RNA-Sequencing data supports the existence of novel VEGFA splicing events but not of VEGFAxxxb isoforms

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

SCIENTIFIC REPORTS

RNA-Sequencing data supports the existence of novel VEGFA splicing events but not of VEGFA$_{xxx}$b isoforms

Stephen Bridgett, Margaret Dellett & David A. Simpson

Vascular endothelial growth factor (VEGFA), a pivotal regulator of angiogenesis and valuable therapeutic target, is characterised by alternative splicing which generates three principal isoforms, VEGFA$_{121}$, VEGFA$_{165}$ and VEGFA$_{189}$. A second set of anti-angiogenic isoforms termed VEGFA$_{xxx}$b that utilise an alternative splice site in the final exon have been widely reported, with mRNA detection based principally upon RT-PCR assays. We sought confirmation of the existence of the VEGFA$_{xxx}$b isoforms within the abundant RNA sequencing data available publicly. Whilst sequences derived specifically from each of the canonical VEGFA isoforms were present in many tissues, there were no sequences derived from VEGFA$_{xxx}$b isoforms. Sequencing of approximately 50,000 RT-PCR products spanning the exon 7–8 junction in 10 tissues did not identify any VEGFA$_{xxx}$b transcripts. The absence or extremely low expression of these transcripts *in vivo* indicates that VEGFA$_{xxx}$b isoforms are unlikely to play a role in normal physiology. Our analyses also revealed multiple novel splicing events supported by more reads than previously reported for VEGFA$_{145}$ and VEGFA$_{148}$ isoforms, including three from novel first exons consistent with existing transcription start site data. These novel VEGFA isoforms may play significant roles in specific cell types.

The pivotal role of VEGFA in angiogenesis has been recognised for many years[1] and it is one of the most extensively studied growth factors. Inappropriate angiogenesis is a key factor in a range of conditions including cancer and proliferative eye diseases; blockade of VEGFA signalling to inhibit angiogenesis and reduce vascular leakage forms the basis of multiple effective clinical treatments[2]. Anti-VEGFA therapies can maintain and even improve visual acuity in neovascular age-related macular degeneration[3].

VEGFA provides a paradigm of how alternative splicing can give rise to a set of related proteins with shared domains and variable regions which confer differing biological properties[4–6]. The VEGFA splice isoforms (Fig. 1a) are named according to the number of amino acids they contain (e.g. VEGFA$_{121}$ or VEGFA$_{165}$). They differ in their bioavailability, the shorter isoforms being freely diffusible while the longer isoforms are highly basic and bound by the extracellular matrix[7].

The report that splicing could occur to a downstream site within exon 8 (creating exon 8b) and generate an additional family of anti-angiogenic isoforms dubbed VEGFA$_{xxx}$b[8] added an intriguing further level of complexity to the VEGFA splicing story. Many publications have since described the presence of VEGFA$_{xxx}$b transcripts *in vivo* and defined the anti-angiogenic properties of the proteins they encode[9–11]. Indeed it has been suggested that these isoforms are implicated in human diseases such as systemic sclerosis[10], peripheral artery disease[9] and maternal gestational diabetes[12]. These findings raise the alarming possibility that many studies of VEGFA need to be re-evaluated due to potential overlap between detection of VEGFA$_{xxx}$ or VEGFA$_{xxx}$b isoforms[13], with potential implications for VEGFA-based therapies[14].

All PCR-based assays are prone to artefacts due to the risk of amplification following hybridisation of primers to regions with incomplete complementarity and the validity of the RT-PCR assays which provide most of the evidence for the expression of VEGFA$_{xxx}$b mRNA *in vivo* has been disputed[15]. In response it was suggested that specific experimental conditions and use of appropriate controls are critical for detection of VEGFA$_{xxx}$b isoforms[16].
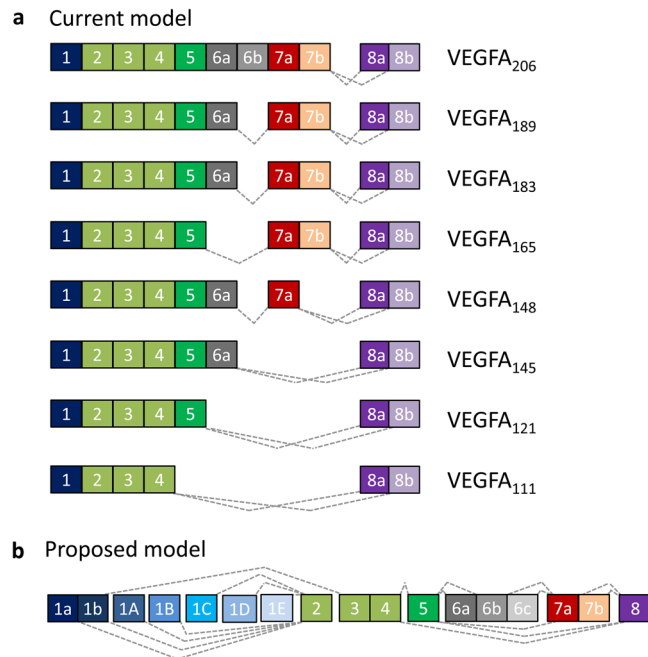
Figure 1. Overview of VEGFA splicing. In the current model exons 1-4 are present in all VEGFA mRNA transcripts. Inclusion or exclusion of different combinations of exons 5 to 7, or variants thereof, creates transcripts encoding VEGFA peptides of differing lengths (indicated by subscript numbers) and biological properties. It has been suggested that use of an alternative 3′ splice site for exon 8 can result in a shorter terminal exon (8b) encoding a peptide of the same length but with a different C-terminus which bestows anti-angiogenic properties. Numerous VEGFA$_{xxx}$b variants have been reported, including VEGFA$_{111b}$[55]; VEGFA$_{165b}$[8]; VEGFA$_{121b}$, VEGFA$_{183b}$, VEGFA$_{145b}$[34]; VEGFA$_{189b}$[56] and VEGFA$_{206b}$[57]. Based upon the analysis of RNA-Seq data presented in this study, we propose an even more complex model for VEGFA transcription which includes alternative first exons and additional splicing events (a novel exon 6 splice site creates exon 6b and the existing exon 6b is renamed 6c). There is a single exon 8 with no splicing to the 8b site.

Although the VEGFA$_{xxx}$b proteins have been overexpressed and demonstrated to be functional, for example reducing KDR (also known as VEGFR-2) phosphorylation[10], this does not bear upon whether they are expressed *in vivo*. Antibodies raised against the VEGFA$_{xxx}$b c-terminus have detected signals in tissues and blood[9, 10, 12, 17–19] but this could be due to translational read-through of exon 8a-containing transcripts[20, 21] or cross-reactivity.

If VEGFA$_{xxx}$b isoforms exist *in vivo* then modulation of their expression provides a very attractive therapeutic approach. Indeed it has been suggested that the balance between pro- and –anti-angiogenic VEGFA$_{xxx}$b isoforms synthesised by cancer cells could be targeted as a therapy[22]. VEGFA$_{xxx}$b proteins have been demonstrated to have anti-angiogenic properties, but if they are not endogenous there is no rationale to limit investigations to these isoforms and other exogenous VEGFA variants may prove to be even more beneficial. It is therefore important to clarify whether VEGFA$_{xxx}$b isoforms exist *in vivo*. High throughput next generation sequencing (NGS) has enabled the transcriptome to be investigated with unprecedented depth (eg ENCODE (Encyclopedia of DNA Elements) Consortium[23] and GTEx[24, 25]) and we reasoned that if VEGFA$_{xxx}$b isoforms do exist *in vivo* they should be represented in RNA sequencing data. Examination of RNA-Seq reads from a wide range of tissues and cancer samples revealed sequences derived from the canonical exon 8a-containing VEGFA isoforms in all datasets, but no sequences indicative of any VEGFA$_{xxx}$b isoform. The absence of sequences spanning the proposed VEGFA exon 8b junction suggests that VEGFA$_{xxx}$b transcripts do not exist or are expressed at extremely lowly levels *in vivo* and therefore that these isoforms are unlikely to play a significant role in normal physiology.

Results

Reports of VEGFA$_{xxx}$b transcripts have relied predominantly upon RT-PCR assays[8, 26], which although very sensitive can be difficult to interpret due to the potential for mis-priming[15]. Seeking independent verification, we reasoned that if the putative VEGFA$_{xxx}$b transcripts exist, they would be represented within the wealth of RNA sequencing data now publicly available[27].

