

SCIENTIFIC REPORTS



OPEN

Smooth Muscle Transcriptome Browser: offering genome-wide references and expression profiles of transcripts expressed in intestinal SMC, ICC, and PDGFR α ⁺ cells

Adrienne Breland, Se Eun Ha, Brian G. Jorgensen, Byungchang Jin, Treg A. Gardner, Kenton M. Sanders & Seungil Ro 

Transcriptome data on the quantitative numbers of transcriptional variants expressed in primary cells offer essential clues into specific cellular functions and biological processes. We have previously collected transcriptomes from primary smooth muscle cells (SMC), interstitial cells of Cajal (ICC), and PDGFR α ⁺ cells (fibroblast-like cells) isolated from murine jejunal and colonic smooth muscle and/or mucosal tissues as well as transcriptomes from the associated tissues (jejunal smooth muscle, colonic smooth muscle, and colonic mucosa). In this study, we have built the Smooth Muscle Transcriptome Browser (SMTB), <https://med.unr.edu/physio/transcriptome>, a web-based, graphical user interface that offers genetic references and expression profiles of all transcripts expressed at both the cellular (SMC, ICC, and PDGFR α ⁺ cells) and tissue level (smooth muscle and mucosal tissue). This browser brings new insights into the cellular and biological functions of the cell types in gastrointestinal smooth muscle biology.

The proper regulation of gut activity is vital to homeostasis and survival. The digestive process of absorbing nutrients and releasing them into the bloodstream is achieved through a series of synchronized involuntary movements (motility) of gastrointestinal (GI) smooth muscle, which mixes food and propels the digested content through the GI tract¹. Smooth muscle tissue is comprised of a diverse range of unique cellular subpopulations that require isolation for individual study to aid in the elucidation of each subpopulations contribution to the functioning of the overall tissue. Motility in the GI tract is controlled by several types of cells including smooth muscle cells (SMC), interstitial cells of Cajal (ICC), PDGFR α ⁺ cells (fibroblast-like cells), as well as the enteric nervous system (ENS)¹. ICC generate spontaneous electrical slow waves², the ENS generates complex rhythmic motor behavior³, and PDGFR α ⁺ cells mediate enteric inhibitory responses^{4,5}, all of which control SMC, the final effectors for muscle contraction and muscle relaxation¹. The three cell types, SMC, ICC, and PDGFR α ⁺ cells (SIP cells), are electrically coupled via gap junctions and create an electrical syncytium, which collectively regulate GI motility¹. Developmental abnormalities and pathophysiological damage to these cells are directly linked to GI neuromuscular diseases such as Hirschsprung's disease⁶, diabetic gastroenteropathy⁷, gastrointestinal stromal tumor⁸, intestinal fibrosis⁹, and chronic intestinal pseudo-obstruction¹⁰. All of these motility diseases are thought to be developed from the remodeling of the smooth muscle in the GI tract, leading to abnormal growth (hypertrophy or tumor), myopathy, or death of the cells.

Genome-scale expression profiles of specific cell types provide indispensable information regarding cellular identity and function. To access the genetic information of SMC, ICC, and PDGFR α ⁺ cells within the small

Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, Nevada, USA. Adrienne Breland, Se Eun Ha and Brian G. Jorgensen contributed equally. Correspondence and requests for materials should be addressed to S.R. (email: sro@med.unr.edu)

intestine and colon, we launched a Smooth Muscle Transcriptome Sequencing Project. For this project, we isolated primary jejunal and colonic SMC, ICC, and PDGFR α^+ cells (mucosa and muscularis) from cell-specific GFP reporter mouse lines, and obtained a transcriptomic profile of each cell type and associated tissue^{11–14}. In analyzing each cell type's transcriptome, we identified new markers and signature genes for each cell type that are linked to cellular functions^{11–14}.

To help to explore this complex dataset, we built the SMTB. This graphical, web-based, browser displays the comprehensive expression profile and genomic map of each cell type and associated tissue within the colon and jejunum. The browser is available online, hosted by the University of Nevada, Reno at <https://med.unr.edu/physio/transcriptome>. This resource provides genome-wide genetic references and expression levels, enabling insight into genetic structure, expression profile, and isoforms of each gene expressed in key GI cell and tissue populations.

Results

The SMTB offers genome-wide genetic references and unique graphical images that can reveal new insights into the genetic structures, expression profiles, and isoforms of each gene expressed in key GI cell populations (SMC, ICC, and PDGFR α^+ cells) and GI tissues (jejunal SM, colonic SM and mucosa) for functional studies.

Applications

- Expression levels of various genes within GI tissues and GI cell types.
- Expression levels, and numbers, of transcriptional gene variants in GI tissues and GI cell types.
- Observing genomic structure (promoter, exons and introns) of transcriptional variants.
- Primer design for RT-PCR or qPCR (designing primers to span exon to exon junctions in order to minimize genomic DNA amplification and to detect specific transcriptional variants).
- Viewing splicing donor and acceptor sequence sites of transcriptional variants.
- Obtaining cDNA sequences for transcriptional variants.
- Finding open reading frames within transcriptional variants.

User's guide

The SMTB is accessible at <https://med.unr.edu/physio/transcriptome/smooth-muscle-transcriptome-browser>.

