Published in final edited form as:

Ann Hum Genet. 2009 July ; 73(Pt 4): 422–428. doi:10.1111/j.1469-1809.2009.00525.x.

Identifying putative promoter regions of Hermansky-Pudlak syndrome genes by means of phylogenetic footprinting

H. Stanescu¹, T.G. Wolfsberg², R.T. Moreland², M.H. Ayub¹, E. Erickson¹, W. Westbroek¹, M. Huizing¹, W.A. Gahl¹, and A. Helip-Wooley^{1,*}

¹ Section on Human Biochemical Genetics, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

² Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Summary

HPS is an autosomal recessive disorder characterized by oculocutaneous albinism and prolonged bleeding. Eight human genes are described resulting in the HPS subtypes 1–8. Certain HPS proteins combine to form Biogenesis of Lysosome-related Organelles Complexes (BLOCs), thought to function in the formation of intracellular vesicles such as melanosomes, platelet dense bodies, and lytic granules. Specifically, BLOC-2 contains the HPS3, HPS5 and HPS6 proteins. We used phylogenetic footprinting to identify conserved regions in the upstream sequences of *HPS3*, *HPS5* and *HPS6*. These conserved regions were verified to have *in vitro* transcription activation activity using luciferase reporter assays. Transcription factor binding site analyses of the regions identified 52 putative sites shared by all three genes. When analysis was limited to the conserved footprints, seven binding sites were found shared among all three genes: Pax-5, AIRE, CACD, ZF5, Zic1, E2F and Churchill. The *HPS3* conserved upstream region was sequenced in four patients with decreased fibroblast *HPS3* RNA levels and only one *HPS3* mutation in the coding exons and surrounding exon/intron boundaries; no mutation was found. These findings illustrate the power of phylogenetic footprinting for identifying potential regulatory regions in non-coding sequences and define the first putative promoter elements for any HPS genes.

Keywords

Hermansky-Pudlak syndrome; lysosome-related organelle; phylogenetic footprinting; promoter; mutation analysis

Introduction

In 1959 Hermansky and Pudlak described “two unrelated albinos with lifelong bleeding tendency and peculiar pigmented reticular cells in the bone marrow as well as in lymph node and liver biopsies” (Hermansky & Pudlak, 1959). The subsequent description of similar cases expanded Hermansky-Pudlak Syndrome (HPS) into a well-recognized, though rare, autosomal recessive disorder of oculocutaneous albinism and bleeding. Some patients also exhibit intracellular accumulation of lipofuscin, and some manifest pulmonary fibrosis, granulomatous colitis, or neutropenia (Gahl et al. 1998; Huizing & Gahl, 2002; King et al. 2001).

* Address correspondence to: Amanda Helip-Wooley, Ph.D., 10 Center Drive, MSC 1852, Building 10, Room 10C-103, NHGRI, NIH, Bethesda, MD 20892-1852, 301-402-2354, FAX 301-402-2740, ahwooley@mail.nih.gov.

A portion of the clinical variability of HPS derives from the fact that HPS is genetically heterogeneous, resulting from mutations in one of eight genes (Wei, 2006). The functions of the proteins encoded by the HPS genes are not known, with the exception of *AP3B1*, which is associated with HPS-2. *AP3B1* codes for the β 3A subunit of adaptor protein complex 3, a heterotetrameric protein responsible for the formation of nascent vesicles from the trans-Golgi network or endosomes (Dell'Angelica et al. 1999).

Because HPS mutations lead to defects in lysosome-related organelles, such as melanosomes and platelet dense bodies, the various HPS proteins are thought to be involved in the biogenesis, trafficking, tethering, or docking of these vesicles. To achieve these functions, various HPS proteins combine with each other and with other proteins to form complexes called Biogenesis of Lysosome-related Organelles Complexes, or BLOCs (Bonifacio, 2004; DiPietro & Dell'Angelica, 2006; Gautam et al. 2006; Wei, 2006). BLOC-1 contains HPS7, HPS8, and other proteins (Starcevic & Dell'Angelica, 2004). BLOC-2 consists of the HPS3, HPS5, and HPS6 proteins (DiPietro et al. 2004), and BLOC-3 is composed of HPS1 and HPS4 (Martina et al. 2003; Nazarian et al. 2003). The proteins composing BLOC-2 and BLOC-3 have no sequence similarity to each other or to other known genes. Mutations in BLOC-3 genes (*HPS1* and *HPS4*) confer a risk of developing pulmonary fibrosis that is fatal in the fourth or fifth decade (Huizing & Gahl, 2002). In contrast, patients with BLOC-2 mutations (*i.e.*, in *HPS3*, *HPS5*, or *HPS6*) have not been found to have lung disease (Huizing et al. 2001; Huizing et al. 2004).