Analysis of Expressed Sequence Tags (ESTs). For many years 'Expressed sequence tags' (ESTs) have been collected by Sanger sequencing of individual cDNA clones isolated from a range of cells and tissues. Alignment of ESTs deposited in Genbank[28] to the human genome using blat[29] identifies all those containing a canonical intron and therefore derived from a spliced transcript. Of 57 spliced ESTs representing the VEGFA gene which include exon 8, 48 contain the standard 8a splice site and there are none with the putative VEGFA$_{xxx}$b exon 8b start site (Fig. 2). One EST, CN256175 represents a rare putative novel isoform similar to transcript variant 6 (NM_001025370; which encodes isoform f with a non-AUG (CUG) translation initiation codon) which is spliced
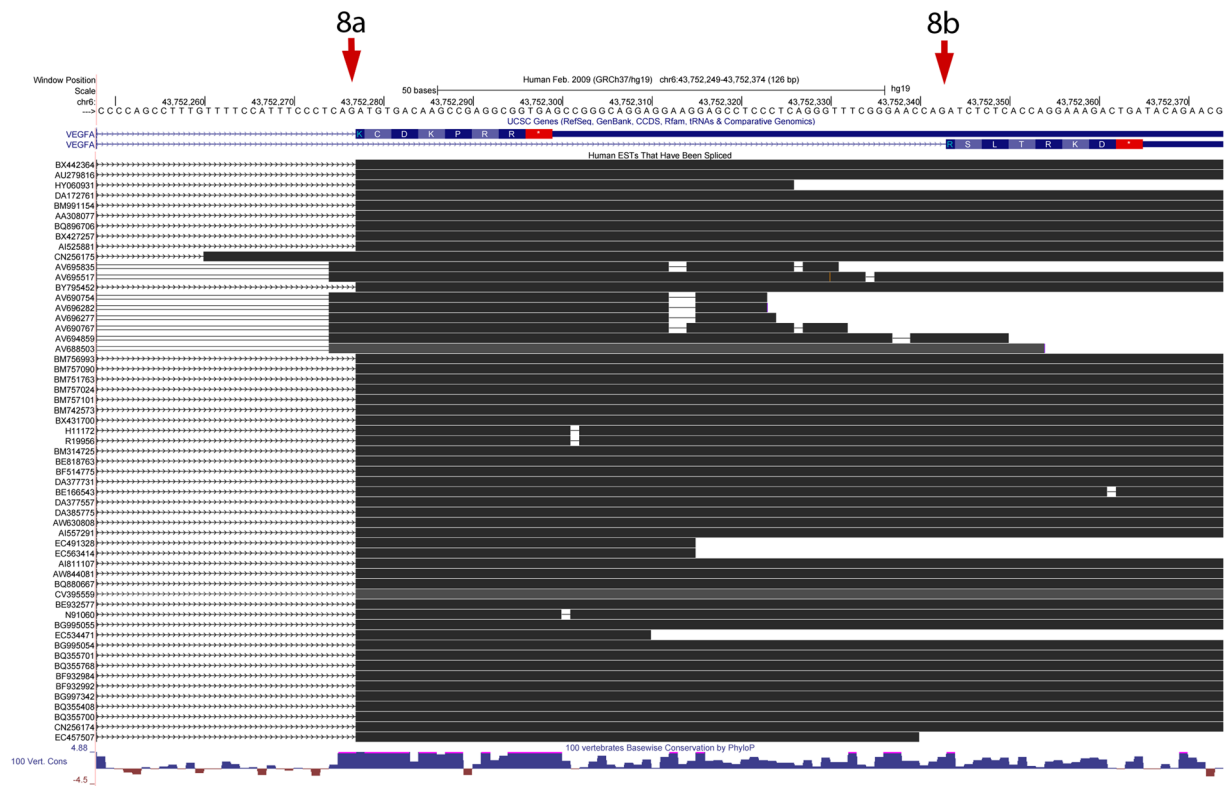
Figure 2. ESTs spanning the exon 8 splice junction viewed in the UCSC genome browser. All available spliced ESTs indicate the exclusive use of exon 8a, with none supporting the existence of exon 8b.

from exon 5 to a position upstream of the usual exon 8a site. Another putative novel upstream site is shared by the remaining 8 ESTs. These are all from the same hepatocellular carcinoma library (LIBEST_006533)[30] and may represent a tumour-related transcript not represented in the cognate library from corresponding non-cancerous liver tissue (LIBEST_005601).

Searches of three publicly available NGS datasets using BLAST and the STAR aligner. It is conceivable that the VEGFA$_{xxx}$b splice site might not be represented within the small number of human ESTs available (<50 containing the canonical exon 8a splice site). Fortunately the advent of NGS has led to the creation of vast repositories of publicly available RNA-Seq data. We analysed three independent datasets comprising sequences from multiple human tissues; (1) 16 tissues from the Illumina Body Map (E-MTAB-513; Bioproject PRJEB2445; Study ERP000546) (adipose, adrenal, brain, breast, colon, heart, kidney, leukocyte, liver, lung, lymph node, ovary prostate, skeletal muscle, testis, thyroid); (2) 27 different tissues from 95 individuals analysed as part of the Human Protein Atlas (www.proteinatlas.org)[31] (ArrayExpress accession E-MTAB-1733; PRJEB4337; ERP003613)(same tissues as data set 1 with addition of appendix, bladder, bone marrow, duodenum, endometrium, esophagus, fallopian tube, gall bladder, pancreas, placenta, rectum, salivary gland, skin, small intestine, smooth muscle, spleen, stomach and tonsil, but without breast or leukocytes); and (3) Selected tissues that contain cells reported to have expression of VEGFA$_{xxx}$b: Normal kidney (reported to be expressed in glomerular podocytes[26, 32, 33]); retina[34]; placenta[35]; adipose (macrophages, reported source of VEGFA$_{xxx}$b in peripheral artery disease[9]); and retinal pigment epithelial (RPE) cells[36–38]. Details of the datasets are provided in Supplementary Dataset 1.

A BLAST database (Supplementary Table 1) was created containing sequences of all the following potential exon junctions which contain the exon 8a or 8b splice sites (VEGFA isoform(s) encoded shown in brackets); 4-8 (VEGFA$_{111}$), 5-8 (VEGFA$_{121}$), 6a-8 (VEGFA$_{145}$), 7a-8 (VEGFA$_{148}$); and 7b-8 (VEGFA$_{165}$, VEGFA$_{183}$, VEGFA$_{189}$, VEGFA$_{206}$). BLAST searches were performed with each NGS read.

BLAST searches of the 5 billion (5×10^9) reads in dataset 1 identified 10,188 reads containing perfect matches to the canonical VEGFA exon 8a splice junction and none to the VEGFA exon 8b junction. Similarly, of a total of 6.9 billion (6.9×10^9) individual reads (3.5 billion pairs) in dataset 2, 31,614 reads matched the canonical VEGFA exon 8a splice junction and none matched the VEGFA exon 8b junction. Of a total of 2.8 billion (2.8×10^9) reads in dataset 3 there were 8,428 matching exon 8a and none matching 8b. Dataset 3 included 6 ribosome profiling samples from kidney[39], in which 11 reads containing the VEGFA exon 8a splice site were identified, but none with the VEGFA exon 8b site. The results of the BLAST analyses are summarised in Table 1 and provided in full in Supplementary Dataset 2. One read containing the VEGFA$_{xxx}$b splice site would have represented 0.002% of the VEGFA transcripts present across all the tissues examined; the failure to detect any suggests that the overall proportion of VEGFA exon 8b transcripts is less than this. It is possible that VEGFAxxxb transcripts are more

Dataset	Source	Num Reads	BLAST hits to VEGFA junctions		STAR alignments to VEGFA junctions	
			8a	8b	8a	8b
1	E-MTAB-513	5×10^9	10,188	0	11,931	0 (24)
2	E-MTAB-2836	6.9×10^9	31,614	0	34,131	0 (56)
3	adipose, kidney, placenta, macrophages, retina, RPE	1.6×10^9	1,380	0	2,406	0 (2)

Table 1. Summary of datasets used, BLAST exact 22-base hits and STAR unique alignments to VEGFA junctions 8a and 8b. Paired reads are counted here as two separate reads. No reads supported existence of an 8b splice sites. Prior to removal of the two VEGFA$_{xxx}$b transcripts from the 'genes.gtf' annotation file a small number of reads (indicated in brackets) were reported to align to the annotated 8b splice site but these had very limited overlap in exon 7 (see Supplementary Fig. 1).

abundant in an individual tissue, but with an average of 1807 reads containing a canonical VEGFA exon 8a splice junction in each of 14 tissues represented in both datasets 1 and 2, the proportion is unlikely to be >0.06% that one read would indicate.

An analysis was performed to reveal the tissue distribution of isoforms across the 14 tissues represented in both datasets and including several with reported high expression of VEGFA$_{xxx}$b isoforms (eg kidney[8] and adipose[40]). The vast majority (>99%) of all reads including an exon 8 splice junction in the 14 tissues common to both datasets included either the exon 5-8a junction characteristic of VEGFA$_{121}$ or the exon 7b-8a junction present in multiple VEGFA isoforms. No reads containing any of the putative VEGFA$_{xxx}$b junctions were detected. The numbers of hits per million reads to each splice junction are shown in Fig. 3a (data for all tissues listed in Supplementary Dataset 2). Fewer reads were detected for the exon 5-8a junction than the exon 7b-8a junction, with the proportion for each tissue generally being consistent between the two independent datasets ($R^2 = 0.26$) and ranging from 0.1–0.8 (Fig. 3c).

