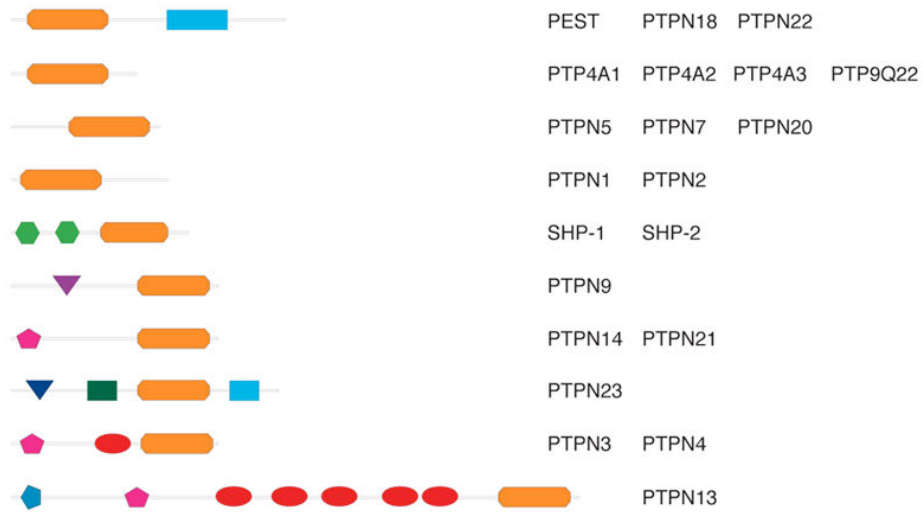
An exon absent in the characterized isoforms may be present in an alternative isoform as a novel or cryptic exon. We have used two different approaches for the identification of novel exons. One is through EST analysis and the other is comparative genomics using mouse and rat sequences (see Materials and Methods).

3.2.2.1 Identification of novel exons through EST analysis

A total of 24 novel exons were found through EST analysis. For example, two novel exons were identified in PTPRZ, one between exons 15 and 16 (AL534804)

A

Non-receptor tyrosine phosphatases



B

Receptor tyrosine phosphatases

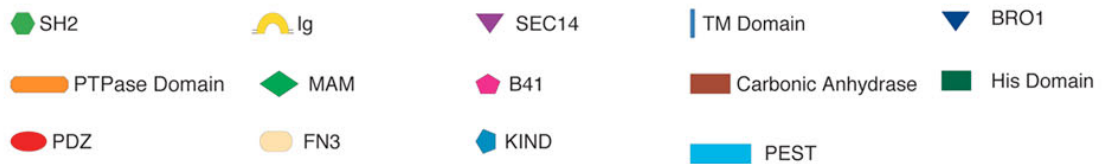
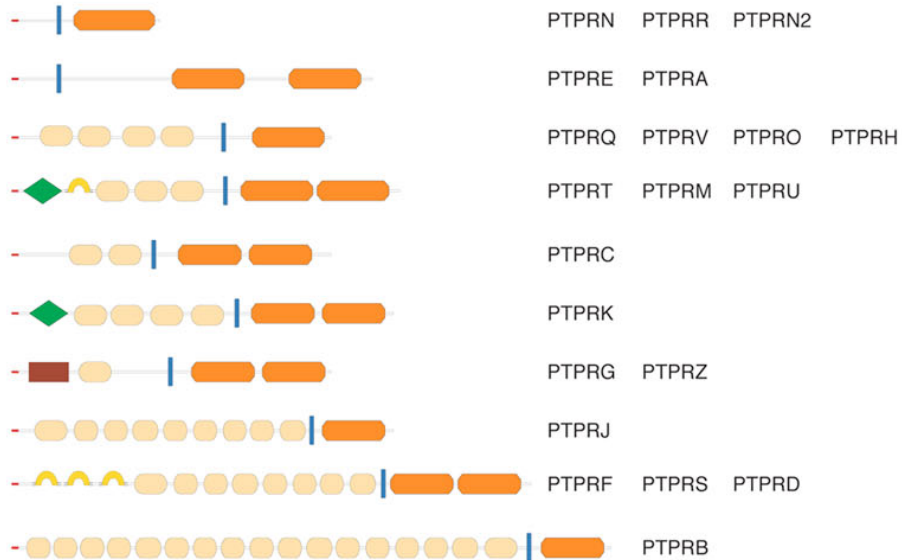
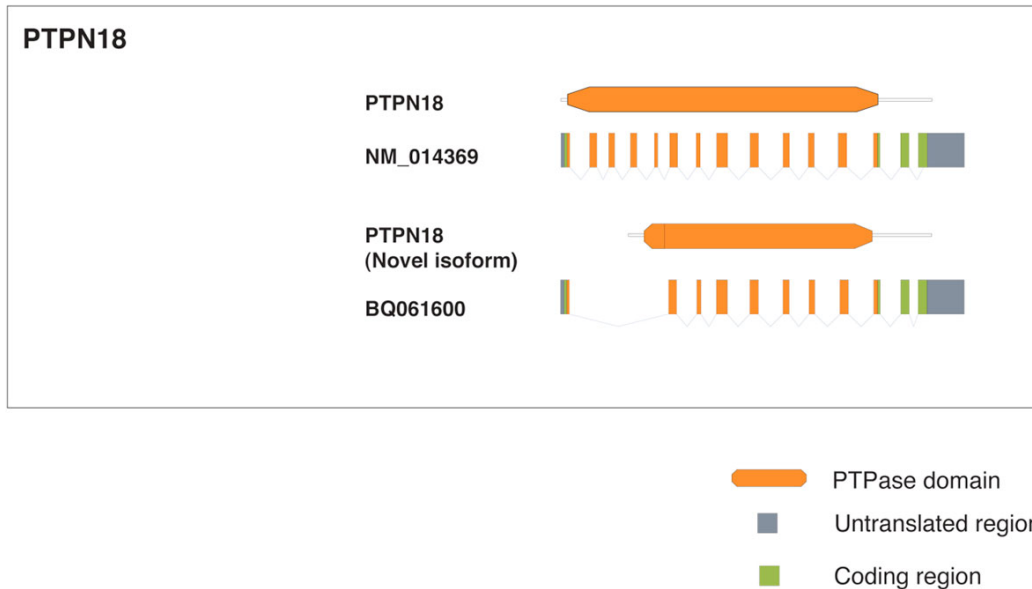