1. Once arrived at the home page, click “Access the Smooth Muscle Transcriptome Browser” to take you to the browser.
2. Go to “Select Track” as shown in Fig. 1a. There are two references of the mouse genome, mm9 (NCBI37, July 2007) and mm10 (GRCm38, Dec. 2011). Select one reference from the “Reference” section. As an example, mm9 was selected in Fig. 1a. Under the transcripts section, there are seven cell types (SMC Jejunum, ICC Jejunum, PDGFR α C Jejunum, SMC Colon, ICC Colon, PDGFR α C Colon, and PDGFR α C Mu Colon), three tissue types (SM Jejunum, SM Colon, and Mu Colon), and combined transcripts (GI All). Select cell(s)/tissue(s) as interested. For example, SMC Colon and SMC Jejunum were selected (Fig. 1a). Once all selections have been made, click “Back to Browser”.
3. The browser itself contains sections include “Search,” “Overview,” “Region,” and “Details” (Fig. 1b). Type a gene name in “Gene Name or Location” under the Search section (overwrite the chromosome location) and click “Search.” As an example, Acta2 (aka alpha smooth muscle actin) was typed in and searched for (Fig. 1c). The browser retrieves and displays the genomic map of Acta2 and the transcriptional variants expressed in both colonic and jejunal SMC. In the “Overview” and “Region,” it displays the location of Acta2 on chromosome 19 with the zoomed in location highlighted in a light blue bar in the “Region” section. This bar can be clicked and dragged to shift the displayed chromosomal location. In the “Details” section, the transcriptional maps of four isoform variants expressed in colonic and jejunal SMC are displayed. Exons in the variants are marked by green (colonic SMC) or yellow (jejunal SMC) boxes. For Acta2, the directionality of the arrows on the lines of each variant indicates that the cDNAs match to the antisense strand. Among the four Acta2 transcriptional variants, TCONS_00138735 is the longest in both colonic and jejunal SMC which is the same as the reference transcript of Acta2 shown in black (Fig. 1c). Wider boxes on the line of the reference gene mark a coding cDNA region while narrow boxes show 5' and 3' non-coding regions.
4. The genomic view can be zoomed in and out on any chromosomal location (Fig. 2). Fixed ranges for zooming are from 100 bp to 1 Mbp (Fig. 2a). In addition, zoom levels can be changed in 10% increments using “+” or “-” buttons in the “Scroll/Zoom” section just above the browser. Furthermore, the view can be freely moved to left or right by clicking any of the “<<”, “<”, “◀” (to left), “>>”, “>”, “▶” (to right) buttons, or by holding and dragging the light blue bar in either direction (Fig. 2b). Chromosomal location and distance can be displayed and measured by clicking the ruler and dragging to the area of interest (Fig. 2b). The reference genes displayed are linked to Gene Search Results in NCBI showing a list of a related gene associated with a full gene report when clicked (Fig. 2b).
5. The genomic map image of Acta2 can be saved in “Snapshots” by clicking the “Save Snapshot” button just above the “Scroll/Zoom” section (Fig. 3a). The saved image can be reloaded, removed, or stored in the “Snapshots” tab above the browser by clicking it (Fig. 3b). The stored image can be also downloaded (Fig. 3c).
6. Each transcriptional variant on the map is linked to detailed data. The longest transcriptional variant of Acta2, TCONS_00138735, will be used as an example for analysis (Fig. 3a). By clicking on the structural display of TCONS_00138735 under the “Details” section, you will be linked to a page containing

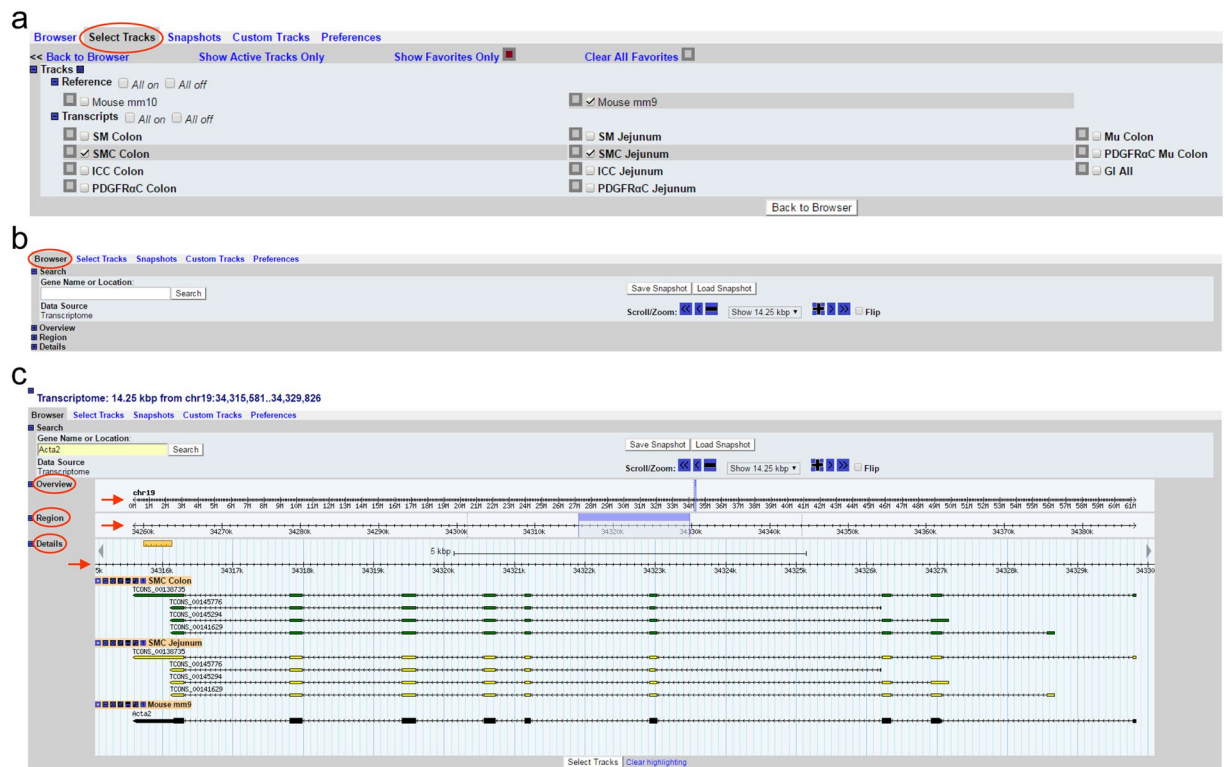


Figure 1. Smooth Muscle Transcriptome Browser built with Gbrowse 2.0. (a) “Select Tracks” tab showing the two selectable mouse reference genomes, mm9 and mm10, as well as the various selectable transcripts from each cell type and tissue. (b) The home screen of the “Browser” tab. (c) The search result of *Acta2* in the “Browser” tab. Shown are the “Overview,” “Region,” and “Details” section showing the chromosomal location, a highlighted region marker (light blue), and a graphical representation of *Acta2* transcriptional variants expressed in colonic (green) and jejunal (yellow) SMC. A map of the reference gene is also shown under the transcripts (black).

a summary of exonic numbers and locations, expression profiles, and cDNA sequence (Fig. 4). Exonic number, length, and associated chromosomal locations are summarized in Fig. 4a. Each exon contains a hyperlink to a map view of that exon. Expression levels (FPKM) of TCONS_00138735 in jejunal cells (JSMC, JICC, and JP α C), jejunal muscle tissue (JSM), colonic cells (CSMC, CICC, and CP α C), colonic muscle tissue (CSM), colonic mucosal PDGFR α + cells (CMuP α C), and colonic mucosal tissue (CMu) are shown in as various histograms (Fig. 4b). This variant is dominantly expressed in JSMC and CSMC. It is also noticeably expressed in JP α C and CMuP α C. Total expression levels of all four transcriptional variants of *Acta2* are shown in Fig. 4c. Figure 4d shows the DNA sequence of TCONS_00138735 containing 9 exons and 8 introns. The cDNA sequence can be downloaded as a.doc file by clicking the “Download cDNA Sequence” hyperlink above the displayed sequence. In addition, the cDNA can be further analyzed to search open reading frames (ORF).