Gene regulatory regions have not yet been defined for any HPS gene. These regions remain of interest for several reasons. First, the unique and specialized functions of HPS proteins could mean that the promoter regions of HPS genes might be defined by a unique collection of transcription factor binding sites. Second, HPS genes encoding proteins in the same BLOC might share promoter elements to effect concerted expression. Finally, HPS patients could have disease-causing mutations within the promoter region of their HPS genes, as has previously been reported for other inherited recessive diseases such as cystinosis (Phornphutkul et al. 2001) and congenital erythropoietic porphyria (Solic et al. 2001). To address these issues, we used phylogenetic footprinting (Duret & Bucher, 1997; Zhang & Gerstein, 2003) to define the boundaries of the conserved putative promoter regions of the BLOC-2 genes (*HPS3*, *HPS5*, and *HPS6*) and performed transcription activation studies to verify our predictions. Transcription factor binding site prediction was then used to identify common putative binding sites within the conserved regions of the putative *HPS3*, 5, and 6 promoters.

Methods

Sequence analyses

Regions of interest (ROIs) for the BLOC-2 promoters were defined as the sequence of nucleotides stretching from 500-bp upstream of the transcription start site to 100-bp downstream, except in the case of *HPS5*, where the sequence was trimmed to exclude the 3'UTR of an upstream gene (*GTF2H1*). The transcription start site was obtained from the UCSC Genome Browser (build hg 18) combining RefSeq annotation with the Stanford Promoter dataset (Kent et al. 2002; Trinklein et al. 2003). Sequences were stored, handled and annotated using BioEdit software v.7.0.5 (<http://www.mbio.ncsu.edu/BioEdit/page2.html>).

For footprint studies, MCS sequences for the ROIs were obtained from Elliott Margulies and from the UCSC Genome Browser (build hg 18, conservation track 8XMAF). The homologous sequences used were *Homo sapiens* (hs), *Pan troglodytes* (pt), *Mus musculus* (mm), *Rattus norvegicus* (rn), *Canis familiaris* (cf), *Gallus gallus* (gg) and *Xenopus laevis*

(xl) as follows: *HPS3* hs, pt, mm, rn, cf, gg; *HPS5* hs, pt, mm, rn; *HPS6* hs, pt, mm, rn, cf, xl. Conserved sequences of length k (k-mers) that share homology across different organisms were identified using the motif discovery algorithm of Blanchette & Tompa (2003) as implemented on the FootPrinter 2.0 webserver. Analyses were performed for k = 6 to 12 using parsimony scores (maximum number of mispairings allowed across homologous k-mers) of 0 to 2. The conserved upstream region for each BLOC-2 gene was obtained by trimming the sequence to encompass the best quality footprints. The resulting regions for *HPS3* were 331 nucleotides (150,329,881-150,330,211) on chromosome 3 (hg18) encompassing 7 footprints, for *HPS5*, 404 nucleotides (18,300,295-18,300,698) on chromosome 11 (hg18) encompassing 6 footprints and for *HPS6*, 259 nucleotides (103,814,918-103,815,176) on chromosome 10 (hg18) encompassing 6 footprints. Sequence logos, which provide a graphical representation of a multiple sequence alignment, were generated for each of the *HPS3*, *HPS5*, and *HPS6* footprints (Supplemental Figure 1) using weblogo with default settings (<http://weblogo.berkeley.edu/>).

Putative transcription factor binding sites in the conserved upstream regions of *HPS3*, *HPS5* and *HPS6* were identified by searching the TRANSFAC Professional 10.4 transcription factor database with MATCH 10.4 using the following parameters: Profile: vertebrates, use only high quality matrices, minimize the sum of both error rates, mat. sim.: 0.7, core sim.: 0.75. When we narrowed the search to focus only on binding sites that overlapped with footprints, we chose only those sites whose conserved core was contained within the footprint.

