Title:

Transcriptomic analysis of the lesser spotted catshark (*Scyliorhinus canicula*) pancreas, liver and brain reveals molecular level conservation of vertebrate pancreas function

Authors:

John F Mulley[1*], Adam D Hargreaves[1], Matthew J Hegarty[2], R. Scott Heller[3] and Martin T Swain[2]

Author affiliations:

1. School of Biological Sciences, Bangor University, Brambell Building, Deiniol Road, Bangor, Gwynedd, LL57 2UW, United Kingdom

2. Institute of Biological, Environmental & Rural Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3DA, United Kingdom

3. Imaging Team, Novo Nordisk A/S, Novo Nordisk Park, DK2760 Måløv, Denmark

* corresponding author

Author email addresses:

John Mulley - j.mulley@bangor.ac.uk

Adam Hargreaves - bspa1b@bangor.ac.uk

Matthew Hegarty – ayh@aber.ac.uk

R. Scott Heller - shll@novonordisk.com

Martin Swain - mts11@aber.ac.uk

Abstract

Background

Understanding the evolution of the vertebrate pancreas is key to understanding its functions. The chondrichthyes (cartilaginous fish such as sharks and rays) have been suggested to possess the most ancient example of a distinct pancreas with both hormonal (endocrine) and digestive (exocrine) roles, although the lack of genetic, genomic and transcriptomic data for cartilaginous fish has hindered a more thorough understanding of the molecular-level functions of the chondrichthyan pancreas, particularly with respect to their "unusual" energy metabolism (where ketone bodies and amino acids are the main oxidative fuel source) and their paradoxical ability to both maintain stable blood glucose levels and tolerate extensive periods of hypoglycemia. In order to shed light on some of these processes we have carried out the first large-scale comparative transcriptomic survey of multiple cartilaginous fish tissues: the pancreas, brain and liver of the lesser spotted catshark, *Scyliorhinus canicula*.

Results

We generated a mutli-tissue assembly comprising 86,006 contigs, of which 44,794 were assigned to a particular tissue or combination of tissue based on mapping of sequencing reads. We have characterised transcripts encoding genes involved in insulin regulation, glucose sensing, transcriptional regulation, signaling and digestion, as well as many peptide hormone precursors and their receptors for the first time. Comparisons to published mammalian pancreas transcriptomes reveals that mechanisms of glucose sensing and insulin regulation used to establish and maintain a stable internal environment are conserved across jawed vertebrates and likely pre-date the vertebrate radiation. Conservation of pancreatic hormones and genes encoding digestive proteins support the single, early evolution of a distinct pancreatic gland with endocrine and exocrine functions in vertebrates, although the peptide diversity of the early vertebrate pancreas has been overestimated as a result of the use of cross-reacting antisera in earlier studies. A three hormone islet organ is therefore the basal vertebrate condition, later elaborated upon only in the tetrapod lineage.

Conclusions

The cartilaginous fish are a great untapped resource for the reconstruction of patterns and processes of vertebrate evolution and new approaches such as those described in this paper will greatly facilitate their incorporation into the rank of "model organism".

Key words

Catshark, pancreas transcriptome, Pdx1, Pdx2, insulin regulation

Background

Chondrichthyans (cartilaginous fish such as sharks, skates, rays (elasmobranchs) and chimeras (holocephalans)) possess the earliest example of a distinct pancreatic gland containing multiple cell types with both endocrine and exocrine functions in vertebrates [1, 2]. The more basal ("primitive") vertebrate lineages such as the jawless hagfish and lampreys (Figure 1) possess only small islet organs containing insulin- and somatostatin-producing endocrine cells and these islets lack any glucagon-producing cells or exocrine function [2-4]. The accumulation of multiple cell types into a single compact gland was an important step in pancreas evolution (and can be considered to be a vertebrate innovation) [5, 6] and it has been suggested that a switch from sensing gut-glucose to blood-glucose to establish a "stable inner *milieu*" via homeostatic mechanisms may have been an important factor in the evolution of a more complex glucose-dependent brain in vertebrates, protected from hyper- and hypoglycaemia [7-9]. However, the fact that insulin-like peptides in insects seem to fulfil similar roles in glucose metabolism and other physiological processes such as growth and reproduction suggests that at least elements of these mechanisms may have a more ancient origin [10].