Following the failure to identify VEGFA$_{xxx}$b splice junctions by searching for the predicted sequences we employed the alternative strategy of aligning the reads to the hg19 genomic DNA reference using the Spliced Transcripts Alignment to a Reference (STAR) software[41]. Almost two million reads (1,994,279) aligned to the VEGFA region (chr6, from 43737946 to 43754223 on hg19), and of these there were 215,382 unique read alignments to splice events within the VEGFA gene. When the VEGFA$_{xxx}$b transcripts were included in the guide gene.gtf file several reads did align to the VEGFA$_{xxx}$b splice sites. However, on closer inspection these all had minimal overlap with exon 7 due to misalignment of the end three bases of the reads which are common to the end of exon 7 and the end of exon 8a immediately before the start of the proposed exon 8b splice site (see Supplementary Fig. 1 showing Integrated Genome Viewer (IGV) alignment). When the VEGFA$_{xxx}$b transcripts (NM_001171629 and NM_001033756) were removed from the genes.gtf file, and the STAR alignment rerun with the novel splice option for the samples with putative VEGFAxxb sites, no reads aligned to the VEGFA$_{xxx}$b splice site. The results of the STAR alignment to the VEGFA gene are summarised in Table 1 and listed in full in Supplementary Dataset 3. As observed from BLAST searches, more reads aligned to the exon 5-8a junction than to the exon 7b-8a junction (Fig. 3b), with ratios ranging from 0.1 (heart and skeletal muscle) to 0.9 (ovary). The values for the same tissue from the independent datasets were strongly correlated ($R^2 = 0.83$) (Fig. 3c). The greater correlation between the same tissues observed with STAR alignments than with BLAST searches probably reflects the more effective normalisation possible with this approach (to bases aligned to chromosome 6) and the requirement for an exact match across the whole junction sequence with BLAST.

Analysis of Genotype-Tissue Expression (GTEx) Tophat alignments. As part of the Genotype-Tissue Expression (GTEx)[24, 25] project RNA-Seq reads from approximately 7000 samples from 50 tissues (Release V6) have been mapped to the human genome using Tophat). Whilst the relative tissue expression of exon 8a-containing VEGFA transcripts is in good agreement with our analyses of the Illumina Body Map and Human Protein Atlas data, with thyroid having the highest expression in all datasets (Fig. 3 and Supplementary Fig. 2), no reads are reported for either of the ENSEMBL transcripts containing the exon 8b splice site (ENST00000482630 and ENST00000518824).

Sequencing of RT-PCR products spanning exon 7-8 junction. To assess the splicing between exons 7 and 8 with greater sensitivity, total RNA from a range of tissues and HEK293 cells expressing recombinant VEGFA$_{165}$ or VEGFA$_{165}$b transcripts was reverse transcribed from an oligonucleotide specific for exon 8 and PCR then performed with a primer pair located further upstream in exon 8 and in exon 7 (Fig. 4a). Single products of the sizes predicted for VEGFA$_{165}$ (exon 7b-8a) or VEGFA$_{165}$b (exon 7b-8b) were amplified from cells overexpressing the respective transcripts (Fig. 4b). From all tissues a single product consistent with exon 7b-8a was observed following agarose gel electrophoresis (Fig. 4c). However, use of next generation sequencing to determine the sequences of approximately 50,000 products from each PCR reaction revealed that while the majority (>99%) of reads were of the canonical exon 7b-8a splice site present in VEGFA$_{165}$, VEGFA$_{183}$, VEGFA$_{189}$, and VEGFA$_{206}$ a small number were of the shortened exon 7, 7a-8a splice site reported in VEGFA$_{148}$[26] (Fig. 4d). No sequences of splicing events including exon 8b were detected in any of the eleven tissues tested.
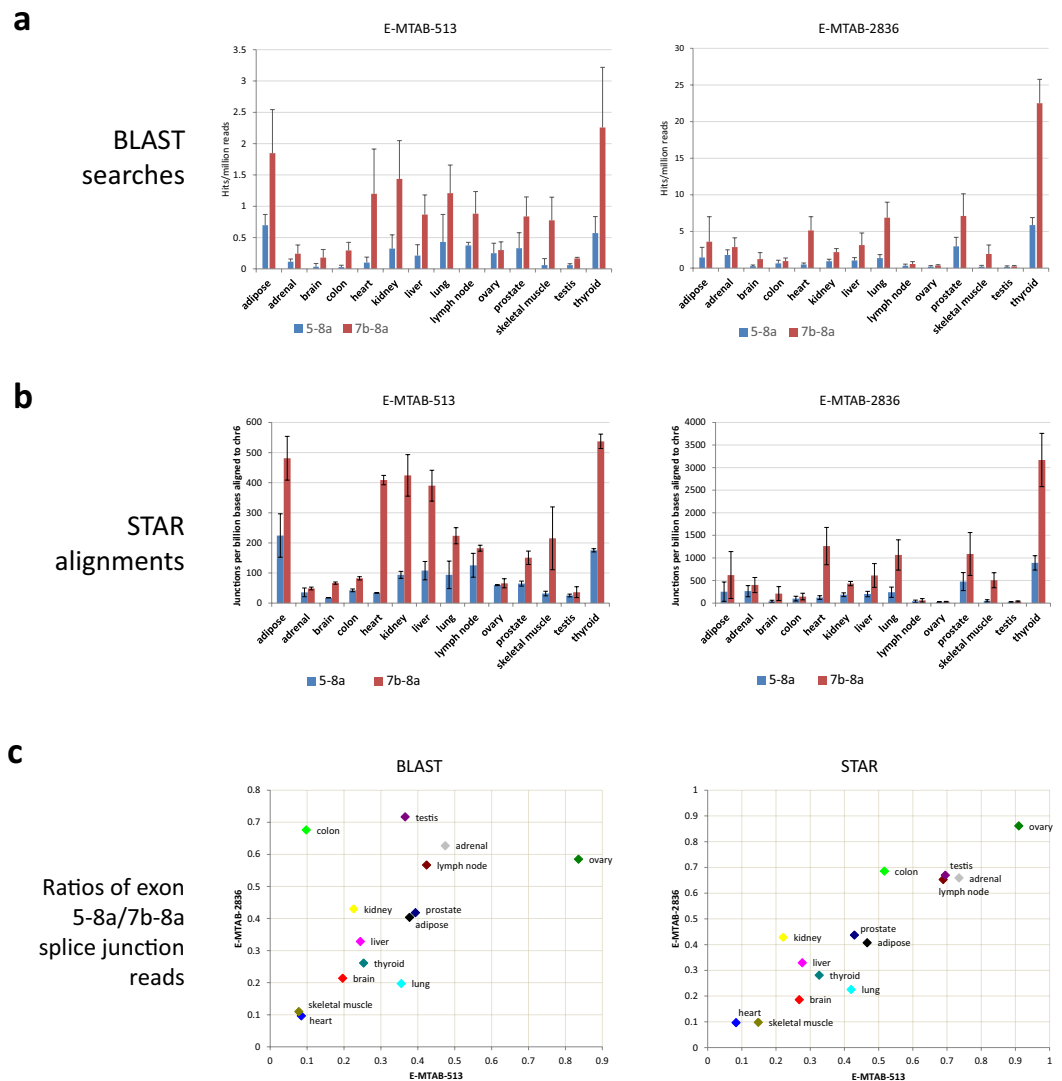
Figure 3. Occurrence of exon 8 splice junction sequences in RNA-Seq data from multiple tissues. Two multi-tissue RNA-Seq datasets (E-MTAB-513 and E-MTAB-2836) were interrogated for the presence of VEGFA exon 8 splice junctions. The number of reads reflecting splicing from exons 4, 6a and 7a were relatively few compared to those for the principal 5-8a and 7b-8a junctions (Supplementary Datasets 2 and 3). No sequences supporting the existence of the 8b site were detected. (**a**) The numbers of reads identified by BLAST searches to contain exon 5-8a and 7b-8a junction sequences are depicted across 14 tissues common to both datasets. (**b**) The number of exon 5-8a and 7b-8a splice junctions detected per billion bases aligned using the STAR aligner across the 14 common tissues. (**c**) The ratios of reads reflecting exon 5-8a: exon 7b-8a splice junctions for each tissue are plotted against the same tissue in the independent dataset. Both BLAST analysis (left) or STAR alignment (right) demonstrate correlation, with R^2 values of 0.26 and 0.83 respectively. This demonstrates that specific tissues have characteristic VEGFA splicing patterns, with the abundance of transcripts containing exon 5-8a splicing relative to those with 7b-8a splicing ranging from 0.1–0.9.