Figure 1. Domain architecture of protein tyrosine phosphatases. The domains are represented schematically with different shapes and colors as indicated. The proteins are grouped based on the presence, number and position of each domain. (A): Non-receptor class of PTPs; (B): Receptor class of PTPs.

Table 1. Summary of alternative splicing in protein tyrosine phosphatases.

Tyrosine Phosphatase genes analyzed for splicing analysis	35
Tyrosine phosphatases with splice forms identified	25
PTPases identified with new exons	14
PTPases with skipped exons found	12
PTPases containing cryptic splice acceptor and donor sites	15

**Figure 2.** Novel isoform of PTPN18. The domain architecture and corresponding exon structures of PTPN18 and PTPN18B are illustrated. PTPN18B is a novel isoform identified through alternatively splicing analysis. The PTPase domain, coding regions and the untranslated regions are represented in orange, light green and grey respectively.

and one between exons 22 and 23 (BQ338787) (Fig. 3A). Both the novel exons (indicated by arrows) are found in the coding region, the latter encodes an extension of the second phosphatase domain of PTPN18. Details of all novel exons are given in Tables 2 and 3.

3.2.2.2 Identification of novel exons using comparative genomics

A comparative analysis of PTPRM transcript between human and mouse genomic and transcriptomic sequences resulted in the identification of an additional exon in the mouse transcript AK039261. When the corresponding human genomic sequence was searched, this exon was identified between exons 18 and 19 of the PTPRM gene, as defined by the transcript NM_002845 (Fig. 3B). Alternative transcripts derived from PTPRM revealed a number of alternative domain structures. A human EST (AV723421) showed a domain arrangement where the two PTPase domains were fused together. Due to cryptic acceptor and donor sites, the part of the exon coding for the linker regions was spliced, resulting in the union of two domains. Splicing in the vicinity of or within the catalytic domain governs the substrate specificity and func-