Randomly selected sequences upstream of transcription start sites

The UCSC Table Browser was used to obtain the strand, transcription start and end, and gene symbol for 3267 randomly selected human NCBI mRNA Reference Sequences (RefSeqs) (Mar 2006 assembly, hg18). On average, the HPS conserved upstream regions are 330 nucleotides in length, and contain 50 nucleotides of the first exon. Therefore, we extracted sequence for each randomly chosen RefSeq that begins 280 nucleotides upstream of the annotated transcription start site, and ends 50 nucleotides downstream of it. For the purpose of this analysis we assumed that the actual transcription start site is close to the annotated transcription start site. The output was limited to RefSeqs that map to a single location in the genome. A total of 3115 random promoters were analyzed with Match 10.4 using the same parameters as for the HPS conserved upstream regions. The number of randomly selected upstream sequences that have at least one copy of each of the seven binding sites shared by the HPS conserved upstream regions was determined by using the binding matrices found in one or more HPS genes.

Transcriptional activation assays

The conserved upstream regions, with respect to the transcription start site, were nucleotides -115 to +206 for *HPS3* (321-bp), -396 to -13 for *HPS5* (383-bp), and -215 to +34 for *HPS6* (249-bp). These regions were amplified from control genomic DNA using sequence-specific primers (for *HPS3*: 5'-GCAGGGACCCAGCGGTC-3' and 5'-CGACGTCCGGCGGGATG-3'; for *HPS5*: 5'-GTCTGAGGGCACGTGAC-3' and 5'-CTTGCTGACGTCATCCAG-3') or were directly synthesized (*HPS6*); all were cloned into pCR-Blunt II-TOPO (Invitrogen, Carlsbad, CA, USA). The conserved upstream regions were then subcloned into a firefly luciferase reporter vector (pGL4.10[luc2]; Promega, Madison, WI, USA). These constructs, as well as positive and negative controls, were transfected into HEK293 cells with an Amaxa nucleoporator using kit V and program Q-01 (Amaxa, Gaithersburg, MD, USA). Cells were co-transfected with pGL4.74[hRluc/TK] encoding *Renilla* luciferase to normalize for transfection efficiency. Both firefly and *Renilla* luciferase activities were assayed approximately 16 h post-transfection using the Dual-

Luciferase Reporter Assay System (Promega, Madison, WI, USA) on a microplate Centro LB 960 luminometer (Berthold Technologies, Oak Ridge, TN, USA). The relative transcriptional activation of each construct was calculated by subtracting background readings, as determined from mock transfected cells, and dividing the firefly luciferase reading by the *Renilla* luciferase reading. The activities of the *HPS3*, *HPS5* and *HPS6* conserved upstream regions are reported as fold of control, determined by dividing the relative activity of each construct by that of the control with no promoter. The activity of each construct was assayed in three independent experiments and the means and standard errors were calculated.

Molecular analyses of HPS patients

HPS patients were enrolled in a protocol approved by the National Human Genome Research Institute's Institutional Review Board, and written informed consent was obtained. Genomic DNA was isolated from peripheral blood using standard techniques. Mutation analysis of an approximately 1.1-kb region (−875 to +270, with respect to the transcription start site) including the predicted *HPS3* conserved promoter region (−115 to +206) was performed on genomic DNA of four HPS-3 patients in which only one *HPS3* mutation had been identified in the protein-coding exons and their boundaries. These patients also displayed absent or decreased expression of *HPS3* RNA by northern blotting. Previously identified mutations were: in patient HPS#34-3 [IVS16+2500G>A] (Huizing et al., 2001); in patient HPS#32-3 [c.2464C>T; p. R822X]; in patient HPS#50-3 [c.1195A>G; c.1199insATTGTC]; and in patient HPS#92-3 [c.1189C>T; p. R396W]. Standard PCR and sequencing methods were used with the following forward primers: 5'-CATCTGAAGGGCCTAGGAAA-3', 5'-AACAGCATCCCTCAATCCAT-3' and 5'-GCAGGGACCCAGCGGTC-3' and the reverse primers: 5'-AACGCCTCCACCTTGCAG-3', 5'-CTGACCGCGAATGTAGGACT-3' and 5'-CGACGTCCGGCGGGATG-3'.