RNA-Seq analysis of cells expressing recombinant VEGFA$_{165}$ or VEGFA$_{165}$b transcripts. RNA from HEK293 cells expressing either recombinant VEGFA$_{165}$ or VEGFA$_{165}$b transcripts was subjected to RNA sequencing. Alignment of the resulting reads to the genome detected the expected 7b-8a or 7b-8b splice sites respectively. A coverage of approximately 20,000 reads along the length of each of the recombinant VEGF transcripts confirms that when over-expressed, both splice variants are detected with similar efficiency by RNA sequencing (Supplementary Fig. 3).

Novel splice sites. Further analysis of the STAR alignments of the public datasets to the VEGFA gene identified numerous novel splice sites, in many cases supported by more reads than several previously annotated splice sites (Supplementary Dataset 3). One novel splice site is within exon 1 and is most frequently spliced to exon 2, creating a shortened exon 1 we have termed exon 1a (Fig. 1b). This site is also supported by several reads in which it is spliced to exon 3 and eight reads in which it is spliced to downstream within exon 1 creating a short
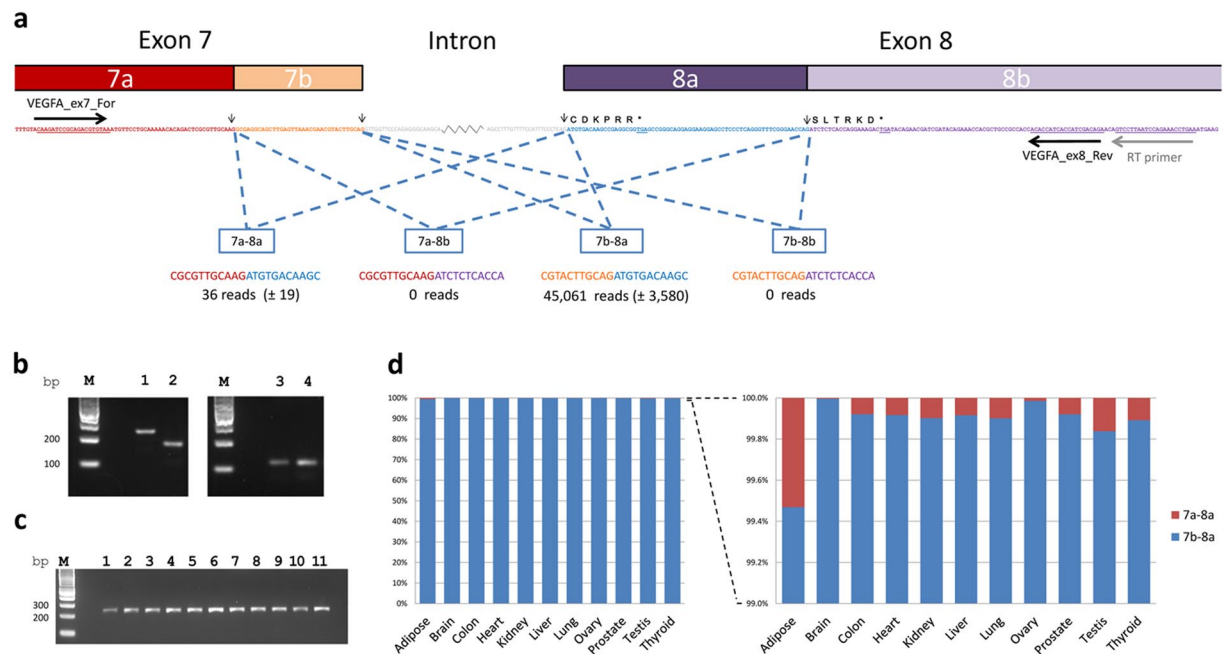
Figure 4. RT-PCR spanning VEGFA exons 7 to 8 and sequencing of amplification products. (**a**) The potential splice junctions between VEGFA exons 7 and 8 are indicated diagrammatically. The mean numbers of reads across all tissues (excluding adipose) containing each splice junction in RT-PCR products amplified using the primers shown in exons 7 and 8 are indicated. (**b**) Single PCR products of the sizes predicted for VEGFA$_{165}$ (exon 7b-8a; 235bp) or VEGFA$_{165}$b (exon 7b-8b; 169bp) were amplified from cells overexpressing the respective transcripts. Lane 1: HEK293 expressing VEGFA$_{165}$, Lane 2: HEK293 expressing VEGFA$_{165}$b, Lane M: Size marker. Products of the same size (115bp) were amplified using primers spanning exons 4 to 5 and common to both transcripts. Lane 3: HEK293 expressing VEGFA$_{165}$, Lane 4: HEK293 expressing VEGFA$_{165b}$. (**c**) The only RT-PCR products amplified by primers in exons 7 and 8 in all tissues and observed by agarose gel electrophoresis are consistent with amplification of exon 7b-8a transcripts (235bp). Lane 1: Adipose, Lane 2: Brain, Lane 3: Colon, Lane 4: Heart, Lane 5: Kidney, Lane 6: Liver, Lane 7: Lung, Lane 8: Ovary, Lane 9: Prostate, Lane 10: Testis, Lane 11: Thryoid, Lane M: Size marker. (**d**) Percentage of reads from transcripts with 7a-8a and 7b-8a splice events.

intron and a transcript with CDS predicted to start at the canonical ATG. Of the novel splice junctions supported by more than 20 uniquely mapped reads four are between putative novel first exons and exon 2. The existence of the putative novel first exons is supported by the presence of transcript start sites just upstream of the pre-dicted splice sites (Fig. 5). These were determined by Transcriptional Start Site-Sequencing (TSS-Seq), which uses oligo-capping to determine transcript 5′ ends[42] and are reported in the DataBase of Transcriptional Start Sites (DBTSS: http://dbtss.hgc.jp/). The exon 1a splice site is before the first potential start codon and the putative novel first exons are likely to be non-coding (they either have no ATG codon or ones located very near the beginning of the exon which create ORFs differing from the canonical VEGFA peptide in subsequent exons and leading to premature termination). The CDS of these transcripts is therefore likely to start in exon 2, with the encoded VEGFA isoforms having a shorter N-terminus lacking a signal peptide and presumably retained within the cell.

Multiple reads supporting the novel exon 1B-2 splicing event were detected in heart and kidney (Supplementary Dataset 3) and are matched by several potential cognate TSS's in adult heart and kidney and many more in foetal heart and kidney[43] (Fig. 5). Other novel splice junctions were detected between exons 1 and 3 and between exon 5 and a novel site within exon 6, 8 nucleotides downstream from the canonical 5′ site, although both events alter the reading frame and are predicted to cause premature termination of translation. The novel splicing events have been incorporated into an updated model of VEGFA splicing (Fig. 1b). The novel rare splice variants do not correlate with overall expression of VEGFA (Supplementary Fig. 4) and therefore are unlikely to reflect 'mis-splicing' of a fixed small fraction of all transcripts. The tissue specificity of the novel splicing events, together with previously reported splice junctions within the same range of expression, is illustrated in Fig. 6.

Discussion

The use of alternative splicing to generate VEGFA protein isoforms with varying bioavailability is a key mech-anism in the control of vascular development and function. The suggestion that additional alternative splicing to generate antiangiogenic VEGFA$_{xxx}$b isoforms provides a further layer of regulation for this pivotal growth factor is therefore very plausible. However the analyses of RNA sequencing data presented in this study suggest that VEGFA$_{xxx}$b transcripts are present at extremely low concentrations or do not exist *in vivo*. It seems unlikely that the absence of reads spanning the putative exon 8b junction is due to technical reasons associated with the laboratory procedures. Although all sequences are not represented equally by RNA-Seq reads due to the
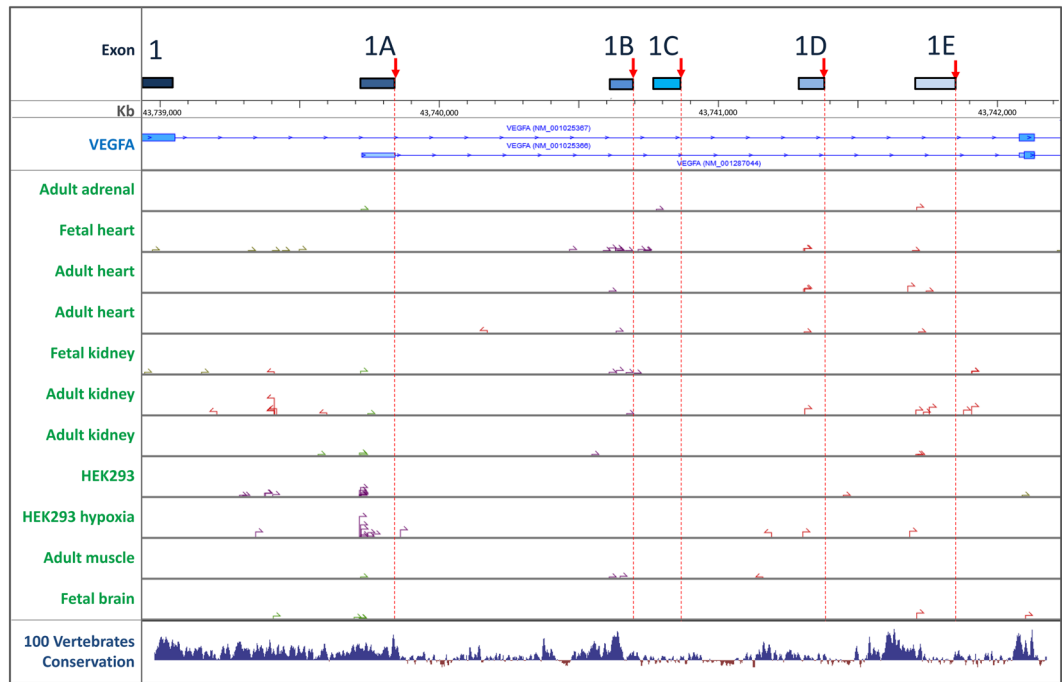
Figure 5. Novel exon 1 splice sites and predicted transcription start sites. Positions of novel splice sites are indicated by red arrows and dotted red lines and transcription start sites (TSS) from selected tissues by arrows.