- Instructions for searching ORF is shown in Fig. 5. *Acta2* TCONS_00138735 cDNA is 1,780 bp in length. The cDNA sequence was copied, pasted, and submitted in the Enter Query Sequence in the NCBI ORF Finder¹⁵ (Fig. 6a). The search produced 12 possible ORFs (ORF1-12, Fig. 6b). ORF1 is the longest one with 1,134 bp (377 amino acids) starting at nucleotide (nt) 71 to nt 1,204. The amino acid sequence can be further analyzed by protein BLAST (SMARTBLAST or BLAST) searching for homologous proteins.

Discussion

The Smooth Muscle Transcriptome Browser (SMTB) built in this study provides both the number and expression level of each transcriptional variant for all mapped genes expressed in SIP cells (SMC, ICC, and PDGFR α + cells) within the mucosa and muscularis layers of the jejunum and colon. The cellular transcriptomes and expression profiles are provided along with the expression profiles of each cell’s respective associated tissue including the jejunal muscularis, colonic muscularis, and mucosa. This browser allows researchers to analyze each transcriptional variant in terms of gene structure (promoter region, exons, and introns), expression levels within each cell type and tissue, and open reading frames (coding protein) within each variant.

We previously built UCSC Smooth Muscle Genome Browser (SMGB)¹¹ that can interact with our SMC transcriptome, SRF (SMC-specific transcription factor) binding CA α G boxes, and the ENCODE data contained within the UCSC Genome Browser¹⁶. Later, we added the transcriptomes of ICC, and PDGFR α + cells from both

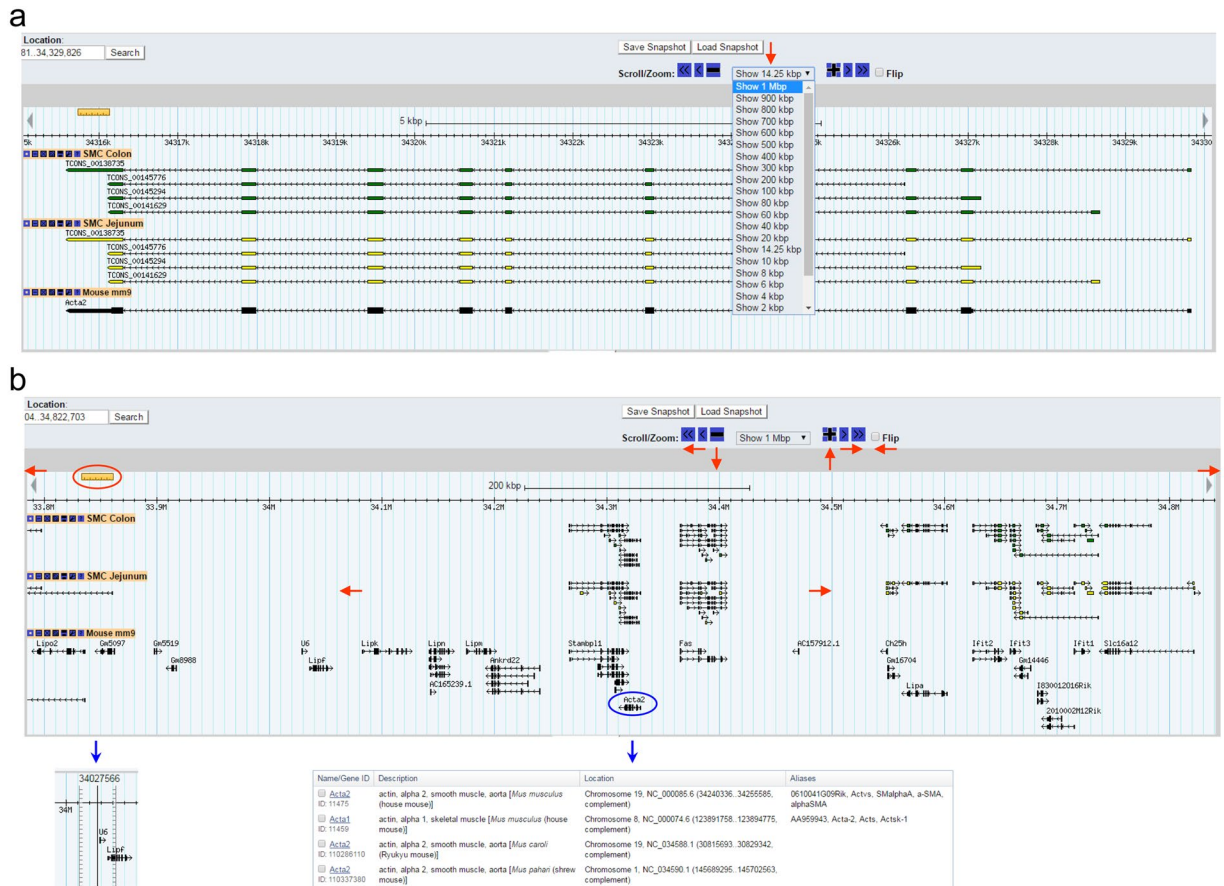


Figure 2. A map view of *Acta2* transcripts in the genome. **(a)** Zooming capabilities range from 1 Mbp to 100 bp at the *Acta2* locus (red arrow). **(b)** A genomic view of 1 Mbp around the *Acta2* gene locus. This map view can be freely moved left (←), or right (→) in the genome by the “Scroll/Zoom” menu options of “<”, “>”, “◀”, “▶”, “>>”, “<<”. Zoom is controlled by “-” or “+” or out (⏏) and in (⏏). “Flip” allows the user to switch between sense strand to anti-strand view or vice versa. Reference genes on the view are linked to the NCBI (*Acta2*, encircled in blue, is linked to reference genes at the NCBI). The ruler on left can be moved and denotes an exact chromosomal location of the cursor.

the muscularis and mucosa to our browser^{12–14}. This interactive database can analyze genetic/epigenetic structures and the regulation of individual genes expressed in SIP cells by integrating the abundant bioinformatics data in the UCSC Genome Database¹⁷. Now, as this paper has reported, we have built the SMTB that is capable of searching the expression profile of individual transcriptional variants of all mapped genes expressed in SIP cells, something that was not available with any of our previously reported SIP cell transcriptome studies. The two browsers serve as cooperative and powerful tools for use in functional studies of SIP cells.