Results

We first searched for conserved sequence upstream of human *HPS3* (NM_032383; chromosome 3q24), *HPS5* (NM_181507; chromosome 11p15.1), and *HPS6* (NM_024747; chromosome 10q24.32) by aligning their putative promoter regions. We extracted sequence from −500 to +100 relative to the transcription start site for each gene and compared them by BLAST, but found no significant sequence similarity. We then used phylogenetic footprinting to identify conserved blocks of sequence (footprints) within each of these three regions (Duret & Bucher, 1997; Zhang & Gerstein, 2003). We ran the upstream sequences from *Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, *Canis familiaris*, *Gallus gallus* and *Xenopus laevis* *HPS3*, *HPS5*, and *HPS6* through FootPrinter 2.0. The best quality footprints were used to establish the boundaries of each conserved upstream region (331 nucleotides/7 footprints for *HPS3*, 404 nucleotides/6 footprints for *HPS5* and 259 nucleotides/6 footprints for *HPS6* (Figure 1). To verify that each of the three conserved upstream regions had a transcription activating function, we subcloned them into a firefly luciferase reporter vector and performed transcriptional activation assays. The cloned regions used for luciferase assays were slightly larger than the conserved upstream regions in order to accommodate suitable primers for PCR. All three BLOC-2 conserved upstream regions had significantly greater relative promoter activity than the control luciferase vector containing no promoter (*HPS3* $p \leq 0.001$, *HPS5* $p \leq 0.002$, and *HPS6* $p \leq 0.003$). The *HPS5* conserved upstream region had the highest activity (141 ± 49 fold of control); *HPS3* had the lowest activity (14 ± 6 fold of control) and *HPS6* had intermediate activity (36 ± 17 fold of control) when transfected into HEK293 cells (Figure 2). These activities are significantly

higher than the control lacking a promoter but quite modest compared with SV40 promoter activity (5669 ± 2377 folds of control).

The conserved upstream regions were then analyzed for transcription factor binding sites (TFBS) using the Match algorithm provided with the Transfac Professional 10.4 dataset. This analysis yielded 52 putative TFBS shared among all three genes. Refining the search for TFBS to the overlaps between predicted sites and the footprints, we obtained seven putative TFBS that occurred in all three conserved upstream regions, i.e., Pax-5, AIRE, CACD, ZF5, Zic1, E2F and Churchill.

To determine the uniqueness of these seven transcription factor binding sites, we searched the putative promoter regions of 3,115 human genes with the transcription factor binding site matrices from Transfac Professional 10.4. The putative promoter regions were defined simply by nucleotide length (330) and proximity to the transcription start site (-280 to $+50$), both factors determined by the average length and position of the HPS conserved upstream regions. Of the 3,115 genes, 2,128 (68%) had at least one copy of each of these seven predicted transcription factor binding sites. For technical reasons it was not possible to automate FootPrinter 2.0 to determine how many of these fall within conserved regions.

Finally, we determined if a mutation in the *HPS3* conserved upstream region might be clinically important. In our group of over 20 HPS-3 patients (Anikster et al. 2001; Huizing et al. 2001) there were four patients with decreased *HPS3* RNA levels in fibroblasts, in which only one mutation in a coding exon or surrounding exon/intron boundaries was found. We sequenced the *HPS3* conserved upstream region in these patients, but found no additional mutations that would explain their HPS phenotype. We had no good candidate patients for *HPS5* or *HPS6* promoter mutations.

Discussion

Genes encoding proteins that interact or function in the same pathway are often transcribed using a similar set of promoter signals, usually located in noncoding regions (Castillo-Davis, 2005; Zuckerland, 2002). Identification of noncoding cis-acting sequences, such as transcription factor binding sites, is difficult because their patterns are not apparent. The length of a conserved transcription factor binding site is approximately 20 base pairs, of which only 5–10 bp constitute the “recognition core” (Latchman, 2005). The ends of such a motif are often poorly defined, meaning that recognition and docking of transcription factors on the DNA promoter elements occurs at least in part by pattern matching. Short patterns are distributed across the genome by chance, and could constitute false-positives. To diminish the rate of false-positives, the technique of phylogenetic footprinting was used. This method relies on the fact that regulatory elements may be under selective pressure, making them less prone to mutation than the non-functional surrounding sequence. Hence, conserved motifs are more likely than nonconserved motifs to have a regulatory function (Castillo-Davis, 2005; Dickmeis & Muller, 2005; Xie et al. 2005; Zhang & Gerstein, 2003).