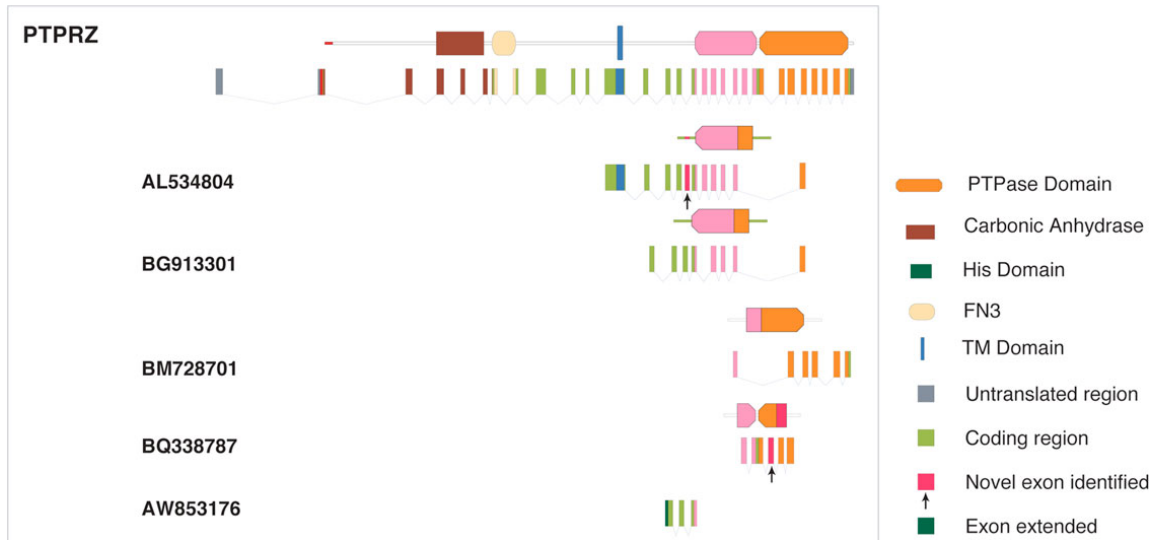
tion of the PTPase.¹⁷ Thus, a drastic change resulting from the fusion of two catalytically distinct PTPase domains is likely to yield a catalytically inactive form that might act as a dominant negative phenotype *in vivo*.

Analysis of the PTPRT gene between the human and mouse orthologs resulted in the detection of a novel exon between the 30th and 31st exons in the human transcript (NM_133170). The mouse transcript of PTPRT (AY026863) shows the presence of an exon which is missed in all the known variants of PTPRT in humans (Fig. 3C). It is likely that this exon is present in mRNA but has not been cloned thus far. The novel exons of PTPRM and PTPRT were experimentally validated by RT-PCR analysis using one primer specific to the novel exon and the other derived from a known exon (Fig. 4).

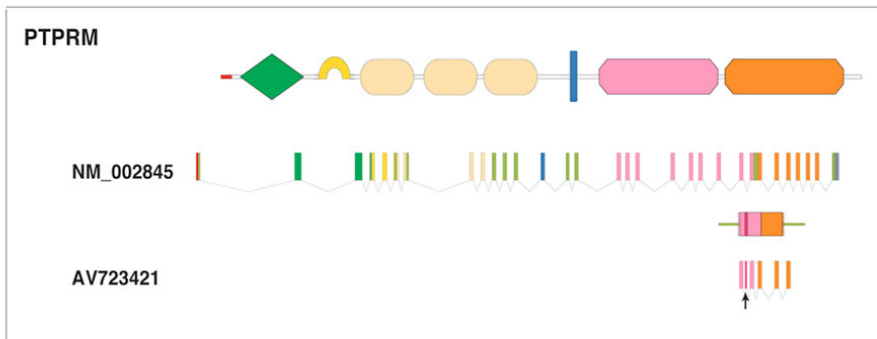
3.2.3. Identification of skipped exons

A skipped exon is one that is part of one isoform but is absent in other alternative forms. We found 19 novel events of skipped exons in 12 proteins among the 35 PTPs analyzed. Nine of the 19 novel skipping events were found to occur in 5 NRPTs and 10 events in 7 RPTs with PTPN4, PTPN12 and PTPRZ each