As an example, the *Acta2* gene was chosen for analysis using the SMGB and SMTB (Fig. S1). Both browsers can carry the mouse mm9 genome showing the same region chr19: 34,315,581–34,329,826 (14,246 kbp), within which four transcriptional variants, V1–4, of the gene are expressed in JSMC and CSMC. Within the SMGB, SRF binding CarG boxes, ENCODE data (SRF, H3K27ac, Pol2, JunD, and c-Jun binding sites, as well as DNase I hypersensitivity) were selected (Fig. S1a). Both V1 and V2 start in the proximal promoter regions that contain two SRF binding sites, within which match two conserved CarG boxes for each variant (CCATATAGGG and CCAAACAAGG for V1, CCATATTTAG, CCTAATTAGG for V2). The region around the SRF binding sites coincide with H3K27ac and Pol2 sites (Fig. S1a), suggesting that the gene is active, and V1 and V2 are transcribed by Pol2 while being regulated by SRF (due to the presence of SRF binding sites and CarG boxes) in intestinal SMC. In addition, a binding site for the transcription factor subunits JunD and c-Jun can be found within intron 7, which also matches a DNase I hypersensitive region (Fig. S1a), suggesting the transcription of V1–4 could be regulated by the Jun family in JSMC and CSMC. The corresponding region (240 bp, chr19: 34,318,625–34,318,864) of the Jun binding site was located on SMTB (Fig. S1b), the sequence of 240 bp was analyzed for the presence of Jun binding sites in the transcriptional regulatory element search database, “PROMO”¹⁸. This search identified two binding sequences, TATGTCA and GAAGTCA, for JunD and c-Jun (Fig. S1b). Next, expression levels of the four variants of *Acta2* were found on SMTB. V1 and V2, likely regulated by SRF, show selective high expression in JSMC and CSMC, compared to colonic and jejunal ICC and PDGFR α^+ cells while V3 and V4 are dominantly expressed in CSMC and Jejunal PDGFR α^+ cells (Fig. 2a). V1, V2, and V4 are also expressed in mucosa PDGFR α^+ cells at relatively low levels. Each cDNA sequence for V1–4 was downloaded and an open reading frame (the

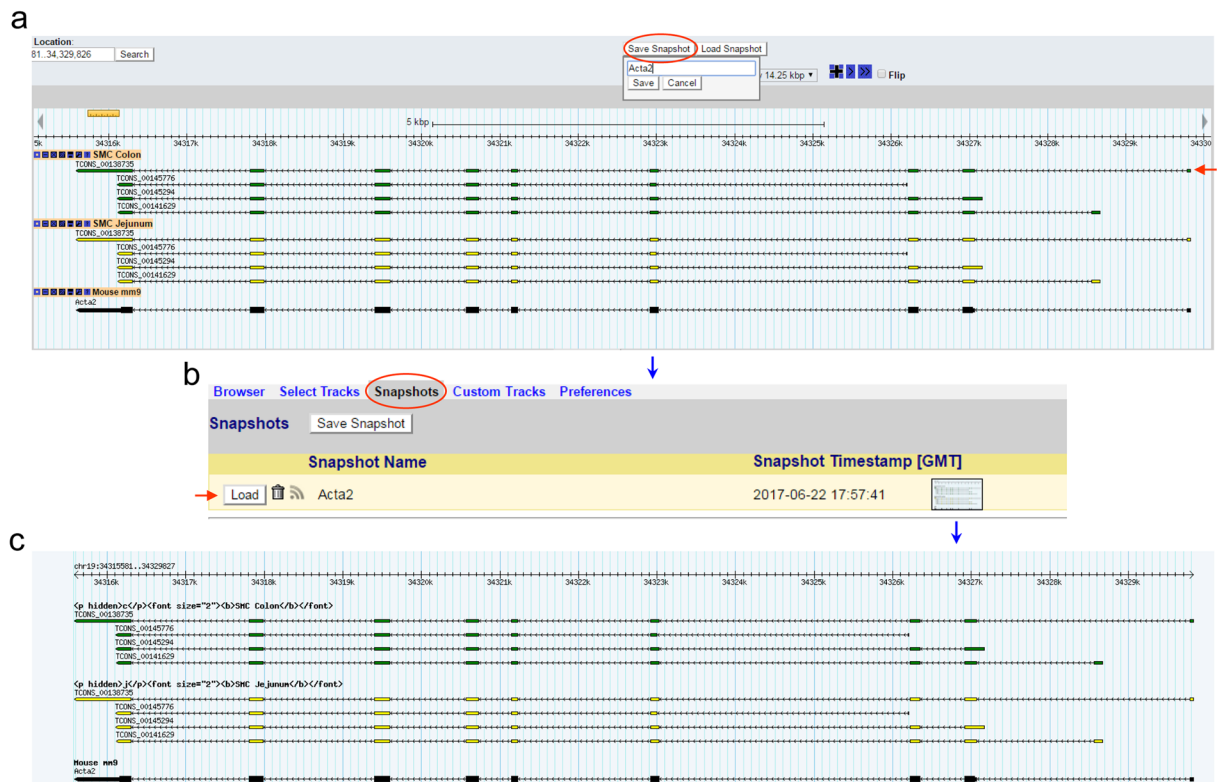


Figure 3. Saving and reloading a map of *Acta2*. (a) The map of *Acta2* to be saved as Acta2 by clicking “Save Snapshot”. (b) An image of the map view saved under the “Snapshots” tab. This view can be reloaded by “Load” or removed by clicking the trash can. (c) A saved image of *Acta2*. The reloaded view (a) shows four transcriptional variants found in “SMC Colon” and in “SMC Jejunum”. The longest variant is TCONS_00138735 and is indicated by a red arrow in (a).

longest one) was obtained from the NCBI ORF Finder¹⁵. Alignment of their amino acid sequences is shown in Fig. S2b, V1, V2, and V3 encode the full length of 377 amino acids which contains six post-translationally modified residues at N-terminus as shown in UniProt¹⁹. V1 encodes a truncated protein of 257 amino acids at the N-terminus that misses the six residues. As all of this integrated information shows, both the SMGB and SMTB are interactive browsers that provide genetic and epigenetic references that can be incorporated together for further functional studies in intestinal SIP cells and tissues.

While our transcriptome browser SMTB does share some similar features with other previously published genome/transcriptome browsers^{20,21}, SMTB contains a myriad of analyses that these resources do not contain the entirety that SMTB encompasses. SMTB allows for graphical representation of FPKM values and number of transcripts at whole tissue (smooth muscle and mucosa) and individual SIP cell type (SMC, ICC, and PDGFR α ⁺ cells) as well as being able to download cDNA sequences for both the mm9 and mm10 murine genome. Other browsers will allow for visualizations of FPKM data, such as the well-designed and valuable Brain RNA-seq project²¹, but it only uses a single genome reference (mm9), does not allow for cDNA sequence download or visualizations of alternatively spliced transcripts and has no overall FPKM for the entire tissue, only individual cell types. In contrast, a tissue reference genome, such as the RETINAL genome²⁰, only contains transcriptomic information from the total retinal tissue and not individual cell types and does not have graphical representations or cDNA sequence downloads and only uses a single genome reference. RNA-seq data from both the small and large intestine (colon) are also available through ENCODE (RNA-seq from LICR and UW, respectively) at the UCSC Genome Browser¹⁷. These RNA-seq data were obtained from whole small or large intestine tissue consisting of smooth muscle and mucosa which are functionally distinct and contain a myriad of different cell types. In contrast, we separated jejunal and colonic smooth muscle from their respective mucosa and mRNA from each was independently sequenced and deposited in the SMTB. Furthermore, while our transcriptome browser is definitively tissue specific to GI SIP types, it contains a plethora of data and user-friendly capabilities that have not been contained within one specific browser to date.