Decades of research using light and electron microscopy and immunohistochemistry has revealed a great deal about the structure and organisation of the chondrichthyan pancreas and more recent studies have characterised the protein sequence and structure of some of the key pancreatic hormones [11-15]. The endocrine islets of chondrichthyans are typically scattered within exocrine tissue, sometimes associated

with minor ducts and contain a number of distinct cell types, thought to include the four major pancreas cell types: α-cells producing glucagon to increase blood glucose; β-cells producing insulin to reduce blood glucose; δ-cells producing somatostatin to regulate pancreatic hormones such as insulin and glucagon and γ-cells producing pancreatic polypeptide [16-18]. In addition to structural, cellular and hormonal conservation of the chondrichthyan pancreas compared to other vertebrates, there is also a conservation of function, with glucose-sensitive insulin release [11], pancreatectomy-induced hyperglycemia [19, 20] and exogenous insulin-induced hypoglycaemic effects [11, 21-23]. However, although blood plasma glucose levels are maintained at a fairly constant level during feeding and fasting (even over periods of up to 150 days without food) [21, 24, 25], actual plasma glucose levels in elasmobranchs are lower than in teleost fish of comparative size and with similar metabolic rates [26]. It has also been found that elasmobranchs have an impressive tolerance of hypoglycaemia, including an ability to cope with a virtual absence of circulating glucose for at least 24 hours, a 75% reduction for at least a week and sub-normal plasma glucose levels for extended periods [22, 27]. There is an obvious paradox associated with an impressive ability to cope with long periods of hypoglycaemia existing in conjunction with the maintenance of apparently stable plasma glucose levels and others have pondered the necessity of central glucose-sensing mechanisms in these species [26].

The chondrichthyan pancreas represents an important model for studies of vertebrate pancreas evolution and function, particularly with reference to glucose homeostasis. However, full analysis of these areas has been hindered by a lack of genetic information and resources - almost the entirety of our current understanding of the chondrichthyan pancreas is based on what might be considered somewhat “old fashioned” (although still vital, important and informative) biological techniques, including descriptive gross anatomy and light and electron microscopy, enzymatic assays typically involving the injection of peptides derived from other (often mammalian) species and immunohistochemistry involving the use of antibodies raised against short mammalian peptide epitopes (see [2, 3, 26, 28] for reviews). There is currently a dearth of data regarding molecular level functions of the chondrichthyan pancreas, including mechanisms of transcriptional and translational control of gene regulation, signaling both in terms of cell-

cell communication within the pancreas and in terms of response to neuroendocrine signals and even more basic information such as the sequences of mRNA and protein precursors of previously identified or characterised digestive enzymes and pancreatic peptides.

The lesser spotted catshark (*Scyliorhinus canicula*, often referred to as the lesser spotted dogfish) has recently become the chondrichthyan model of choice for a wide range of genetic, developmental and evolutionary analyses [29] and transcriptomic and genomic sequencing projects are currently underway for this species at Genoscope (www.genoscope.cns.fr). An enigmatic second member of the *Pancreas and duodenal homeobox* (Pdx) gene family (called *Pdx2*) was recently identified in the *S. canicula* pancreas [30] but further studies of the role of this gene or the identification of the presence of additional members of other gene families involved in pancreas development, cell-specification and insulin regulation are impossible without more comprehensive molecular analyses. In order to shed further light on the possible role of the *Pdx2* gene and to go some way to addressing the current dearth of data we set out to determine the pancreas transcriptome of the lesser spotted catshark and to carry out comparative expression analyses with other adult body tissues (liver and brain). These data represent the first large-scale transcriptomic analysis of multiple cartilaginous fish tissues and will be invaluable in understanding the functions of the cartilaginous fish pancreas, as well as shedding light on the evolution of the vertebrate pancreas itself.

Results

A total of 6,260,398; 32,106,318 and 12,201,682 paired-end sequencing reads were generated for the pancreas, liver and brain respectively and these were pooled to generate a single assembly (Additional file 1) which contained 86,006 contigs (when trimmed to remove all contigs <300bp, which likely represent single pairs of sequencing reads). The tissue distribution of these transcripts was determined by mapping sequencing reads from each tissue to this assembly and abundance values of ≥1 fragments per kilobase per million mapped reads (FPKM) were taken to confirm expression of a particular transcript in each tissue (Table 1). In this way 44,794 contigs were assigned to one or more tissues (Figure 2, Additional files 2-8). All transcripts contained an ORF encoding 20 amino acids or more (Figure 3), of which roughly 4-7%