Traditional phylogenetic footprinting involves constructing a global multiple alignment of the sequences of interest and identifying conserved aligned regions. This approach is not accurate if the regions under consideration are very divergent, because insertions and deletions ‘shuffle’ the conserved sequences, modifying the distances between them and their relative positions. Hence, we employed a modification of this method that relies upon detecting motifs of a relatively short sequence length. We applied this technique to three genes (*HPS3*, *HPS5*, and *HPS6*) whose protein products interact in a complex called BLOC-2 but otherwise have no sequence similarity to each other. Dysfunction of any

BLOC-2 protein results in Hermansky-Pudlak syndrome (Huizing et al. 2008; Wei et al. 2006).

The putative regulatory regions of *HPS3*, *HPS5*, and *HPS6* were identified using phylogenetic footprinting, and the transcriptional activation activities of these conserved upstream regions were verified by expression studies using a reporter gene. The conserved upstream region of *HPS5* had much greater activity than that of *HPS3* or *HPS6*. The reason for this remains unknown; the *HPS5* conserved upstream region had the greatest sequence length but did not have an increased number of recognized transcription factor binding sites or footprints compared with the *HPS3* or *HPS6* conserved upstream regions (Figure 1).

Transcription factor binding site prediction using MATCH 10.4 yielded 52 putative sites common to the conserved upstream regions of *HPS3*, *HPS5*, and *HPS6*. By combining the techniques of transcription factor binding site prediction and phylogenetic footprinting, we reduced the number of common putative binding sites from 52 to seven (Pax-5, AIRE, CACD, ZF5, Zic1, E2F and Churchill). A review of the literature did not reveal any known role for these transcription factors in lysosome-related organelle biogenesis or in cell types particularly affected in HPS, i.e. melanocytes and megakaryocytes. A search of 3,115 randomly selected sequences upstream of transcription start sites found that the coincidence of these seven predicted transcription factor binding sites is not unusual. An important distinction, however, is that in the case of the HPS upstream sequences, these sites were identified in the conserved upstream regions, which are both considerably shorter and highly conserved across species. In the randomly selected upstream sequences, conserved sequences could not be determined using the same methodology; therefore, binding sites were sought in all 330 nucleotides. The bulk of those discovered are likely to be false positives.

To determine which footprints contain the most highly conserved sequence, and may therefore have the most regulatory significance, *HPS3*, *HPS5*, and *HPS6* footprints were analyzed by sequence logos (Supplemental Figure 1). Those footprints containing a 'core' of at least two completely conserved contiguous residues include *HPS3* footprints 2 and 3, *HPS5* footprints 1,2,3,5, and 6 and *HPS6* footprint 6. It should be noted that for *HPS3* and *HPS6* sequences from six species were available for comparison versus only four species available for *HPS5*, potentially explaining the difference in the number of "highly conserved" footprints. Interestingly, only one putative transcription factor binding site, AIRE, is found overlapping just these "highly conserved" footprints in all three HPS genes. It has yet to be determined if these highly conserved footprints contribute more to the regulation of HPS gene expression.

This work defines the putative promoter regions of HPS genes for the first time, along with their predicted shared transcription factor binding sites. Identification of the putative promoters for *HPS3*, *HPS5*, and *HPS6* provides an additional region in which to search for disease-causing mutations in HPS patients. The study also demonstrates that transcription factor binding site searches alone are not a reliable method for the identification of promoter regions. However, we show that phylogenetic footprinting can be used to refine the search for regulatory regions in conserved, non-coding DNA in genes that share no other sequence similarity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors appreciate the advice and assistance of Richard Hess, Robert Kleta, Elliot Margulies and Anthony Antonellis. This research was supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.