influences of factors such as ligation bias[44] and varying reverse transcription efficiency, reads spanning exon 8b would be derived from RNAs with varying GC contents and a wide range of secondary structures, particularly if spliced to different upstream exons; therefore at least some would be amenable to sequencing. The detection using RNA-Seq of a similar number of reads spanning the exon 7b-8a and exon 7b-8b splice junctions from recombinant $VEGF_{165}$ and $VEGFA_{165b}$ constructs respectively confirms that both variants are detectable with similar efficiency using this technology. The exon 8a and exon 8b splice sites were also detected with comparable efficiency by RT-PCR, further confirming a lack of stable secondary structures which might block reverse transcription. Other factors which might cause mild under-representation of certain sequences differ between protocols and yet no $VEGFA_{xxx}b$ reads were observed in libraries prepared from multiple samples and in different laboratories. Even in tissues reported to have high levels of $VEGFA_{xxx}b$ protein (up to 45% of total VEGFA protein) such as kidney[33] and placenta[35], no exon 8b splice sites were detected.

The use of two complementary approaches, BLAST searches for predefined short exon junction sequences and STAR spliced alignment of reads to genome sequence, provide high confidence that exon 8b-spanning reads were not missed during the bioinformatic analyses. The value of this dual approach was demonstrated by the observation of multimapping reads mapping to the shared 3 bp sequence at the end of exon 7 and exon 8a (when the putative exon 8b splice site was provided in the STAR index guide file). Although this was detected as an artefact due to the lack of reads with longer overlap, the issue did not arise in the BLAST analysis. Furthermore, a different aligner (TopHat) was used in the analysis of the GTEX data and no $VEGFA_{xxx}b$-containing reads were identified. It is possible that $VEGFA_{xxx}b$ isoforms are expressed at levels below the limit of detection of the available RNA-Seq datasets. However, the depth provided by the combined RT-PCR and sequencing approach enabled robust detection of the rare ($<0.1\%$) exon 7a-exon 8 splicing event characteristic of $VEGFA_{148}$, but still did not identify exon 8b splicing. If $VEGFA_{165}b$ does exist below this level of detection its biological significance is most likely limited.

A re-evaluation of the literature is required in light of these findings, which suggest that $VEGFA_{xxx}b$ transcripts do not exist *in vivo*. Firstly, antibodies specific for the $VEGFA_{xxx}b$ C-terminal peptide have detected antigen across a range of tissues[11, 20, 45, 46]. Secondly, a number of studies have suggested that $VEGFA_{165}b$ has anti-angiogenic properties. Recently overexpression of $VEGFA_{165}b$ in adipose issue has been linked with impaired angiogenesis[40]. Although VEGFA was highly expressed in adipose tissue according to RNA-Seq data, we detected no sequences specific for $VEGFA_{165}b$ transcripts. Similarly it has been reported that $VEGFA_{165}b$ is upregulated in peripheral artery disease[9]. Overexpression of any truncated or modified protein might cause dominant negative effects upon the endogenous protein, and this could be the explanation for the observed anti-angiogenic effects of $VEGFA_{165}b$. However, the demonstration by Eswarappa, S. M. *et al*.[21] that read-through of the VEGFA stop codon leads to translation of an extended protein (VEGF-Ax) with the same C-terminus as the putative $VEGFA_{xxx}b$ isoforms provides an alternative explanation. VEGF-Ax could be responsible for the signal observed with antibodies raised against the $VEGFA_{xxx}b$ terminal peptide[21]. VEGF-Ax has been shown to be antiangiogenic[20, 21] and $VEGFA_{xxx}b$ may exert its anti-angiogenic effects through a similar mechanism, indeed both proteins are defective in binding to neuropilin-1[21, 47].
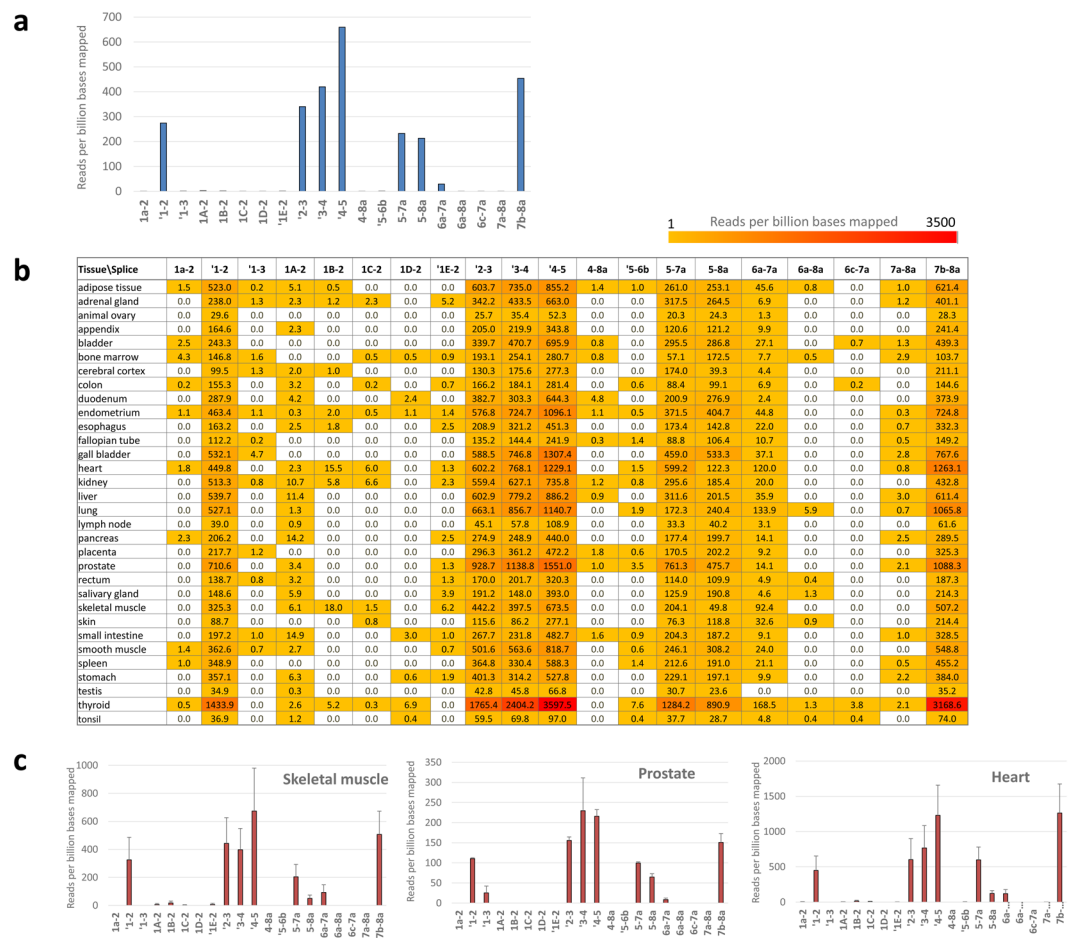
Figure 6. Occurrence of novel splicing events across different tissues. (**a**) All splice sites supported by 20–200 uniquely aligned reads from all datasets combined are listed and the number of reads per million aligned to each indicated. (**b**) The pattern of splice site use across different tissues (dataset 2) is indicated by the number of reads containing each site per billion bases mapped. (**c**) Examples of differential splice site use across tissues.