encoded a signal peptide and so are likely to be secreted (Table 1). Typically, between 15-53% of transcripts had a BLAST hit in the RefSeq collection and 12-49% were annotated with GO terms (Table 2). Although low, these figures are broadly comparable with a similar analysis of the white shark heart transcriptome [31], which found matches of 23.5% and 21.5% respectively. In order to provide a broad overview of the assigned gene ontology terms, we carried out a generic GOSlim annotation of the data (Figures 4 and 5), and Fishers exact tests showed that the pancreas was enriched for seven GO terms compared to liver ('cell', 'reproduction', 'transcription, DNA-dependent', 'embryo development', 'growth', 'intracellular non-membrane-bounded organelle' and 'non-membrane-bounded organelle') and two terms compared to brain ('transcription, DNA-dependent' and 'cellular amino acid metabolic process' – the full comparative enrichment results are provided in Additional file 9).

A more detailed search strategy was carried out for particular categories of genes that would shed light on the similarities or differences of pancreas function in *S. canicula* compared to other vertebrates. The results of these analyses are outlined in the following sections.

Pancreatic hormones and their receptors

A large amount of immunohistochemical research has putatively identified the presence of a number of pancreatic peptide hormones in cartilaginous fish, including insulin, glucagon, somatostatin and pancreatic polypeptide, and the presence of at least some of these peptides has been confirmed by proteomic studies [11, 12, 15, 32]. Our transcriptomic data confirms the presence of mRNA transcripts encoding preproinsulin, preproglucagon (Figure 6) and preprosomatostatin (corresponding to the *SSa* gene [33]) in the pancreas and preprosomatostatin b and c in both the brain and pancreas, but not pancreatic polypeptide (PP – see next section), gastrin, gastric inhibitory polypeptide (GIP) or secretin. Vasoactive intestinal polypeptide (VIP) is expressed by both pancreas and brain, cholecystokinin by only brain and, as previously suggested [3], ghrelin appears to be absent from the shark pancreas and, indeed, from all three tissues. We find transcripts encoding the insulin receptor (IR) only in brain, the glucagon receptor only in liver and somatostatin receptor 1 (SSTR1) and SSTR5 in brain and both pancreas and brain respectively. Contrary to the findings of Larsson *et al.* [34], we find Neuropeptide Y receptors Y1, Y5 and Y6 in brain

only and Y8 in both pancreas and brain, suggesting that the expression of these receptors either varies between chondrichthyan species, or is more dynamic than previously assumed.

The presence of PP and γ-cells are key aspects of current schemes for the mode of vertebrate pancreas evolution [1-3, 35]. However, it has been known for some time that PP is tetrapod-specific, produced via duplication of the *Peptide YY* gene sometime prior to the divergence of this lineage [36, 37] and there is therefore a discrepancy between the findings of decades of immunohistochemical research and data from molecular genetic studies and analyses of vertebrate whole genome sequences. We have identified transcripts of two members of the Neuropeptide Y family (which includes Neuropeptide Y (NPY), Peptide YY (PYY) and Pancreatic polypeptide (PP)) in our dataset - a *PYY* gene expressed in pancreas and brain and a *NPY* gene expressed only in brain (Figure 7, both sequences are identical to published catshark sequences for PYY and NPY (accessions P69095 [38], AAB23237 [14]). We therefore suggest that older immunohistochemical studies which claimed to have detected PP+ cells in the cartilaginous fish pancreas may have in fact been relying on antisera that cross-reacted with PYY. A focus on the (often brief) methods sections of several key historical papers revealed that they in fact used the same anti-PP antibody, produced by Ronald Chance at Eli Lily in the 1970's [16, 39, 40]. It therefore appears that this antibody was detecting PYY in the pancreas of cartilaginous fish and that these initial papers and various subsequent papers have repeatedly been cited until the presence of PP in cartilaginous fish is considered to be established fact. In other cases, the misidentification of sequenced peptides has added to the confusion [38].

Our immunohistochemical surveys of the catshark pancreas using high-affinity anti-PP antibodies (Table 1 in Additional file 10) showed varying results. The PP antisera from Sigma weakly stained the catshark pancreas but the staining was completely absorbed with PP, NPY and PYY peptides (Figure 8). While the Millipore anti-PP failed to stain (except on the mouse control pancreas, data not shown), experiments with anti-PYY antibodies detected strong signals, co-localising with insulin but not glucagon or somatostatin (Figure 8). NPY antisera immunoreactivity was detected in the same pattern as PYY (data not shown) and was absorbed with either PYY or NPY peptides (Table 2 in Additional file 10).