A



B



C

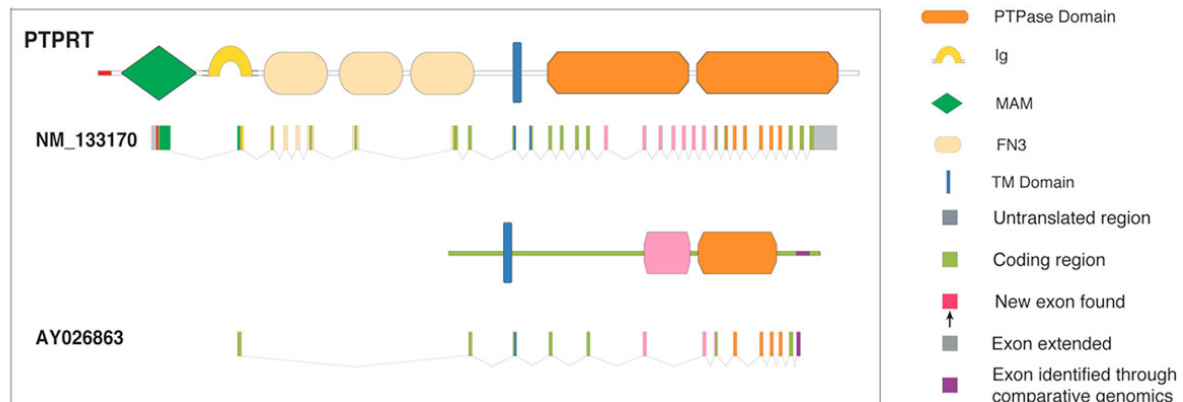


Figure 3. The alternative splicing events in PTPs. The domain architecture of the longest protein isoforms with their corresponding exon arrangements are shown along with the novel splicing events identified in the study. The color codes used for the domains are indicated. The exons are color coded according to the part of domains they encode. The untranslated regions and coding regions are shown in grey and light green, respectively. The novel alternative spliced events such as novel exons, exon extensions and exons identified through comparative genomic analysis are shown in pink, green and purple, respectively. (A) Splicing events in PTPRZ. (B) Splicing events in PTPRM. (C) Splicing events in PTPRT.

Table 2. Alternative splicing events among non-receptor protein tyrosine phosphatases.

	Gene Symbol	Whether Splice forms identified	Number of novel exons identified (EST/cDNA Accession #)	Number of transcripts with exons skipped (EST/cDNA Accession #)	Number of transcripts with evidence of cryptic splice acceptor/donor sites (EST/cDNA Accession #)	Whether splicing events affect coding region
1.	PTPN1	No	0	0	0	No
2.	PTPN2	Yes	3 (BG723710; AL041460; BM972725)	0	0	Yes
3.	PTPN3	No	0	0	0	No
4.	PTPN4	Yes	0	3 (AW798520; BM557364; BX342720)	0	Yes
5.	PTPN5	Yes	0	0	1 (BI457969)	Yes
6.	PTPN6	Yes	2 (BU540066; BQ221242)	0	1 (BM825262)	Yes
7.	PTPN7	Yes	0	0	1 (BE243885)	Yes
8.	PTPN9	Yes	1 (BM978412)	0	0	Yes
9.	PTPN11	No	0	0	0	No
10.	PTPN12	Yes	1 (BG829296)	3 (BG829296; BF105998; BM968291)	2 (BF105998; BF818703)	Yes
11.	PTPN13	Yes	1 (BM931852)	1 (BF308428)	0	Yes
12.	PTPN14	No	0	0	0	No
13.	PTPN18	Yes	0	1 (BQ061600)	3 (BE616252; AW849267; BI858031)	Yes
14.	PTPN21	Yes	1 (BM927098)	0	0	Yes
15.	PTPN22	No	0	0	0	No
16.	PTPN23	Yes	0	1 (BE882725)	3 (BX420767; BM893811; AA994453)	Yes

having a maximum of 3 novel exon skipping events each. Diagrammatic representations of all novel events of skipped exons can be visualized in the PTP Resource (<http://ptpr.ibioinformatics.org>).

3.2.4. Identification of cryptic splice sites

Instead of the splice sites used in the majority of transcripts, other alternative sites called cryptic splice sites can also be used (alternate 5' donor and 3' acceptor sites). Selection of cryptic splice sites by the spliceosome complex is a process that can alter mRNA structure. Mutations near the authentic splice sites allow the spliceosome

to select alternate 5' or 3' splice sites leading to alternatively spliced transcripts¹⁸ For example, a substitution of G to A in the longer form of human type I hair keratin hHa1 (hHa-1) results in the retention of a 41-nucleotide intronic sequence past the exon in the mature transcript. This leads to a frameshift leading to a premature stop codon and a shorter protein that lacks the tail domain in the keratin domain resulting in the pathological hair phenotype in hHa-1-t individuals.¹⁹ Several diseases are associated with cryptic 5' splice sites that lead to several splice versions of longer transcript either with deletion of an existing region of exons or incorporation of an intronic

Table 3. Alternative splicing events among receptor protein tyrosine phosphatases.