For all our previously reported transcriptomes, we obtained and analyzed the RNA-seq data from SIP cells contained within the murine jejunum and colon (Table 1) that unveiled a plethora of genetic information at the cellular levels. This report is the first to compile and compare these transcriptomes. SIP cells express anywhere from 15,192–17,172 known genes (Table 2), which account for 66–75% of all known mouse genes. In addition, these genes are transcribed into multiple transcriptional variants through alternative start sites and splicing. For example, the L-type Ca²⁺ channel *Cacna1c* is expressed into 22 different variants in jejunal and colonic SMC²². The average number of transcriptional variants per gene is 3 (Table 2). Most variants appear to be

Open Reading Frame (ORF) Finder at National Center for Biotechnology (NCBI)

1. Copy the cDNA sequence below
2. Paste it into the sequence text box at [NCBI ORF Finder](#) and submit it
3. Look for the longest ORF in predicted ones or check if any ORF has homology with known proteins in Smart BLAST
4. Translated protein sequence found below

```
>Acta2, TCONS_00138735, 1780 bp
ggacaccaccaccagagtggaagccagccagctcgtcaggaaccctgagacgctcctccagctatgtgtgaagaggaagacagcacagccctg
gtgtgcgacaaatggctctgggctctgtaagccggcttcgctgggtgtagatgctccagggtgttttcccaatcctcgtgggagctccagacacag
gagtaatggttggaaatggccaaaagacagctatgtgggggtagaagccagagcaagagagggatcctgacgctgaagatccgatagaacacggcat
ctcaccacactgggacgacatggaaaagatctggcaccactcttctataacgagctcgtgtggccctgaaagacatccgacactcgtgacagaggca
ccactgaaccctaaagcccaaccgggagaaaatgaccagattatgtttgagacctcaatgtccccgccatgtagtggtctatcaggctgtgtgtccc
tctatgcctctggacgtacaactgggtattgtgctggactctggagatgggtgactcacaacgtgcctatctatgagggctatgcccctgcctcatgccc
catgctctggactggctggcggagatctcaaccgactacctcatgaaagatcctgactgagcgtggctattcctctcgtgactactgcccagcgtgagatt
gtccgtgacatcaaggagaagctgtgctatgtagctctggactttgaaaaagagaaggccagggccctcctctctcctcctggagaagagctacgaac
tgctgacggcgagtgatcaccatggaaacgaaacgcttccgctgcccagagactctctccagccatcttctatgggtaggagtcagcgggcatcca
cgaaaccacctaataacgacatcagaaatgtgataatgacatcaggaaggatctctatgctaaacagctcctgacagggggtaccaccaatgtaccaggc
attgtgacagagtgacagaaggagatcacaagcctcgcaccagcaccatgaaagatcaagatcattgcccctccagaaacgaagctactctgtctggatcg
gtggctccatctggctcgtctctaccctccagcagatgtggaatcagcaaacaggaatacagcagaagctgggcccctccatcgtccaacgcaaatgctt
ctaagtcccccctgctctgcctctagcacaacaactgtgaacgttttggtagcagcgcctccagttcctttccaaatcattcctgcccagaagcttggatt
tgttactcgtggttttttaaaaaataatcagacatgtgctacccttaactttagccctggcctagcaacactgattcaattgtgttctgttaggtga
gaatcatttattataggtggccccctctagtggctcaggaatagatgtgacacaactagaaaagatattccagcagtggaatggcacaatcactcaca
cagaggaaggtccactaaatgtagtaaacggggtatttgagagcgtataggactgtatagattacataataataatattagtagcacaatagggaaa
atcgcaactgacttcacagctagagtgactatttactttatacacaacaggtgaccagataggtaaaaaaacttttactgtattttaaaaatttggcaag
taataatggttttttaaaatcacatcttttattaaagcacaacagctgcatggatgcatgctgacttttaactt
```

Figure 5. The cDNA sequence of TCONS_00138735 and instructions on using the Open Reading Frame (ORF) Finder at NCBI.

cell-specific. Further characterization of each variant is required to explore localization of variants in subpopulations of SIP cells (e.g. longitudinal SMC, circular SMC, intermuscular ICC/PDGFR α^+ cells, and intramuscular ICC/PDGFR α^+ cells), as well as to further understand the functional role of variants (if they are translated into protein or if they are pseudo-substrates for microRNAs).

The transcriptomes of SIP cells are very similar, sharing not only the genes that are expressed in all three types (around 93%), but also showing comparable expression levels for these gene transcripts. These similar transcriptomic profiles lend credence to the notion that these three cell types might have a shared developmental lineage arising from mesenchymal cells²³. In contrast to their largely shared expression profiles, SIP cells also have ~1,500 genes that are unique to each cell within the syncytium. These cell type-specific genes may contribute to the individual phenotypic identity, and unique functioning, of each cell type in the SIP syncytium.

Through previous gene ontology (GO) term analysis of the predominately expressed genes within SIP cells, we found the main functional roles of each cell type. SMC genes showed unique GO terms related to muscle contraction including genes related to the cytoskeleton, actin binding, calcium ion binding, myosin complex, and smooth muscle contractile fibers¹¹. ICC genes showed GO terms related to membrane excitability including membrane integrity, plasma membrane, metal ion binding, and transport¹³. PDGFR α^+ cell genes revealed GO categories related to the function and structure of the extracellular matrix^{12,14}.

The cell-specific markers identified in our previous transcriptome studies provide new tools to study SIP cells^{11–13} and our new SMTB allows for easy visualization and comparison of this elucidated data. We found that the most distinctive markers for detecting primary mature SMC are *Cnn1*, *Mylk*, *Tpm2*, *Tpm1*, *Des*, and *Myh11*¹¹. SMC are phenotypically dynamic with the ability to dedifferentiate back to a myofibroblast-like state induced by pathological conditions causing SMC overgrowth and within *in vitro* conditions²⁴. The six SMC markers should be used together to evaluate SMC phenotype in pathological conditions or cultured conditions. We also identified the new markers *Thbs4* and *Hcn4* that could be used in ICC identification. Currently, and historically, ICC are identified and tagged through the use of KIT (CD117) antibodies. However, the expression of KIT is not consistent across various ICC phenotypic conditions²⁵. For example, in some GI pathologies that cause digestive dysmotility, ICC lose their expression of KIT, making KIT antibodies a poor marker of ICC as they cannot be used to track ICC that undergo phenotypic alterations²⁶. Our newly uncovered ICC-selective marker, THBS4, may allow for the tracking of ICC long after KIT expression is lost. Additionally, we also found that expression of *Cacna1g* may be an identifying marker of hyperplastic PDGFR α^+ cells in the context of various GI diseases (small bowel obstruction, colorectal cancer, Crohn's disease, and diverticulitis)¹². Lastly, we identified the new marker *Adamdec1* in mucosal PDGFR α^+ cells (unpublished data). Expression of *Adamdec1* is induced in the DSS-induced colitis mouse model and human tissue affected by Crohn's disease (unpublished data).