A search of the Blast2GO results for the term 'Hormone activity' identifies three other peptide-encoding transcripts expressed in the catshark pancreas: Gastrin-releasing peptide (GRP, in pancreas and brain), which fulfils a variety of roles in the gastrointestinal tract, including the regulation of hormone release and the secretion of pancreatic enzymes [41, 42]; Neuromedin U (NMU, in all three tissues) which is expressed in nerves throughout the gastrointestinal tract [43] and Enkephalin (in pancreas and brain), an endogenous opioid that functions as a neurotransmitter or neuromodulator [44, 45].

Glucose sensing

Hexokinase type IV, more typically called *Glucokinase* (*GK*), is a glucose-phosphorylating enzyme that has been shown to be the key molecule for glucose sensing in mammalian liver and pancreas cells [46] and mutations in *GK* are known to cause Maturity Onset Diabetes of the Young Type II (MODY2) [47, 48]. Somewhat surprisingly, we find that *GK* is expressed in the shark brain and *glucokinase regulatory protein* (*GKRP*) only in liver. We also identified transcripts in all three tissues corresponding to *6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2* (*Pfkfb2*), another regulator of glucose metabolism via interaction with GK. Finally we detected transcripts encoding two glucose transporters in all three tissues, namely the solute carrier family 2 genes *Glut1* and *Glut2* and the shark pancreas therefore appears to represent an ancestral state where both types of transporter are expressed in the pancreas, as opposed to the the human pancreas (which relies on *GLUT1*) and that of rodents (which rely more heavily on *Glut2*) [49-51].

Insulin regulation

The regulation of the insulin gene has, for obvious reasons, been an area of intensive study and a number of key regulators are now known from studies in rodents and humans. Perhaps one of the most important is the *Pancreas and duodenal homeobox 1* (*Pdx1*) gene, also called *Insulin promoter factor 1* (Ipf1), *Islet/duodenum homeobox 1* (*Idx1*) or *Somatostatin transactivating factor 1* (*Stf1*) [52-54] which (in addition to roles in embryonic development and β-cell specification) binds to the TAAT motif-containing A boxes of the mammalian insulin promoter to stimulate transcription [55-58]. Mutations in *PDX1* have been linked to Maturity Onset Diabetes of the Young Type IV (MODY4) [59] and pancreatic agenesis

[60, 61]. Cartilaginous fish and coelacanths have previously been shown to have retained an ancient paralogue of *Pdx1*, which we termed *Pdx2* [30, 62] and accordingly we find transcripts of both genes in our pancreas transcriptome dataset. The basic helix-loop-helix (bHLH) transcription factor *NeuroD1* (previously called β2 [63]) has roles in pancreas development and islet formation [64, 65] and mutations in this gene have been linked to Type 2 Diabetes mellitus and MODY6 [66]. *NeuroD1* has also been shown to be important for insulin gene expression [67-69] and the LIM-homeodomain transcription factor *Isl1* acts synergistically with *NeuroD1* and the bHLH transcription factor *E47* to bind to and activate the insulin gene promoter [70-72]. The *hepatocyte nuclear factor 1 alpha* (*Hnf1a*) gene was originally described as a liver-specific transcription actor, responsible for the regulation of a number of genes important for liver function [73-75]. However, it was later found that this gene also has a role in glucose homeostasis via regulation of insulin secretion [76, 77] and that mutations in *Hnf1a* were the cause of MODY3 [78-80]. The *Nkx6.1* homeodomain transcription factor is a potent transcriptional repressor with a key role in β-cell differentiation [81, 82] and has also been shown to suppress the expression of glucagon to maintain β-cell identity, as well as being able to regulate glucose-sensitive insulin secretion [83]. Our discovery of *Pdx1* (and *Pdx2*), *NeuroD1* and its partner *E47*, *Isl1*, *Hnf1a* and *Nkx6.1* transcripts expressed in the catshark pancreas suggests an ancient role for these genes in vertebrate pancreas function and hints at early establishment of the insulin gene regulatory network. Additionally, the presence of transcripts encoding *NeuroD1*, *e47*, *Isl1* and *Nkx6.1* in the catshark brain highlights shared ancestry of these tissues in the vertebrate neuroendocrine system. We do not find any transcripts for *MafA*, which has been shown to be a key regulator of glucose-sensitive insulin secretion in humans and rodents [58, 84, 85], although other studies have also had difficulty identifying transcripts of this gene and other pancreas transcription factors in non-PCR based experiments [86, 87], possibly because of the low level of expression of transcription factors in general [88].