	Gene Symbol	Whether splice forms identified	Number of novel exons identified (EST/cDNA Accession #)	Number of transcripts with exons skipped (EST/cDNA Accession #)	Number of transcripts with evidence of cryptic splice acceptor/donor sites (EST/cDNA Accession #)	Whether splicing events affect coding region
1.	PTPRA	Yes	1 (BX481785)	0	1 (BX481785)	Yes
2.	PTPRB	No	0	0	0	No
3.	PTPRC	Yes	0	2 (CD688103; BG756746)	1 (BM146183)	Yes
4.	PTPRD	Yes	0	1 (AI652235; AV652567)	0	Yes
5.	PTPRE	Yes	1 (BM920649)	1 (BQ337341)	1 (BI768443)	Yes
6.	PTPRF	Yes	1 (BM450545)	0	1 (BQ300386)	Yes
7.	PTPRG	Yes	0	1 (BM150416)	0	Yes
8.	PTPRH	Yes	1 (BG248753)	1 (BG248753)	0	Yes
9.	PTPRJ	No	0	0	0	No
10.	PTPRK	Yes	0	1 (AI825662)	2 (BE869224; AV658169)	Yes
11.	PTPRM	Yes	8 (AL589249; CF552499; AK039261; AI040971; AV723421; BM921533; CB051573)	0	0	Yes
12.	PTPRN	Yes	1 (BX115193)	0	0	Yes
13.	PTPRN2	No	0	0	0	No
14.	PTPRO	Yes	0	0	2 (BX451544; AW448986)	Yes
15.	PTPRR	No	0	0	0	No
16.	PTPRS	No	0	0	0	No
17.	PTPRT	Yes	1 (AY026863)	0	1 (BE778141)	Yes
18.	PTPRU	Yes	0	0	1 (BM825262)	Yes
19.	PTPRZ	Yes	2 (AL534804; BQ338787)	3 (AL534804; BM728701; AV721260)	3 (BU737244; BM728701; AW853176)	Yes

region and all these forms are associated with diseases.²⁰

In our analysis, 15 PTPs were found to utilize cryptic splice sites (Tables 2 and 3). Among the 16 NRPTPs analyzed, 6 proteins were found to have a total of 11 novel cryptic acceptor and donor sites. Of the 19 RPTPs analyzed, 13 novel cryptic acceptor and donor sites were

found in a total of 9 proteins. Three novel cryptic splice sites were found in each of PTPN18, PTPN23 and PTPRZ. Details of novel cryptic acceptor and donor sites are given in Tables 2 and 3.

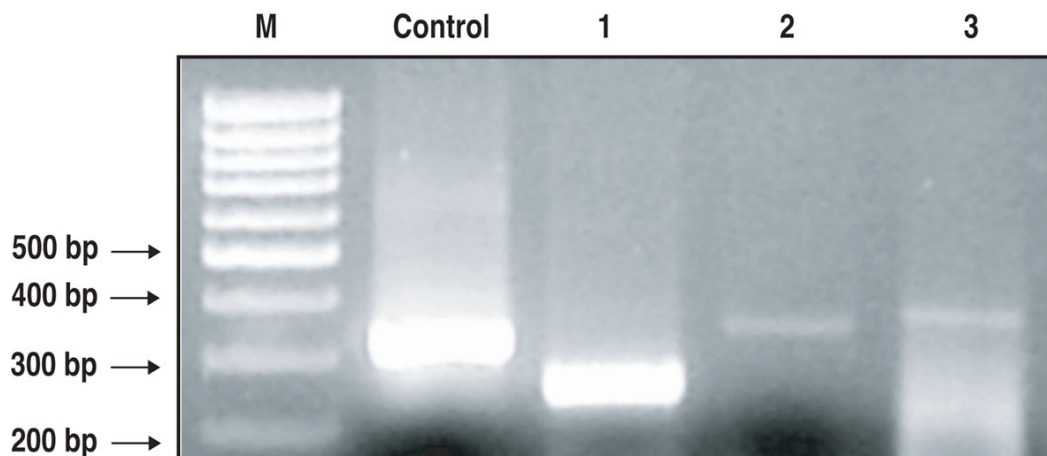


Figure 4. PCR amplification of the novel exons in PTPRM and PTPRT. The figure shows agarose gel electrophoresis of the PCR products containing novel exons. Lanes 1 and 2 represent fragments containing the novel exon of PTPRM; in lane 1, the forward primer used is located within the novel exon, whereas in lane 2, the reverse primer is located within the novel exon. Lane 3 represents the fragment containing the novel exon of PTPRT where the forward primer is located within the novel exon. Numbers on the left indicate the fragment sizes of the marker used (*M*).

3.3. Implications of alternative splicing in determining substrate specificity and localization

Variation in the C-terminus of PTP-S2 transcripts results in two isoforms, each possessing a unique substrate specificity and localization, either to the nucleus or cytoplasm.^{21,22} Interestingly, the alternate splicing in PTPN2 results in isoforms having variations in their C-terminus, as observed in the EST BM972725. Thus, it is possible that these alternatively spliced forms might have different localization and substrate specificity, for example, the PDZ domain containing PTPs.