References

- Blanchette M, Tompa M. Discovery of promoter elements by a computational method for phylogenetic footprinting. *Genome Res.* 2002; 12:739–748. [PubMed: 11997340]
- Blanchette M, Tompa M. FootPrinter: a program designed for phylogenetic footprinting. *Nucleic Acids Res.* 2003; 31:3840–3842. [PubMed: 12824433]
- Bonifacino JS. Insights into the biogenesis of lysosome-related organelles from the study of the Hermansky-Pudlak syndrome. *Ann NY Acad Sci.* 2004; 1038:103–114. [PubMed: 15838104]
- Bucher P, Bryan B. Signal search analysis: a new method to localize and characterize functionally important DNA sequences. *Nucleic Acids Res.* 1984; 12:287–305. [PubMed: 6546421]
- Castillo-Davis CI. The evolution of noncoding DNA: how much junk, how much func? *Trends Genet.* 2005; 21:533–536. [PubMed: 16098630]
- Dell'Angelica EC, Shotelersuk V, Aguilar RC, Gahl WA, Bonifacino JS. Altered trafficking of lysosomal proteins in Hermansky-Pudlak syndrome due to mutations in the beta-3A subunit of the AP-3 adaptor. *Mol Cell.* 1999; 3:11–21. [PubMed: 10024875]
- Dickmeis T, Muller F. The identification and functional characterization of conserved promoter elements in developmental genes. *Brief Funct Genomic Proteomic.* 2005; 3:332–350. [PubMed: 15814024]
- DiPietro SM, Falcon-Perez JM, Dell'Angelica EC. Characterization of BLOC-2, a complex containing the Hermansky-Pudlak syndrome proteins HPS3, HPS5 and HPS6. *Traffic.* 2004; 5:276–283. [PubMed: 15030569]
- DiPietro SM, Dell'Angelica EC. The cell biology of Hermansky-Pudlak syndrome: Recent advances. *Traffic.* 2005; 6:525–533. [PubMed: 15941404]
- Duret L, Bucher P. Searching for promoter elements in human noncoding sequences. *Curr Op Struct Biol.* 1997; 7:399–405.
- Gahl WA, Brantly M, Kaiser-Kupfer MI, Iwata F, Hazelwood S, Shotelersuk V, Duffy LF, Kuehl EM, Troendle J, Bernardini I. Genetic defects and clinical characteristics of patients with a form of oculocutaneous albinism (Hermansky-Pudlak syndrome). *N Engl J Med.* 1998; 338:1258–1264. [PubMed: 9562579]
- Gautam R, Novak EK, Tan J, Wakamatsu K, Ito S, Swank RT. Interaction of Hermansky-Pudlak Syndrome genes in the regulation of lysosome-related organelles. *Traffic.* 2006; 7:779–792. [PubMed: 16787394]
- Hermansky F, Pudlak P. Albinism associated with hemorrhagic diathesis and unusual pigmented reticular cells in the bone marrow: report of two cases with histochemical studies. *Blood.* 1959; 14:162–169. [PubMed: 13618373]
- Huizing M, Anikster Y, Fitzpatrick DL, Jeong AB, D'Souza M, Rausche M, Toro JR, Kaiser-Kupfer MI, White JG, Gahl WA. Hermansky-Pudlak syndrome type 3 in Ashkenazi Jews and other non-Puerto Rican patients with hypopigmentation and platelet storage-pool deficiency. *Am J Hum Genet.* 2001; 69:1022–1032. [PubMed: 11590544]
- Huizing M, Gahl WA. Disorders of vesicles of the lysosomal lineage: the Hermansky-Pudlak syndromes. *Curr Mol Med.* 2002; 2:451–467. [PubMed: 12125811]
- Huizing M, Hess R, Dorward H, Claassen DA, Helip-Wooley A, Kleta R, Kaiser-Kupfer MI, White JG, Gahl WA. Cellular, molecular and clinical characterization of patients with Hermansky-Pudlak syndrome type 5. *Traffic.* 2004; 5:711–722. [PubMed: 15296495]
- Huizing M, Helip-Wooley A, Westbroek W, Gunay-Aygun M, Gahl WA. Disorders of Lysosome-Related Organelle Biogenesis: Clinical and Molecular Genetics. *Annu Rev Genomics Hum Genet.* 2008; 9:359–386. [PubMed: 18544035]