Transcription factors

In addition to the transcription factors involved in insulin regulation discussed above, KEGG orthology (KO) analysis [89-91] has identified 13 transcription factors expressed in just the pancreas (including

Pdx1 and *Pdx2*, *FoxA1* (*Hnf3a*) and *Pancreas-specific transcription factor 1a* (*Ptf1a*)), 14 expressed in pancreas and liver, 33 in pancreas and brain and 51 in all three tissues (Tables 2 and 3).

In a survey of the expression of 790 human DNA-binding transcription factors, Kong et al. [88] identified 80 with expression restricted to the fetal pancreas, 32 restricted to the adult pancreas and 18 shared by both. Of the 31 adult-specific genes, we find evidence that 6 are also expressed in the adult catshark pancreas, although this number increases to 15 if members of the same gene family are considered (the possibility of divergent resolution of gene duplicates following the whole genome duplications [92] in early vertebrate ancestry must be considered). Since transcription factors are known to be expressed at low levels in cells (less than 20 copies per human adult cell [88]) it is likely that our figure is an underestimate and a more comprehensive survey of candidate transcription factor expression in this species is needed.

Signalling

Our KEGG orthology analysis identified 38 transcripts involved in signal transduction that are expressed only in the catshark pancreas, 11 in both pancreas and liver, 104 in pancreas and brain and 187 in all three tissues (Tables 2 and 3). Among these are representatives of the major vertebrate signalling pathways, including ligands and receptors for Fgf, Wnt, Notch, Vegf, Tgfβ and Pdgf. Members of all of these pathways have previously been identified in the human pancreas transcriptome [87].

Homeobox gene diversity

Homeobox genes are a group of transcription factors that encode a 60 amino acid DNA-binding homeodomain and that are involved in a wide variety of gene regulatory events in embryonic and adult tissues. A number of homeobox genes are known to be expressed during endodermal regionalisation and pancreas development, including *Islet 1* and *2* (*Isl1*, *Isl2*), *Pancreatic and duodenal homeobox 1* (*Pdx1*), *Nkx6.1*, *Nkx2.2*, *Pituitary homeobox 2* (*Pitx2*), *Motor neuron and pancreas homeobox 1* (*Mnx1*), *Onecut homeobox 1* (*Onecut/Hnf6*) and *Paired box genes 4* and *6* (*Pax4*, *Pax6*) [93, 94]. Some older studies have detected a variety of homeobox genes in mammalian pancreas cell lines, including *Cdx4*, *Hox1.4* (*HoxA4*), *Chox7* (*Gbx1*), *Hox2.6* (*HoxB4*), *Cdx3* (*Cdx2*), *Cdx1*, *Hox4.3* (*HoxD8*), *Hox1.11* (*HoxA2*), *Hox4a* (*HoxD3*), *Hox1.3* (*HoxA5*) in the somatostatin-producing rat insulinoma cell line RIN1027-B2 [53] and

Isl1, *Lmx2*, *Alx3*, *HoxA4*, *HoxA13*, *Ipf1* (*Pdx1*), *Nkx2.2*, *Nkx6.1*, *En2* and *Vdx* in a hamster insulinoma cell line [95]. More recently, microarray and RNA-seq studies have identified a much larger number of homeobox genes expressed in the pancreas and especially the β-cell, with over 60 different homeobox genes identified by Kutlu et al. [87]. We used the homeodomain sequences of all human homeobox genes from HomeoDB [96, 97] and all vertebrate homeobox gene sequences from Pfam [98] as BLAST queries against our catshark transcriptome data and identified 11 different homeobox genes expressed in the pancreas, including five in just pancreas (*HoxB5*, *HoxB7*, *Mox1*, *Pdx1*, *Pdx2*), two in pancreas and liver (*Hlx*, *Hhex*), three in pancreas and brain (*Arx*, *Zfhx3*, *Zfhx4*) and one in all three tissues (*Cut-like 2*). These include genes known to be restricted to, or highly expressed in, β-cells (*Pdx1*), α-cells (*Arx*) and acinar cell types (*Cut-like 2*) [86].