Overall, we found that 71% of the human PTPs analyzed have alternative splice variants (Tables 2 and 3). The majority of alternative splicing events affected the coding region of PTPs. Among non-receptor PTPs, all 16 novel splicing events were discovered at coding regions, except in the case of PTP18 where both coding and non-coding regions were found to have novel splice sites. Among the receptor PTPs, only 5 novel splicing events were detected in non-coding regions whereas 18 novel splicing events occurred in coding regions. In the case of PTPRK, two new splice sites were detected in the non-coding region of transcript. Among the PTPs analyzed, PTPRM has 8, which is the highest number of novel alternate splice sites found in the present study, of which 7 are within coding regions. PTPRZ1 and PTPN12 were found to have 6 and 4 novel alternate splice sites, respectively, all in coding regions of their respective transcripts. The overall statistics of splicing events are summarized in Tables 2 and 3.

3.4. Tyrosine phosphatase pseudogenes in the human genome

We have analyzed pseudogenes using the criteria outlined in Material and Methods. A total of 11 pseudogenes were found among classical PTPs. Our analysis is in agreement with that published by Anderson and coworkers.⁴ All of these pseudogenes were found to be processed except for a non-processed pseudogene observed in PTPRD. Two of the pseudogenes corresponding to PTPN11 (BQ022482 and CD244998) and one corresponding to PTPN2 (AW401979) were also found to be transcribed.

3.5. Creation of a web-based protein tyrosine phosphatase resource

We developed a novel database to serve as a community resource for information regarding molecular information, alternative splice forms, and pseudogenes of PTPs. This database can be accessed at <http://ptpr.ibioinformatics.org>. Results of the detailed analysis of alternative splicing and pseudogene were incorporated in the database resource. These molecules are linked to the Human Protein Reference Database (HPRD)¹⁵ where information about domain architecture, post-translational modifications, protein-protein interactions and enzyme-substrate relationships is available. Figure 5 shows a screen shot of PTPN18 as depicted in the PTP Resource.

The information can be accessed by clicking on each PTPase from the list provided on the home page of the resource. Upon clicking an entry, the user can view three hyperlinked tabs: Molecular information, Pseudogenes and Alternative splicing. The molecular information section provides information about alternative names, gene