- Jegga AG, Gupta A, Gowrisankar S, Deshmukh MA, Connolly S, Finley K, Aronow BJ. CisMols Analyzer: identification of compositionally similar cis-element clusters in ortholog conserved regions of coordinately expressed genes. *Nucleic Acids Res.* 2005; 33:W408–411. [PubMed: 15980500]
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The Human Genome Browser at UCSC. *Genome Res.* 2002; 12:996–1006. [PubMed: 12045153]
- Khokha MK, Loots G. Strategies for characterizing cis-promoter elements in *Xenopus*. *Brief Funct Genomic Proteomic.* 2005; 4:58–68. [PubMed: 15975265]
- Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. *Nature.* 2005; 436:11. [PubMed: 16001029]
- King, RA.; Hearing, VJ.; Creel, DJ.; Oetting, WS. Albinism. In: Scriver, CR.; Beaudet, AL.; Valle, DL.; Sly, WS., editors. *The Metabolic and Molecular Bases of Inherited Disease*. New York: McGraw-Hill; 2001. p. 5587-5627.
- Latchman, D. *Gene Regulation*. Vol. 206. New York: Taylor & Francis; 2005.
- Martina JA, Moriyama K, Bonifacio JS. BLOC-3, a protein complex containing the Hermansky-Pudlak syndrome gene products HPS1 and HPS4. *J Biol Chem.* 2003; 278:29376–29384. [PubMed: 12756248]
- Mellor J. The dynamics of chromatin remodeling at promoters. *Mol Cell.* 2005; 19:147–157. [PubMed: 16039585]
- Nazarian R, Falcon-Perez JM, Dell'Angelica EC. Biogenesis of lysosome-related organelles complex 3 (BLOC-3): a complex containing the Hermansky-Pudlak syndrome (HPS) proteins HPS1 and HPS4. *Proc Natl Acad Sci USA.* 2003; 100:8770–8775. [PubMed: 12847290]
- Phornphutkul C, Anikster Y, Huizing M, Braun P, Brodie C, Chou JY, Gahl WA. The promoter of a lysosomal membrane transporter gene, CTNS, binds Sp-1, shares sequences with the promoter of an adjacent gene, CARKL, and causes cystinosis if mutated in a critical region. *Am J Hum Genet.* 2001; 69:712–721. [PubMed: 11505338]
- Sipos L, Gyurkovics H. Long-distance interactions between enhancers and promoters. *FEBS J.* 2005; 272:3253–3259. [PubMed: 15978032]
- Solis C, Aizencang GI, Astrin KH, Bishop DF, Desnick RJ. Uroporphyrinogen III synthase erythroid promoter mutations in adjacent GATA1 and CP2 elements cause congenital erythropoietic porphyria. *J Clin Invest.* 2001; 107:753–762. [PubMed: 11254675]
- Starcevic M, Dell'Angelica EC. Identification of snapin and three novel proteins (BLOS1, BLOS2, and BLOS3/reduced pigmentation) as subunits of biogenesis of lysosome-related organelles complex-1 (BLOC-1). *J Biol Chem.* 2004; 279:28393–28401. [PubMed: 15102850]
- Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol.* 1998; 203:439–455. [PubMed: 3199442]
- Trinklein N, Force Aldred S, Saldanha A, Myers RM. Identification and functional analysis of human transcriptional promoters. *Genome Res.* 2003; 13:308–312. [PubMed: 12566409]
- Wei M. Hermansky-Pudlak syndrome: a disease of protein trafficking and organelle function. *Pig Cell Res.* 2006; 19:19–41.
- Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Landers ES, Kelis M. Systematic discovery of promoter motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature.* 2005; 434:17.
- Zhang Z, Gerstein M. Of mice and men: phylogenetic footprinting aids the discovery of promoter elements. *J Biol.* 2003; 2:11. [PubMed: 12814519]
- Zuckerland E. Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. *Genetica.* 2002; 115:105–129. [PubMed: 12188043]

HPS3

HPS5

HPS6

Figure 1.

HPS3, *HPS5*, and *HPS6* conserved upstream regions.

Custom tracks in the UCSC Genome Browser on Human March 2006 Assembly (hg18). The location of each HPS conserved upstream region on the respective chromosome is indicated in red; coordinates are indicated in the first track. Footprints (black) are shown in the second track. The cloned region used for luciferase assays is in track 3 (gray). Putative transcription factor binding sites are designated in track 4 as follows: ZF5 (red), Zic1 (purple), E2F (orange), Pax-5 (aqua), AIRE (green), Churchill (blue) and CACD (gray). Transcripts of each gene are shown in blue; for *HPS5* the translation start site (ATG) is marked by a thickened blue line. *HPS3* and *HPS6* are on the plus strand, *HPS5* is on the minus strand.

Figure 2.

Activity of *HPS3*, *HPS5*, and *HPS6* conserved upstream regions as determined by luciferase activity.

The promoter activities of *HPS3*, *HPS5* and *HPS6* are reported as fold of control, determined by dividing the relative promoter activity of each construct by that of the no promoter control. The activity of each construct was assayed in three independent experiments and the means and standard errors were calculated.