Digestion

In addition to its endocrine roles, the pancreas is also an important exocrine organ, fulfilling key functions in the digestion of proteins, lipids and carbohydrates. In the carnivorous elasmobranchs protein and lipids are the main energy sources [99] and it has been shown that ketone bodies and amino acids are the main oxidative fuel source for muscles and several other tissues, in preference to fatty acids [24, 28, 99]. Carbohydrates are thought to be utilised as oxidative fuels in elasmobranch heart muscle, as well as brain, red muscle and rectal gland [28, 100, 101]. It is therefore perhaps reasonable to assume that proteases and lipases are the most significant digestive enzymes produced by the elasmobranch pancreas and indeed this appears to be the case. Some form of chymotrypsinogen and trypsinogen have long been known to be produced by the elasmobranch pancreas, as has carboxypeptidase B, although these enzymes have not been fully characterised or isolated and sequenced [102-105]. We find transcripts of *Elastase 2a* and *3b*, *Chymotrypsinogen b1* (*Ctrb1*), *Chymotrypsin-like* (*Ctrl*), *Chymotrypsin-like elastase family, member 1* (*Cela1*) and *Chymotrypsin-like elastase family, member 3B* (*Cela3b*), *Trypsin 1*, *2* and *3*, as well as the digestive carboxypepetidases (A1, A2, B1) and those involved in activation and processing of other proteins, such as *carboxypeptidase B2*, *D* and *E* [106, 107].

Some form of triacylglycerol lipase activity has previously been detected using crude enzyme preparations from the pancreas of skate (*Raja* (now *Amblyraja*) *radiata*) [108] and Leopard shark, *Triakis semifasciata* [109]. However, we find no evidence of Pancreatic liapse in the catshark pancreas transcriptome and instead find only Pancreatic lipase-related proteins 1 (Pnliprp1) and 2 (Pnliprp2). It is likely therefore that the triacylglycerol lipase activity found previously is a result of the action of Carboxyl ester lipase (CEL or bile salt-stimulated lipase) [110]. We also find colipase, in agreement with earlier studies of a range of cartilaginous fish and other basal vertebrates [111-113] and Hepatic and Hormone-sensitive lipases. Several lipid transporting apolipoproteins are also expressed by the catshark pancreas, including apolipoproteins A-IV, E, M and O. Finally, we have identified transcripts of genes involved in the digestion of carbohydrates (Pancreatic alpha amylase) and nucleic acids (deoxyribonuclease I and various ribonucleases).

Microsatellites

It has recently been suggested [31] that a high frequency of dinucleotide simple sequence repeats (SSRs, microsatellites) is a general feature of shark genomes. We find 6,843 transcripts containing one or more di-, tri- or tetranucleotide microsatellites of five perfect repeats or more in our catshark data, with 482 of these only in pancreas, 3,083 only in brain and 473 only in liver (Table 4). In accordance with previous suggestions [31] we find dinucleotide repeats to be the most common type of SSR in both coding and non-coding regions of catshark transcripts.

Discussion

Our analysis of the catshark pancreas transcriptome reveals the presence of genes known to be involved in glucose sensing and regulation of the insulin gene in other vertebrates and illustrates that functional conservation of these aspects of the vertebrate pancreas is reflected at the molecular-level. We therefore propose that these molecular-level mechanisms are a common feature of jawed vertebrates and that this lends support to the theory that the evolution of blood-glucose sensing and regulatory mechanisms may