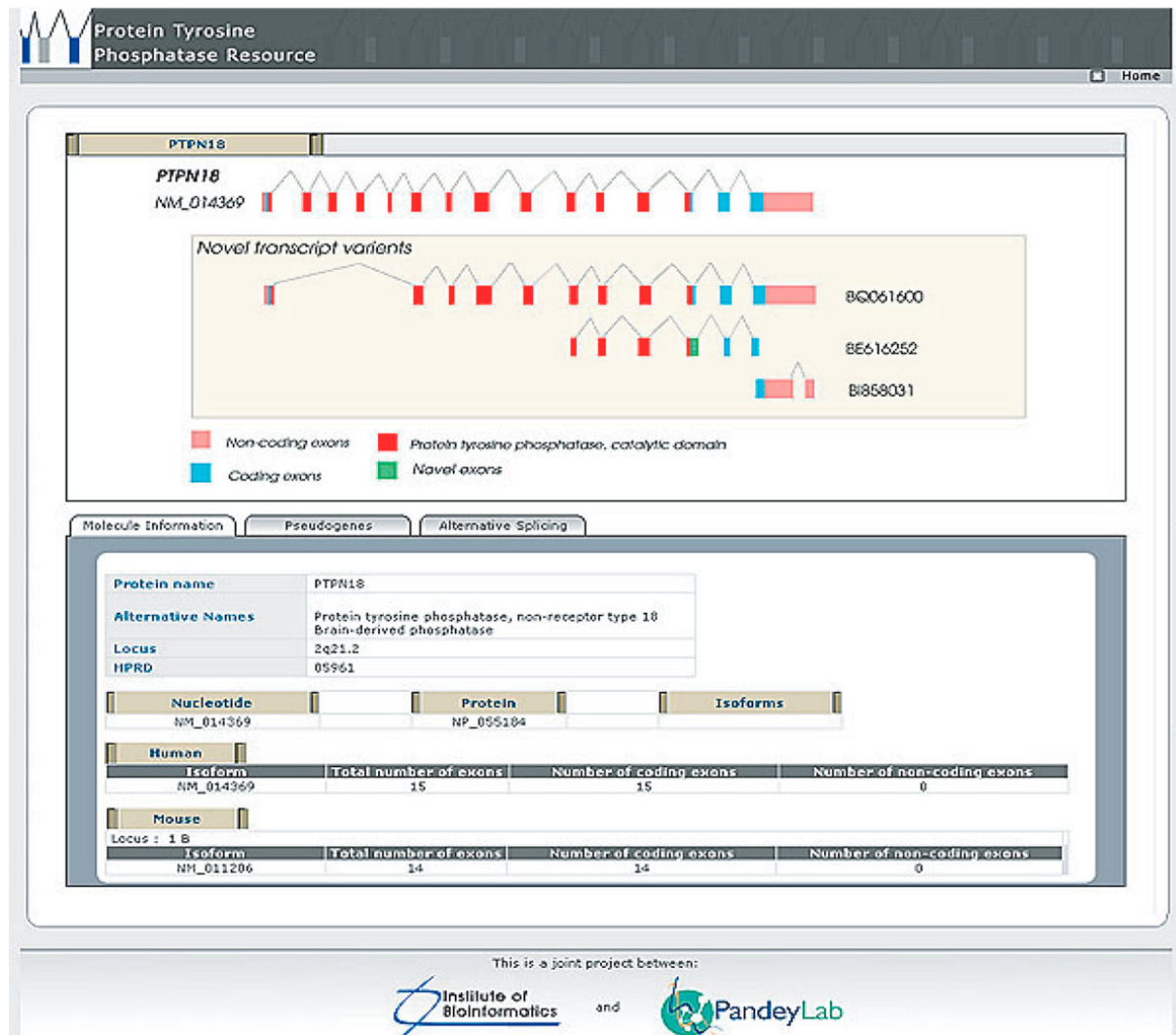


Figure 5. Screen shot of the Protein Tyrosine Phosphatase Resource (PTPR). The figure shows the PTPN18 molecule page. The hyperlinked tabs lead to molecular information, pseudogenes and alternative splicing pages. A link to the Human Protein Reference Database (HPRD) is provided from the molecular information page. The image shows the gene structure of all isoforms and the novel splice variants identified.

and protein accession numbers for all isoforms, and the number of coding and non-coding exons. Links to other databases such as HPRD and OMIM are also provided. In the alternative splicing section, information regarding the total number of transcript variants identified in the analysis can be viewed along with details on the number of new exons identified, the number of exons skipped and other splicing events. EST and protein sequence alignments are accessible for further reference by clicking the alignments hyperlink. An image depicting the gene structure of the PTP gene and its isoforms is displayed above each PTP entry. Novel splice forms are marked as novel transcripts with the accession number of the corresponding EST. This data will be useful for researchers to characterize and understand the role of novel PTP isoforms.

4. CONCLUSIONS

In this study, 61 alternative splice forms for both receptor and non-receptor tyrosine phosphatase genes and a novel isoform of PTPN18 (designated as PTPN18 β) have been characterized and documented. Novel exons of PTPRM and PTPRT predicted through comparative genomics were verified by RT-PCR amplification. The remarkable number of alternative splicing events among PTPs is a means to diversify the activities of a small set of proteins. We have developed a Protein Tyrosine Phosphatase Resource, which should prove to be a useful community resource about genomic and proteomic information related to PTPs.

Acknowledgements: Akhilesh Pandey is supported by a Career Development Award from the Breast Cancer SPORE at Johns Hopkins (CA 88843) and a Kimmel

Scholar Award from the Sidney Kimmel Foundation for Cancer Research. The authors thank Troels Kristiansen and Jun Zhong for helpful comments. Dr. Pandey serves as Chief Scientific Advisor to the Institute of Bioinformatics. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.