Tissue\Splice	1a-2	'1-2	'1-3	1A-2	1B-2	1C-2	1D-2	'1E-2	'2-3	'3-4	'4-5	4-8a	'5-6b	5-7a	5-8a	6a-7a	6a-8a	6c-7a	7a-8a	7b-8a
adipose tissue	1.5	523.0	0.2	5.1	0.5	0.0	0.0	0.0	603.7	735.0	855.2	1.4	1.0	261.0	253.1	45.6	0.8	0.0	1.0	621.4
adrenal gland	0.0	238.0	1.3	2.3	1.2	2.3	0.0	5.2	342.2	433.5	663.0	0.0	0.0	317.5	264.5	6.9	0.0	0.0	1.2	401.1
animal ovary	0.0	29.6	0.0	0.0	0.0	0.0	0.0	0.0	25.7	35.4	52.3	0.0	0.0	20.3	24.3	1.3	0.0	0.0	0.0	28.3
appendix	0.0	164.6	0.0	2.3	0.0	0.0	0.0	0.0	205.0	219.9	343.8	0.0	0.0	120.6	121.2	9.9	0.0	0.0	0.0	241.4
bladder	2.5	243.3	0.0	0.0	0.0	0.0	0.0	0.0	339.7	470.7	695.9	0.8	0.0	295.5	286.8	27.1	0.0	0.7	1.3	439.3
bone marrow	4.3	146.8	1.6	0.0	0.0	0.5	0.5	0.9	193.1	254.1	280.7	0.8	0.0	57.1	172.5	7.7	0.5	0.0	2.9	103.7
cerebral cortex	0.0	99.5	1.3	2.0	1.0	0.0	0.0	0.0	130.3	175.6	277.3	0.0	0.0	174.0	39.3	4.4	0.0	0.0	0.0	211.1
colon	0.2	155.3	0.0	3.2	0.0	0.2	0.0	0.7	166.2	184.1	281.4	0.0	0.6	88.4	99.1	6.9	0.0	0.2	0.0	144.6
duodenum	0.0	287.9	0.0	4.2	0.0	0.0	2.4	0.0	382.7	303.3	644.3	4.8	0.0	200.9	276.9	2.4	0.0	0.0	0.0	373.9
endometrium	1.1	463.4	1.1	0.3	2.0	0.5	1.1	1.4	576.8	724.7	1096.1	1.1	0.5	371.5	404.7	44.8	0.0	0.0	0.3	724.8
esophagus	0.0	163.2	0.0	2.5	1.8	0.0	0.0	2.5	208.9	321.2	451.3	0.0	0.0	173.4	142.8	22.0	0.0	0.0	0.7	332.3
fallopian tube	0.0	112.2	0.2	0.0	0.0	0.0	0.0	0.0	135.2	144.4	241.9	0.3	1.4	88.8	106.4	10.7	0.0	0.0	0.5	149.2
gall bladder	0.0	532.1	4.7	0.0	0.0	0.0	0.0	0.0	588.5	746.8	1307.4	0.0	0.0	459.0	533.3	37.1	0.0	0.0	2.8	767.6
heart	1.8	449.8	0.0	2.3	15.5	6.0	0.0	1.3	602.2	768.1	1229.1	0.0	1.5	599.2	122.3	120.0	0.0	0.0	0.8	1263.1
kidney	0.0	513.3	0.8	10.7	5.8	6.6	0.0	2.3	559.4	627.1	735.8	1.2	0.0	295.6	185.4	20.0	0.0	0.0	0.0	432.8
liver	0.0	539.7	0.0	11.4	0.0	0.0	0.0	0.0	602.9	779.2	886.2	0.9	0.0	311.6	201.5	35.9	0.0	0.0	3.0	611.4
lung	0.0	527.1	0.0	1.3	0.0	0.0	0.0	0.0	663.1	856.7	1140.7	0.0	1.9	172.3	240.4	133.9	5.9	0.0	0.7	1065.8
lymph node	0.0	39.0	0.0	0.9	0.0	0.0	0.0	0.0	45.1	57.8	108.9	0.0	0.0	33.3	40.2	3.1	0.0	0.0	0.0	61.6
pancreas	2.3	206.2	0.0	14.2	0.0	0.0	0.0	2.5	274.9	248.9	440.0	0.0	0.0	177.4	199.7	14.1	0.0	0.0	2.5	289.5
placenta	0.0	217.7	1.2	0.0	0.0	0.0	0.0	0.0	296.3	361.2	472.2	1.8	0.6	170.5	202.2	9.2	0.0	0.0	0.0	325.3
prostate	0.0	710.6	0.0	3.4	0.0	0.0	0.0	1.3	928.7	1138.8	1551.0	1.0	3.5	761.3	475.7	14.1	0.0	0.0	2.1	1088.3
rectum	0.0	138.7	0.8	3.2	0.0	0.0	0.0	1.3	170.0	201.7	320.3	0.0	0.0	114.0	109.9	4.9	0.4	0.0	0.0	187.3
salivary gland	0.0	148.6	0.0	5.9	0.0	0.0	0.0	3.9	191.2	148.0	393.0	0.0	0.0	125.9	190.8	4.6	1.3	0.0	0.0	214.3
skeletal muscle	0.0	325.3	0.0	6.1	18.0	1.5	0.0	6.2	442.2	397.5	673.5	0.0	0.0	204.1	49.8	92.4	0.0	0.0	0.0	507.2
skin	0.0	88.7	0.0	0.0	0.0	0.8	0.0	0.0	115.6	86.2	277.1	0.0	0.0	76.3	118.8	32.6	0.9	0.0	0.0	214.4
small intestine	0.0	197.2	1.0	14.9	0.0	0.0	3.0	1.0	267.7	231.8	482.7	1.6	0.9	204.3	187.2	9.1	0.0	0.0	1.0	328.5
smooth muscle	1.4	362.6	0.7	2.7	0.0	0.0	0.0	0.7	501.6	563.6	818.7	0.0	0.6	246.1	308.2	24.0	0.0	0.0	0.0	548.8
spleen	1.0	348.9	0.0	0.0	0.0	0.0	0.0	0.0	364.8	330.4	588.3	0.0	1.4	212.6	191.0	21.1	0.0	0.0	0.5	455.2
stomach	0.0	357.1	0.0	6.3	0.0	0.0	0.6	1.9	401.3	314.2	527.8	0.0	0.0	229.1	197.1	9.9	0.0	0.0	2.2	384.0
testis	0.0	34.9	0.0	0.3	0.0	0.0	0.0	0.0	42.8	45.8	66.8	0.0	0.0	30.7	23.6	0.0	0.0	0.0	0.0	35.2
thyroid	0.5	1433.9	0.0	2.6	5.2	0.3	6.9	0.0	1765.4	2404.2	3597.5	0.0	7.6	1284.2	890.9	168.5	1.3	3.8	2.1	3168.6
tonsil	0.0	36.9	0.0	1.2	0.0	0.0	0.0	0.0	59.5	69.8	97.0	0.0	0.4	37.7	28.7	4.8	0.4	0.4	0.0	74.0

Use of antibodies against the C-terminal peptide encoded by VEGFA exon 8a has been suggested as an isoform-specific therapy that would not target anti-angiogenic VEGFA$_{xxx}$b isoforms. Although likely effective it would now be more pertinent to confirm lack of reactivity with VEGF-Ax than VEGFA$_{xxx}$b. It has been reported that the serine arginine protein kinase (SRPK) inhibitor, SRPIN340, modulates splicing of the VEGFA pre-mRNA to reduce formation of VEGFA$_{165}$ with a concomitant relative increase in VEGFA$_{165}$b[48]. Similar modulation of splicing to exon 8b has been proposed as an approach to reduce neuropathic pain[49]. mRNA-Seq should be performed in these models to determine whether VEGFA$_{165}$b is present.

Several novel putative first exons, each supported by more splice junction reads than were observed for the previously characterised VEGFA$_{148}$ were identified. Although relatively rare, these variants exhibited tissue-specific expression and may be highly expressed in a small subset of cells. The exquisite expression patterns could be regulated through specific promoters which are active only in particular cell types. The function of the encoded proteins, which lack the N-terminal region present in canonical VEGFA proteins, is unclear but could be significant for such cells.

The existence and anti-angiogenic properties of VEGFA$_{xxx}$b protein isoforms with an alternative C-terminus have been widely reported. The lack of RNA-Seq evidence for the cognate transcripts within approximately fifteen billion (14.7×10^9) reads suggests that they do not exist or are present at extremely low concentrations *in vivo*. However, many of these observations can be attributed to the presence of VEGF-Ax proteins which have the same C terminus due to programmed translational readthrough rather than alternative splicing. The physiological implications of anti-angiogenic VEGFA isoforms and potential therapeutic applications therefore remain valid, but with the focus shifting to VEGF-Ax. The characterisation of additional VEGFA splice variants from RNA-Seq data, notably novel first exons, underlines the role of alternative promoters and splicing in delivering the exquisite regulation required for this pivotal gene.