Prior to the publication of our SIP cell transcriptomes, there was no transcriptome-level comparative tools available for analyzing cells contained within the SIP syncytium; it was only available to be compared through individual extraction of each data point in extremely large and generally opaque spreadsheets. Our new SMTB is a new and dynamic tool that allows for quick, easy, and clear-cut access to genetic information for individual transcripts in individual SIP cells and their associated tissue while allowing for comparison between both transcripts and cells/tissue. This tool will bring insights into new potential pathways for future research in the vast field of GI smooth muscle biology.

Materials and Methods

Intestinal tissues and cells. Pooled total RNAs were isolated from gastrointestinal (GI) tissue (2 males and 2 females, 1–2 months old) or sorted cells (up to 20 males and 20 females, 1–2 months old) from GFP reporter mouse lines: SMC from Myh11-Cre-eGFP²⁷, ICC from Kit-copGFP²⁸, and PDGFR α^+ cells from Pdgfra-eGFP²⁹ (Table 1). All procedures involving animals and their care were performed in accordance with institutional, state

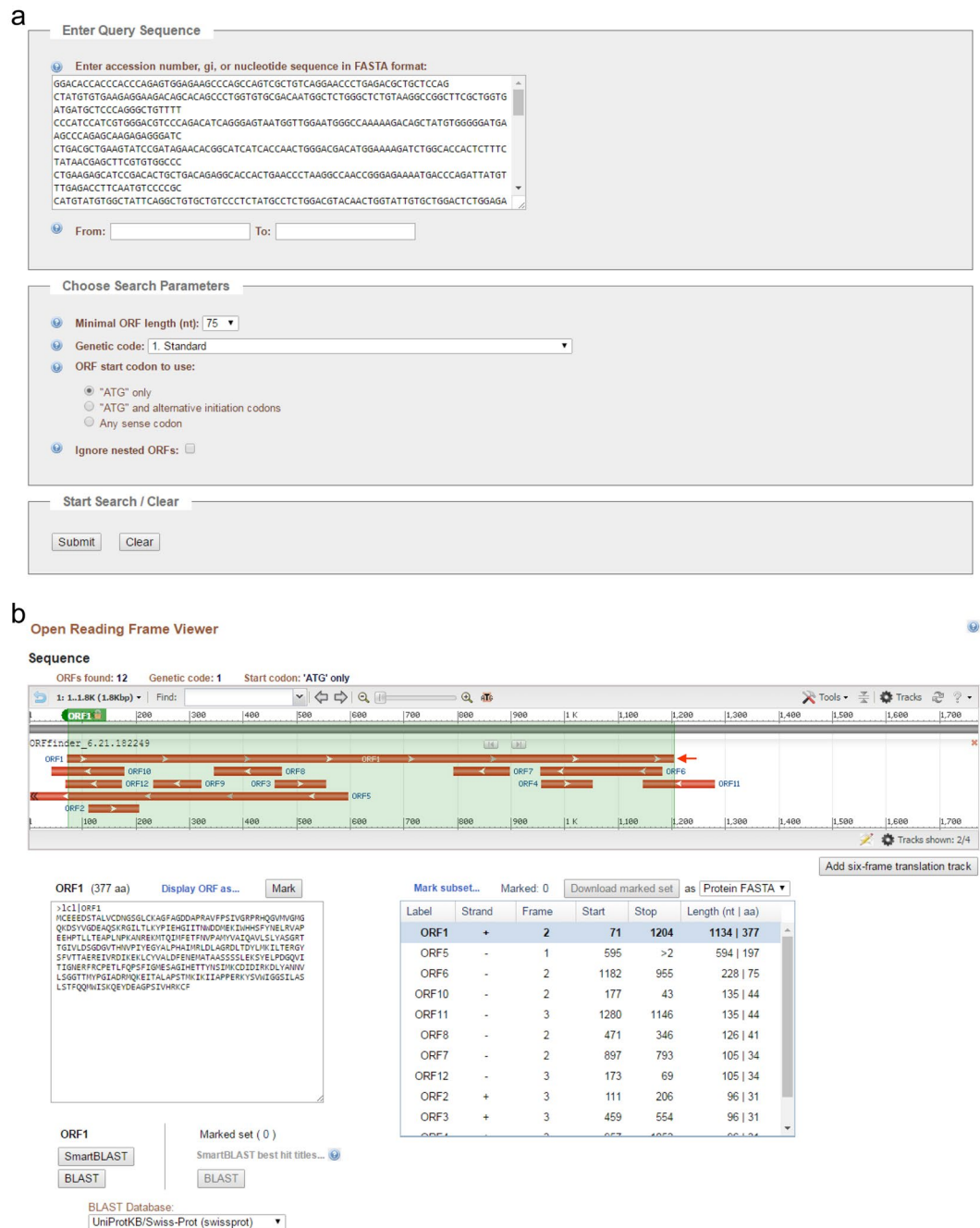


Figure 6. Submission of TCONS_00138735 cDNA to ORF Finder and selection of a putative ORF. **(a)** A snapshot of TCONS_00138735 cDNA pasted in ORF Finder. **(b)** A snapshot of the ORF Viewer showing TCONS_00138735 ORFs. The longest ORF, ORF1, is indicated by an arrow. Translated ORF1 amino acid (aa) sequence and summary of all ORFs are shown.

and national guidelines. The animal protocol was approved by the Institutional Animal Care and Use Committee at the University of Nevada-Reno Animal Resources. RNA-seq libraries were generated and sequenced via Illumina HiSeq 2000 (Illumina, San Diego, CA) at LC Sciences (Houston, TX) as previously described^{11–14}.

RNA-seq data analysis. Paired-end sequencing reads were isolated, processed, and analyzed as previously described^{11–14}. An FPKM = 0.025 was selected as our cutoff expression value as it resulted in equal false positive and false negative ratios of reliability. Transcripts with FPKM values of less than 0.025 were considered to be 0. Sequencing reads were assembled and annotated onto a reference genome (UCSC mm9 or mm10) using TopHat v1.4.1 software.