References

- Zhang, Z. Y. 1998, Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis, *Crit. Rev. Biochem. Mol. Biol.*, **33**, 1–52.
- Li, L. and Dixon, J. E. 2000, Form, function, and regulation of protein tyrosine phosphatases and their involvement in human diseases, *Semin. Immunol.*, **12**, 75–84.
- McPherson, J. D., Marra, M., Hillier, L. et al. 2001, A Physical map of the human genome, *Nature*, **409**, 934–941.
- Andersen, J. N., Jansen, P. G., Echwald, S. M. et al. 2004, A genomic perspective on protein tyrosine phosphatases: gene structure, pseudogenes, and genetic disease linkage, *FASEB J.*, **18**, 8–30.
- Alonso, A., Sasin, J., Bottini, N. et al. 2004, Protein Tyrosine Phosphatases in the Human Genome, *Cell*, **117**, 699–711.
- Bhaduri, A. and Sowdhamini, R. 2003, A genome-wide survey of human tyrosine phosphatases, *Protein Eng.*, **16**, 881–888.
- Boulanger, L. M., Lombroso, P. J., Raghunathan, A., During, M. J., Wahle, P., and Naegele, J. R. 1995, Cellular and molecular characterization of a brain-enriched protein tyrosine phosphatase, *J. Neurosci.*, **15**, 1532–1544.
- Brett, D., Hanke, J., Lehmann, G. et al. 2000, EST comparison indicates 38% of human mRNAs contain possible alternative splice forms, *FEBS Lett.*, **474**, 83–86.
- Longo, F. M., Martignetti, J. A., Le Beau, J. M., Zhang, J. S., Barnes, J. P., and Brosius, J. 1993, Leukocyte common antigen-related receptor-linked tyrosine phosphatase. Regulation of mRNA expression, *J. Biol. Chem.*, **268**, 26503–26511.
- Matthews, R. J., Cahir, E. D., and Thomas, M. L. 1990, Identification of an additional member of the protein-tyrosine-phosphatase family: evidence for alternative splicing in the tyrosine phosphatase domain, *Proc. Natl. Acad. Sci. USA.*, **87**, 4444–4448.
- Zhang, J. S. and Longo, F. M. 1995, LAR tyrosine phosphatase receptor: alternative splicing is preferential to the nervous system, coordinated with cell growth and generates novel isoforms containing extensive CAG repeats, *J. Cell Biol.*, **128**, 415–431.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. 1998, SMART, a simple modular architecture research tool: identification of signaling domains, *Proc. Natl. Acad. Sci. USA.*, **95**, 5857–5864.
- Marshall, M., Moxon, S., Sonnhammer, E. L. et al. 2004, The Pfam protein families database, *Nucleic Acids Res.*, **30**, 276–280.
- Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M., and Miller, W. 1998, A computer program for aligning a cDNA sequence with a genomic DNA sequence, *Genome Res.*, **8**, 967–974.
- Peri, S., Navarro, J. D., Amanchy, R. et al. 2003, Development of human protein reference database as an initial platform for approaching systems biology in humans, *Genome Res.*, **13**, 2363–2371.
- Kim, Y. W., Wang, H., Sures, I., Lammers, R., Martell, K. J., and Ullrich, A. 1996, Characterization of the PEST family protein tyrosine phosphatase BDP1, *Oncogene*, **13**, 2275–2279.
- Matthews, R. J., Cahir, E. D., and Thomas, M. L. 1990, Identification of an additional member of the protein-tyrosine-phosphatase family: evidence for alternative splicing in the tyrosine phosphatase domain, *Proc. Natl. Acad. Sci. USA.*, **87**, 4444–4448.
- Nelson, K. K. and Green, M. R. 1990, Mechanism for cryptic splice site activation during pre-mRNA splicing, *Proc. Natl. Acad. Sci. USA.*, **87**, 6253–6257.
- Winter, H., Hofmann, I., Langbein, L., Rogers, M. A., and Schweizer, J. 1997, A splice site mutation in the gene of the human type I hair keratin hHa1 results in the expression of a tailless keratin isoform, *J. Biol. Chem.*, **272**, 32345–32352.
- Roca, X., Sachidanandam, R., and Krainerr, A. R. 2003, Intrinsic differences between authentic and cryptic 5' splice sites, *Nucleic Acids Res.*, **31**, 6321–6333.
- Kamatkar, S., Radha, V., Nambirajan, S., Reddy, R. S., and Swarup, G. 1996, Two splice variants of a tyrosine phosphatase differ in substrate specificity, DNA binding, and subcellular location, *J. Biol. Chem.*, **271**, 26755–26761.
- McLaughlin, S. and Dixon, J. E. 1993, Alternative splicing gives rise to a nuclear protein tyrosine phosphatase in *Drosophila*, *J. Biol. Chem.*, **268**, 6839–6842.
- Zhao, R., Fu, X., Li, Q., Krantz, S. B., and Zhao, Z. J. 2003, Specific interaction of protein tyrosine phosphatase-MEG2 with phosphatidylserine, *J. Biol. Chem.*, **278**(25), 22609–22614.
- Ogata, M. et al. 1999, Effects of overexpression of PTP36, a putative protein tyrosine phosphatase, on cell adhesion, cell growth, and cytoskeletons in HeLa cells, *J. Biol. Chem.*, **274**(18), 12905–12909.
- Kim, Y. W., Wang, H., Sures, I., Lammers, R., Martell, K. J., and Ullrich, A. 1996, Characterization of the PEST family protein tyrosine phosphatase BDP1, *Oncogene*, **13**(10), 2275–2279.