Methods

Nomenclature.   VEGFA exons are referred to as depicted in Fig. 1a and the co-ordinates in human reference hg19 are: Exon 1a 43737946–43738362; 1b 43738363–43739049; 1A 43739722–43739838; 1B? –43740695; 1C? –43740870; 1D?– 43741375; 2 43742078–43742129; 3 43745206–43745402; 4 43746197–43746273; 5

43746626–43746655; 6a 43748469–43748522; 6b 43748523–43748540; 6c 43748541–43748591; 7a 43749693–43749789; 7b 43749790–43749824; 8a 43752278–43752343; 8b 43752344–43754223.

Splice junctions were named according to the donor and acceptor exons. Junction 7b-8a is present in transcripts encoding the $VEGFA_{165}$, $VEGFA_{183}$, $VEGFA_{189}$ and $VEGFA_{206}$ isoforms. Junction 5-8a is present in transcripts encoding the $VEGFA_{121}$ isoform.

Datasets. RNA-Seq datasets were identified in Gene Expression Omnibus (GEO)[50] (http://www.ncbi.nlm.nih.gov/geo/) or ArrayExpress (http://www.ebi.ac.uk/arrayexpress/). The two principal datasets comprised: 16 human tissues from the Illumina bodyMap2 transcriptome (http://www.ncbi.nlm.nih.gov/bioproject/204271) (dataset 1) and tissue samples of 95 human individuals representing 27 different tissues[31] (dataset 2). Dataset 3 contained additional RNA sequencing datasets from a range of studies from kidney, retina, placenta, adipose tissue, macrophages and RPE[36–38]. These 3 datasets are summarised in Table 1 and in more detail in Supplementary Dataset 1. Data for the VEGFA gene was downloaded from the GTEx portal (http://www.gtexportal.org/home/gene/VEGFA). Expressed sequence tags (ESTs) with evidence of splicing were viewed using the UCSC Genome browser (http://genome-euro.ucsc.edu). Transcription start sites were identified from the Database of Transcriptional Start Sites (DBTSS: http://dbtss.hgc.jp/), an integrative platform for transcriptome, epigenome and genome sequence variation data[51].

Data analysis. *Blastn search for splice sites.* A fasta file of all ten VEGFA splice junctions including exon 8a or 8b was created with each sequence of length 22 bases (ie. 11 bases either side of the splice site), see Supplementary data Table 1). This was indexed using the "*formatdb*" command from a locally installed copy of the NCBI blast-all version 2.2.26 (NCBI *blastall* – ftp.ncbi.nlm.nih.gov/blast/executables/blast/LATEST/)[52]. Fastq files of Illumina RNAseq reads (as listed in Supplementary dataset 1), were downloaded from the EBI's sequence read archive: 16 human tissues, Illumina bodyMap2 transcriptome: http://www.ncbi.nlm.nih.gov/bioproject/204271) and RNA-seq of coding RNA from tissue samples of 95 human individuals representing 27 different tissues[31]. These fastq files of reads were converted to fasta format using a perl script. *Blastn* (from the locally installed NCBI *blastall*) was used to search the exon junctions file (described above) for matches to these reads that had an e-value cutoff 1e-6 or better, with complexity filtering disabled (ie. *blastall* options: *-p blastn -e 1e-6 -F F -m 8*). Read 1 and read 2 from the paired-end reads were used independently. To compare between the different runs and tissues, the numbers of hits were normalised to reads per million in each run. A perl script was used to summarise the results from all runs as a text file which was subsequently imported into Excel for viewing and plotting.

Splice-aware alignment to human reference. The STAR RNAseq aligner (version 2.4.0f1)[41] was used to align the Illumina reads (obtained as in point 1 above) to the human reference hg19 chromosome 6, using options:

STAR–runMode alignReads–genomeDir $HG19chr6–runThreadN 4–genomeLoad LoadAndRemove–readFilesIn $IN1.fastq $IN2.fastq–readFilesCommand zcat–outSAMmode Full–outSAMattributes Standard–outFileNamePrefix $OUT.

A Perl script (available on request) was written to extract the splice sites within the VEGFA gene (nucleotide position 43737945 to 43754224 on chromosome 6 of the hg19 genome) from all the splice-junction files produced by STAR, summarising these in text files subsequently imported into Excel (see Supplementary Dataset 3) for viewing and plotting. To compare between the different runs and tissues, numbers of VEGFA splice sites were normalised as 'per billion bases aligned to chromosome 6'. Using samtools[53], the bam files of STAR alignments were sorted by position and indexed, to view the alignments in the IGV browser[54].

Generation of recombinant VEGF expression plasmids. Sequences encoding the $VEGFA_{165}$ or $VEGFA_{165}$b open reading frames and 205 nucleotides of the 3′ UTR were synthesised and inserted into pcDNA3.1(+)(Thermo Fisher) at the *Hin*dIII/*Eco*RI restriction sites (see Supplementary Dataset 4). The plasmid DNA was purified from transformed bacteria (*E. coli* K12 DH10B™ T1R) and concentration determined by UV spectroscopy. The final construct was verified by Sanger sequencing, confirming 100% sequence congruence within the insertion sites.

Cell culture and RNA extraction. HEK293 cells were cultured in DMEM (Thermo Fisher) supplemented with 10% fetal calf serum (FCS) and 100 μg/ml PrimocinTM (InvivoGen). Transfections were performed with 1 μg plasmid using lipofectamine 3000 (Thermo Fisher) and following the manufacturer's instructions for a 12 well plate. RNA extractions were performed after 24 hours using a miRCURY RNA Isolation Kit (Exiqon) and following the manufacturer's instructions.

RT-PCR and NGS. Total human RNA (FirstChoice Human Total RNA Survey Panel, Life Technologies) (1 μg) or RNA from HEK293 cells was reverse transcribed at 50 °C using an oligonucleotide (TTCAGG TTTCTGGATTAAGGAC) specific for VEGFA exon 8 (Fig. 4) and Superscript III (Life Technologies). PCR was performed with primers spanning the exon7-8 splice junction (VEGFA_ex7_For1 CAAGATCCGC AGACGTGTAA; VEGFA_ex8_Rev TCTGTCGATGGTGATGGTGT) or within exons 4 and 5 of $VEGFA_{165}$ and $VEGFA_{165}$b (VEGF_universal_For TGCGGGATCAAACCTCACCAA; VEGF_universal_Rev GGCCCACA GGGATTTTCTTG). PCR amplification was performed using Platinum PCR SuperMix High Fidelity Taq (Life Technologies) with an initial denaturation at 95 °C for 3 min, followed by 40 cycles of 95 °C for 30 s, 55 °C for 30 s and 68 °C for 1 min. Amplicon size was verified by gel electrophoresis on a 2% agarose gel.

Libraries were prepared from PCR products using the Ion Plus fragment library kit (Life Technologies) following the manufacturer's instructions (MAN0006846) version A.0 with slight modifications. Pooling was not performed until after the libraries were barcoded using the Ion express barcode adaptor kit (Life Technologies). All purification steps were performed using Agencourt AMPure XP Reagent (Beckman Coulter) as indicated in the

protocol. Libraries were quantified using the PerfeCTa NGS Quantification Kit for Ion Torrent libraries (Quanta Biosciences). Template preparation was performed on the OneTouch 2 system (Life Technologies) using the Ion PGM Template OT2 200 Kit and protocol version A.0 as recommended by the manufacturer's instructions (Life Technologies). Sequencing was performed on Ion Torrent PGM (Life Technologies) using the Ion PGM HI-Q Sequencing kit and an Ion 314 Chip Kit V2, following the manufacturer's instructions (version B.0).

TruSeq stranded mRNA-seq libraries were prepared from total RNA by Edinburgh Genomics, The University of Edinburgh and sequenced on a MiSeq (v2 100PE).