Name	Description	Mouse strain	Background	Age	Molecule
SM_Jejunum	Jejunum SM	B6.129S7-Kittm1Rosay/J	C57BL/6	4 wks	mRNA
SMC_Jejunum	SMC in jejunum SM	B6.Cg-Tg(Myh11-cre,-EGFP)2Mik/J	C57BL/6	4 wks	mRNA
ICC_Jejunum	ICC in jejunum SM	B6.129S7-Kittm1Rosay/J	C57BL/6	4 wks	mRNA
PaC_Jejunum	PDGFR α^+ cells in jejunum SM	B6.129S4-Pdgfrat11(EGFP)Sor/J	C57BL/6	4 wks	mRNA
SM_Colon	Colon SM	B6.129S7-Kittm1Rosay/J	C57BL/6	4 wks	mRNA
SMC_Colon	SMC in colon SM	B6.Cg-Tg(Myh11-cre,-EGFP)2Mik/J	C57BL/6	4 wks	mRNA
ICC_Colon	ICC in colon SM	B6.129S7-Kittm1Rosay/J	C57BL/6	4 wks	mRNA
PaC_Colon	PDGFR α^+ cells in colon SM	B6.129S4-Pdgfrat11(EGFP)Sor/J	C57BL/6	4 wks	mRNA
Mu_Colon	Colon mucosa	B6.129S4-Pdgfrat11(EGFP)Sor/J	C57BL/6	4 wks	mRNA
PaC_Mu_Colon	PDGFR α^+ cells in colon mucosa	B6.129S4-Pdgfrat11(EGFP)Sor/J	C57BL/6	4 wks	mRNA

Table 1. List of intestinal tissues and cells used for transcriptome studies.

Name	Total read	Mapped read	Known gene	Total isoform	Average isoform
SM_Jejunum	238,246,980	184,746,380	15,837	53,337	3.4
SMC_Jejunum	150,702,388	139,478,151	15,427	46,299	3.0
ICC_Jejunum	206,711,208	177,168,420	17,172	55,259	3.2
PaC_Jejunum	149,419,352	135,517,106	16,270	48,974	3.0
SM_Colon	212,744,016	164,801,684	16,359	55,345	3.4
SMC_Colon	151,445,926	139,645,488	15,192	47,275	3.1
ICC_Colon	193,155,892	168,384,646	16,820	54,488	3.2
PaC_Colon	136,563,090	123,172,778	15,714	48,894	3.1
Mu_Colon	168,835,236	153,208,218	15,933	52,113	3.3
PaC_Mu_Colon	154,151,508	141,592,530	15,777	51,282	3.3

Table 2. Summary of transcriptome data obtained from intestinal tissues and cells. *GI_All: all transcriptional variants expressed in intestinal tissues and cells above.

Transcriptome. Transcriptomes obtained from RNA-seq data analyses are shown in Table 2^{11–14}. Briefly, we obtained total reads of 136.6 M–238.2 M, 77–93% of which were mapped to the murine genome. The mapping identified 46,299–55,345 transcriptional isoform variants, which were aligned to 15,192–17,172 known genes. The average number of transcriptional variants per gene was 3.

Smooth Muscle Transcriptome Browser. Gbrowse 2.0³⁰ was used to build the SMTB. Gbrowse is a web-based genome browser originally developed in 2002 for use with Wormbase³¹. It enables web-based graphical sequence displays and sequence annotation. The browser has been used for other public data sources such as Flybase³², SGD³³, and SilkDB³⁴. In the browser, data can be downloaded for all transcripts on transcript detail pages. The SMTB was created using the standard Gbrowse 2.0 install with custom modifications. In addition to configuration files, Gbrowse files modified included CSS files and the cgi scripts “gbrowse” and “gbrowse_details”. Gbrowse 2.0 was installed on a Linux Debian 7.6 operating system. Transcript data files were converted to gff3 format and then uploaded to a MySQL backend database included with the install. Images for each transcript containing expression levels and number of splice variants were created with in house software.

Future Directions. We plan to expand the Smooth Muscle Transcriptome Sequencing Project to other GI cell types in mice and human GI cell types, as well as update the SMTB.

Availability of Supporting Data

The RNA-seq data from this study have been deposited to the NCBI Gene Expression Omnibus (GEO): GSM1388406, Jejunum SMC; GSM1388407, Colon SMC; GSM1388408, Jejunum ICC; GSM1388409, Colon ICC; GSM1388410, Jejunum PDGFR α^+ cell; GSM1388411, Colon PDGFR α^+ cell; GSM1388412, Jejunum SM; GSM1388413, Colon SM; GSM1388414, Colon Mu; GSM1388415, Colon Mu PDGFR α^+ cell.

References

- Sanders, K. M., Koh, S. D., Ro, S. & Ward, S. M. Regulation of gastrointestinal motility—insights from smooth muscle biology. *Nature reviews. Gastroenterology & hepatology* **9**, 633–645, <https://doi.org/10.1038/nrgastro.2012.168> (2012).
- Sanders, K. M., Koh, S. D. & Ward, S. M. Interstitial cells of cajal as pacemakers in the gastrointestinal tract. *Annual review of physiology* **68**, 307–343, <https://doi.org/10.1146/annurev.physiol.68.040504.094718> (2006).
- Bornstein, J. C., Costa, M. & Grider, J. R. Enteric motor and interneuronal circuits controlling motility. *Neurogastroenterology and motility: the official journal of the European Gastrointestinal Motility Society* **16**(Suppl 1), 34–38, <https://doi.org/10.1111/j.1743-3150.2004.00472.x> (2004).

4. Kurahashi, M., Mutafova-Yambolieva, V., Koh, S. D. & Sanders, K. M. Platelet-derived growth factor receptor-alpha-positive cells and not smooth muscle cells mediate purinergic hyperpolarization in murine colonic muscles. *American journal of physiology. Cell physiology* **307**, C561–570, <https://doi.org/10.1152/ajpcell.00080.2014> (2014).
5. Baker, S. A., Hennig, G. W., Ward, S. M. & Sanders, K. M. Temporal sequence of activation of cells involved in purinergic neurotransmission in the colon. *The Journal of physiology* **593**, 1945–1963, <https://doi.org/10.1113/jphysiol.2014.287599> (2015).
6. Martucciello, G., Ceccherini, I., Lerone, M. & Jasonni, V. Pathogenesis of Hirschsprung's disease. *J Pediatr Surg* **35**, 1017–1025, <https://doi.org/10.1053/jpsu.2000.7763> (2000).
7. Ordog, T., Hayashi, Y. & Gibbons, S. J. Cellular pathogenesis of diabetic gastroenteropathy. *Minerva gastroenterologica e dietologica* **55**, 315–343 (2009).
8. Miettinen, M. & Lasota, J. Gastrointestinal stromal tumors: review on morphology, molecular pathology, prognosis, and differential diagnosis. *Archives of pathology & laboratory medicine* **130**, 1466–1478, [10.1043/1543-2165\(2006\)130\[1466:GASTROM\]2.0.CO;2](https://doi.org/10.1043/1543-2165(2006)130[1466:GASTROM]2.0.CO;2) (2006).
9. Specia, S., Giusti, I., Rieder, F. & Latella, G. Cellular and molecular mechanisms of intestinal fibrosis. *World journal of gastroenterology: WJG* **18**, 3635–3661, <https://doi.org/10.3748/wjg.v18.i28.3635> (2012).
10. Stanghellini, V., Corinaldesi, R. & Barbara, L. Pseudo-obstruction syndromes. *Bailliere's clinical gastroenterology* **2**, 225–254 (1988).
11. Lee, M. Y. *et al.* Smooth Muscle Cell Genome Browser: Enabling the Identification of Novel Serum Response Factor Target Genes. *PLoS one* **10**, e0133751, <https://doi.org/10.1371/journal.pone.0133751> (2015).
12. Ha, S. E. *et al.* Transcriptome analysis of PDGFRalpha+ cells identifies T-type Ca2+ channel CACNA1G as a new pathological marker for PDGFRalpha+ cell hyperplasia. *PLoS one* **12**, e0182265, <https://doi.org/10.1371/journal.pone.0182265> (2017).
13. Lee, M. Y. *et al.* Transcriptome of interstitial cells of Cajal reveals unique and selective gene signatures. *PLoS one* **12**, e0176031, <https://doi.org/10.1371/journal.pone.0176031> (2017).
14. Ha, S. E. *et al.* Transcriptome profiling of subepithelial PDGFRα+ cells in colonic mucosa reveals several cell-selective markers (In review).
15. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology. *Nucleic acids research* **31**, 28–33 (2003).
16. Mouse, E. C. *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome biology* **13**, 418, <https://doi.org/10.1186/gb-2012-13-8-418> (2012).
17. Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic acids research* **41**, D64–69, <https://doi.org/10.1093/nar/gks1048> (2013).
18. Messeguer, X. *et al.* PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**, 333–334 (2002).
19. Pundir, S., Martin, M. J. & O'Donovan, C. UniProt Protein Knowledgebase. *Methods in molecular biology* **1558**, 41–55, https://doi.org/10.1007/978-1-4939-6783-4_2 (2017).
20. Farkas, M. H. *et al.* Transcriptome analyses of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC genomics* **14**, 486, <https://doi.org/10.1186/1471-2164-14-486> (2013).
21. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **34**, 11929–11947, <https://doi.org/10.1523/JNEUROSCI.1860-14.2014> (2014).
22. Lee, M. Y. *et al.* Serum response factor regulates smooth muscle contractility via myotonic dystrophy protein kinases and L-type calcium channels. *PLoS one* **12**, e0171262, <https://doi.org/10.1371/journal.pone.0171262> (2017).
23. Kurahashi, M. *et al.* Platelet-derived growth factor signals play critical roles in differentiation of longitudinal smooth muscle cells in mouse embryonic gut. *Neurogastroenterology and motility: the official journal of the European Gastrointestinal Motility Society* **20**, 521–531, <https://doi.org/10.1111/j.1365-2982.2007.01055.x> (2008).
24. Owens, G. K., Kumar, M. S. & Wamhoff, B. R. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiological reviews* **84**, 767–801, <https://doi.org/10.1152/physrev.00041.2003> (2004).
25. Torihashi, S. *et al.* Blockade of kit signaling induces transdifferentiation of interstitial cells of cajal to a smooth muscle phenotype. *Gastroenterology* **117**, 140–148 (1999).
26. Vanderwinden, J. M. & Rumessen, J. J. Interstitial cells of Cajal in human gut and gastrointestinal disease. *Microscopy research and technique* **47**, 344–360, [https://doi.org/10.1002/\(SICI\)1097-0029\(19991201\)47:5<344::AID-JEMT6>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0029(19991201)47:5<344::AID-JEMT6>3.0.CO;2-1) (1999).
27. Xin, H. B., Deng, K. Y., Rishniw, M., Ji, G. & Kotlikoff, M. I. Smooth muscle expression of Cre recombinase and eGFP in transgenic mice. *Physiological genomics* **10**, 211–215, <https://doi.org/10.1152/physiolgenomics.00054.2002> (2002).
28. Ro, S. *et al.* A model to study the phenotypic changes of interstitial cells of Cajal in gastrointestinal diseases. *Gastroenterology* **138**(1068–1078), e1061–1062, <https://doi.org/10.1053/j.gastro.2009.11.007> (2010).
29. Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D. & Soriano, P. Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. *Molecular and cellular biology* **23**, 4013–4025 (2003).
30. Stein, L. D. Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief Bioinform* **14**, 162–171, <https://doi.org/10.1093/bib/bbt001> (2013).
31. Rogers, A. *et al.* WormBase 2007. *Nucleic acids research* **36**, D612–D617, <https://doi.org/10.1093/nar/gkm975> (2008).
32. Tweedie, S. *et al.* FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic acids research* **37**, D555–559, <https://doi.org/10.1093/nar/gkn788> (2009).
33. Hong, E. L. *et al.* Gene Ontology annotations at SGD: new data sources and annotation methods. *Nucleic acids research* **36**, D577–D581, <https://doi.org/10.1093/nar/gkm909> (2008).
34. Duan, J. *et al.* SilkDBv2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic acids research* **38**, D453–D456, <https://doi.org/10.1093/nar/gkp801> (2010).

Acknowledgements

The authors would like to thank Tsai-wei Shen, Ph.D. (LC Sciences) for the computational analysis of our RNA-seq data, and Duane Wiley (Computer System Administrator at UNR) for installation of Smooth Muscle Transcriptome Browser. This study was supported by the National Institute of Diabetes and Digestive and Kidney Diseases grants (DK094886 and DK103055 to S. Ro, DK091336 to K.M. Sanders, P01DK041315 to K.M. Sanders and S. Ro) and by the Reno School of Medicine (Faculty Development Support Award to S. Ro.) and Department of Cell Biology and Physiology (Departmental Matching Fund). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. NIH grants R01DK094886 and R01DK103055 to S. Ro, P01-DK41315 to K.M. Sanders and S. Ro.

Author Contributions

A.B. and S.R. designed research; A.B. and T.A.G. built the browser; S.E.H., B.G.J., B.J. and S.R. analyzed and generated data; K.M.S. contributed new reagents/analytic tools; B.G.J., A.B. and S.R. wrote the paper. B.G.J., S.R. and K.M.S. revised the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-36607-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019