References

1. Ferrara, N., Gerber, H.-P. & LeCouter, J. The biology of VEGF and its receptors. *Nat. Med.* **9**, 669–76 (2003).
2. Carmeliet, P. & Jain, R. K. Molecular mechanisms and clinical applications of angiogenesis. *Nature* **473**, 298–307 (2011).
3. Solomon, S. D., Lindsley, K., Vedula, S. S., Krzystolik, M. G. & Hawkins, B. S. Anti-vascular endothelial growth factor for neovascular age-related macular degeneration. *Cochrane Database Syst Rev* **8**, CD005139 (2014).
4. Arcondéguy, T., Lacazette, E., Millevoi, S., Prats, H. & Touriol, C. VEGF-A mRNA processing, stability and translation: a paradigm for intricate regulation of gene expression at the post-transcriptional level. *Nucleic Acids Res.* **41**, 7997–8010 (2013).
5. Farrokh, S., Brillen, A.-L., Haendeler, J., Altschmied, J. & Schaal, H. Critical Regulators of Endothelial Cell Functions: For a Change Being Alternative. *Antioxidants & redox signaling* **22(14)**, 1212–29 (2015).
6. Muñoz-Moreno, L. *et al.* Inhibitory effects of antagonists of growth hormone-releasing hormone on growth and invasiveness of PC3 human prostate cancer. *Int. J. Cancer* **132**, 755–65 (2013).
7. Park, J. E., Keller, G.-A. & Ferrara, N. The vascular endothelial growth factor (VEGF) isoforms: differential deposition into the subepithelial extracellular matrix and bioactivity of extracellular matrix-bound VEGF. *Molecular biology of the cell* **4**, 1317–1326 (1993).
8. Bates, D. O. *et al.* VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma. *Cancer Res.* **62**, 4123–31 (2002).
9. Kikuchi, R. *et al.* An antiangiogenic isoform of VEGF-A contributes to impaired vascularization in peripheral artery disease. *Nature medicine* **20(12)**, 1464–71 (2014).
10. Manetti, M. *et al.* Overexpression of VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, leads to insufficient angiogenesis in patients with systemic sclerosis. *Circ. Res.* **109**, e14–26 (2011).
11. Woolard, J. *et al.* VEGF165b, an Inhibitory Vascular Endothelial Growth Factor Splice Variant Mechanism of Action, *In vivo* Effect On Angiogenesis and Endogenous Protein Expression. *Cancer research* **64**, 7822–7835 (2004).
12. Sultan, S. A. *et al.* The Role of Maternal Gestational Diabetes in Inducing Fetal Endothelial Dysfunction. *Journal of cellular physiology* **230(11)**, 2695–705 (2015).
13. Dokun, A. O. & Annex, B. H. The VEGF165b "ICE-o-form" puts a chill on the VEGF story. *Circ. Res.* **109**, 246–7 (2011).
14. Rennel, E. S., Harper, S. J. & Bates, D. O. Therapeutic potential of manipulating VEGF splice isoforms in oncology. *Future Oncol* **5**, 703–12 (2009).
15. Harris, S. *et al.* Do anti-angiogenic VEGF (VEGFxxxb) isoforms exist? A cautionary tale. *PLoS ONE* **7**, e35231 (2012).
16. Bates, D. O. *et al.* Detection of VEGF-A(xxx)b isoforms in human tissues. *PLoS ONE* **8**, e68399 (2013).
17. Bunni, J. *et al.* Circulating levels of anti-angiogenic VEGF-A isoform (VEGF-Axxxb) in colorectal cancer patients predicts tumour VEGF-A ratios. *American journal of cancer research* **5**, 2083 (2015).
18. Baba, T. *et al.* VEGF 165 b in the developing vasculatures of the fetal human eye. *Dev. Dyn.* **241**, 595–607 (2012).
19. Kim, J.-W. *et al.* Clinical Implications of VEGF, TGF-β1, and IL-1β in Patients with Advanced Non-small Cell Lung Cancer. *Cancer Res Treat* **45**, 325–33 (2013).
20. Eswarappa, S. M. & Fox, P. L. Antiangiogenic VEGF-Ax: A New Participant in Tumor Angiogenesis. *Cancer Research* **75**, 2765–2769 (2015).
21. Eswarappa, S. M. *et al.* Programmed translational readthrough generates antiangiogenic VEGF-Ax. *Cell* **157**, 1605–18 (2014).
22. Harper, S. J. & Bates, D. O. VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat. Rev. Cancer* **8**, 880–7 (2008).
23. The ENCODE (Encyclopedia of DNA Elements) Consortium. Available at: https://www.encodeproject.org/. (Date of access: 26/05/2016).
24. Consortium, T. Gte. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**, 580–585 (2013).
25. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
26. Whittle, C., Gillespie, K., Harrison, R., Mathieson, P. W. & Harper, S. J. Heterogeneous vascular endothelial growth factor (VEGF) isoform mRNA and receptor mRNA expression in human glomeruli, and the identification of VEGF148 mRNA, a novel truncated splice variant. *Clin. Sci.* **97**, 303–12 (1999).
27. Li, W., Dai, C., Kang, S. & Zhou, X. J. Integrative analysis of many RNA-seq datasets to study alternative splicing. *Methods* **67**, 313–324 (2014).
28. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank: update. *Nucleic Acids Res.* **32**, D23–6 (2004).
29. Kent, W. J. BLAT-the BLAST-like alignment tool. *Genome Res.* **12**, 656–64 (2002).
30. Xu, X. R. *et al.* Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc. Natl. Acad. Sci. USA* **98**, 15089–94 (2001).
31. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell Proteomics* **13**, 397–406 (2014).
32. Cui, T.-G. *et al.* Differentiated human podocytes endogenously express an inhibitory isoform of vascular endothelial growth factor (VEGF165b) mRNA and protein. *American Journal of Physiology-Renal Physiology* **286**, F767–F773 (2004).
33. Bevan, H. S. *et al.* The alternatively spliced anti-angiogenic family of VEGF isoforms VEGFxxxb in human kidney development. *Nephron Physiol* **110**, p57–67 (2008).
34. Perrin, R. *et al.* Diabetic retinopathy is associated with a switch in splicing from anti-to pro-angiogenic isoforms of vascular endothelial growth factor. *Diabetologia* **48**, 2422–2427 (2005).
35. Bates, D. O. *et al.* The endogenous anti-angiogenic family of splice variants of VEGF, VEGFxxxb, are down-regulated in pre-eclamptic placentae at term. *Clinical Science* **110**, 575–585 (2006).
36. Santaguida, S., Vasile, E., White, E. & Amon, A. Aneuploidy-induced cellular stresses limit autophagic degradation. *Genes & Development* **29**, 2010–2021 (2015).
37. Peng, S. *et al.* Engineering a blood-retinal barrier with human embryonic stem cell-derived retinal pigment epithelium: transcriptome and functional analysis. *Stem cells translational medicine* **2**, 534 (2013).
38. Radeke, M. J. *et al.* Restoration of mesenchymal retinal pigmented epithelial cells by TGFβ pathway inhibitors: implications for age-related macular degeneration. *Genome medicine* **7**, 1–19 (2015).
39. Loayza-Puch, F. *et al.* Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. *Nature* **530**, 490–4 (2016).

40. Ngo, D. T. M. *et al.* Antiangiogenic actions of vascular endothelial growth factor-A165b, an inhibitory isoform of vascular endothelial growth factor-A, in human obesity. *Circulation* **130**, 1072–80 (2014).
41. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
42. Suzuki, Y. & Sugano, S. Construction of a full-length enriched and a 5′-end enriched cDNA library using the oligo-capping method. *Methods Mol. Biol.* **221**, 73–91 (2003).
43. Yamashita, R. *et al.* Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. *Genome Res.* **21**, 775–89 (2011).
44. Sorefan, K. *et al.* Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**, 4 (2012).
45. Qiu, Y., Hoareau-Aveilla, C., Oltean, S., Harper, S. J. & Bates, D. O. The anti-angiogenic isoforms of VEGF in health and disease. *Biochem. Soc. Trans.* **37**, 1207–13 (2009).
46. Qiu, Y. *et al.* Mammary alveolar development during lactation is inhibited by the endogenous antiangiogenic growth factor isoform, VEGF165b. *FASEB J* **22**, 1104–12 (2008).
47. Delcombel, R. *et al.* New prospects in the roles of the C-terminal domains of VEGF-A and their cooperation for ligand binding, cellular signaling and vessels formation. *Angiogenesis* **16**, 353–371 (2013).
48. Gammons, M., Dick, A. D., Harper, S. J. & Bates, D. O. SRPK1 inhibition modulates VEGF splicing to reduce pathological neovascularization in a rat model of retinopathy of prematurity. *Investigative ophthalmology & visual science* **54**, 5797–5806 (2013).
49. Hulse, R. *et al.* Regulation of alternative VEGF-A mRNA splicing is a therapeutic target for analgesia. *Neurobiology of disease* **71**, 245–259 (2014).
50. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* **41**, D991–5 (2013).
51. Suzuki, A. *et al.* DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic acids research* **43**, D87–D91 (2015).
52. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–10 (1990).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
54. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–6 (2011).
55. Gu, F. *et al.* VEGF111b, a new member of VEGFxxxb isoforms and induced by mitomycin C, inhibits angiogenesis. *Biochem. Biophys. Res. Commun.* **441**, 18–24 (2013).
56. Miller-Kasprzak, E. & Jagodzi'nski, P. 5-Aza-2′-deoxycytidine increases the expression of anti-angiogenic vascular endothelial growth factor 189b variant in human lung microvascular endothelial cells. *Biomedicine & Pharmacotherapy* **62**, 158–163 (2008).
57. Baltes-Breitwisch, M. M. *et al.* Neutralization of vascular endothelial growth factor antiangiogenic isoforms or administration of proangiogenic isoforms stimulates vascular development in the rat testis. *Reproduction* **140**, 319–29 (2010).

Acknowledgements

Author Contributions

D.A.S. conceived the study, wrote the main manuscript and prepared the figures with assistance from M.D. and S.B. S.B. performed BLAST searches and STAR alignments, M.D. performed PCR, library preparation and sequencing. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-00100-3

Competing Interests: The authors declare that they have no competing interests.

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