311 have facilitated the evolution of the complex glucose-dependent brain of vertebrates [7-9]. We further
312 suggest that the early evolution and fixation of these mechanisms has imposed evolutionary constraints on
313 glucose sensing and insulin regulation in vertebrates, including in cartilaginous fish, even in the face of
314 their ability to tolerate extended periods of hypoglycaemia and likely relaxed requirements for these
315 processes.
316 We find that the catshark pancreas produces at least eight peptide hormones (insulin; glucagon;
317 somatostatin; peptide YY; gastrin-releasing peptide, neuromedin U, encephalin and vasoactive intestinal
318 polypeptide, Table 5) and expresses a wide variety of genes involved in digestion, especially the digestion
319 of proteins and lipids. The catshark pancreas therefore clearly has the features of a distinct pancreatic
320 gland with both endocrine and exocrine functions and as such will be of great use in reconstructing the
321 characteristics of the earliest vertebrate pancreas. The similarity in gene expression between the catshark
322 and other vertebrates with respect to hormones, digestive enzymes, transcription factors and signaling
323 pathways again provides support to the theory that there was a single, early origin of the pancreas at the
324 base of the jawed vertebrate radiation. The overlap in peptides produced by the catshark pancreas and
325 brain (Table 5) is a reflection of the shared ancestry of these tissues within the vertebrate neuroendocrine
326 system [114].
327 Based on its co-localisation with insulin-, glucagon-, somatostatin- and PP-cells during mouse
328 development, it has previously been suggested that a PYY+ cell may constitute a common progenitor of
329 the major islet cell types [115]. Recent lineage tracing experiments have demonstrated that PYY+ cells
330 give rise to islet ∂ and PP cells and approximately 40% of pancreatic α and at least some β cells arose
331 from peptide YY+ cells [116]. Most β cells and the majority of α cells are therefore not descendants of the
332 peptide YY+/glucagon+/insulin+ cells that first appear during early pancreas ontogeny. The co-
333 localisation of PYY with insulin in the adult shark pancreas illustrates the diversity of mechanisms that
334 exist in vertebrate pancreas development and function and demonstrates the utility of "non-model" species
335 to study these processes. The catshark PYY+ cells will therefore provide important insights into the
336 evolution of the vertebrate pancreas, and especially progenitors of α, β, δ and γ-cells.

Our experiments make clear that much of the previous work on the presence or absence of peptides in basal vertebrate lineages may be suspect, with many false-positive signals resulting from cross-reacting antisera. Previous schemes of pancreas evolution based on these and similar data, which posited the restriction of various hormones to the alimentary canal (similar to the situation in protochordates such as amphioxus), the accumulation of these into a two-or three peptide islet organ in jawless fish and finally the "classic four-hormone islet tissue" of cartilaginous fish and other vertebrates [2] are therefore incorrect. In fact, it appears that the three hormone (glucagon, insulin, somatostatin) islet organ was established early in vertebrate evolution and remains today in the adult (but not larval) lamprey, cartilaginous fish and actinopterygian (ray-finned) fish, and that it is only in the sarcopterygian (lobe-finned fish) lineage that a four hormone (the above, plus PP) pancreas was formed.

Our analysis of homeobox gene expression reveals a surprising level of variation between the genes known to be expressed in the catshark pancreas, human islets [87] and rat [53] and hamster [95] cell lines. It therefore seems likely that this particular class of transcription factors is extremely variable with respect to their spatial or temporal expression pattern in the vertebrate pancreas (or more likely both) and this is perhaps not too surprising given the variety of roles carried out by the pancreas in response to feeding, digestion and the regulation of blood glucose. As expected we have identified transcripts of both *Pdx1* and *Pdx2* in the catshark pancreas, although we do not find any evidence for the presence of additional duplicates of other genes encoding proteins known to interact with PDX1 in other species. It therefore seems unlikely that the maintenance of paralogous *Pdx2* genes in some vertebrate lineages reflects a wider conservation of duplicated gene regulatory networks produced as a result of whole genome duplication events early in vertebrate evolution. Comparison of the amino acid sequences of PDX1 and PDX2 across vertebrates shows conservation of the Pbx-interacting motif, DNA-binding domain and nuclear localisation signal but not of known transactivation domains and the PCIF1-interaction domain [117-119] (Figure 9). The functions of the *Pdx2* gene and the reasons for its retention in some species and independent loss in others (ray-finned fish and tetrapods) remain unknown.

With the availability of whole genome sequence information for a greater number of taxa and improved coverage of vertebrate pancreas transcriptomes a larger amount of data than ever before is now becoming available. These data, together with an appreciation that early vertebrate evolution was characterised by extensive genetic, developmental and morphological innovation facilitated by multiple whole genome duplications [120, 121] will better enable us to reconstruct pancreas evolution. As an example, we propose that the creation of the paralogous NPY and PYY during these duplications [34] facilitated the separation of the neuronal and gastroenteropancreatic (GEP) endocrine systems. We further suggest that the availability of additional copies of developmentally-important genes produced during the same duplication events [121] enabled the remodelling of the developing gut and the formation of a distinct pancreas with both endocrine and exocrine functions.

Conclusions

We have generated a multi-tissue transcriptomic resource for an up and coming model organism, the lesser spotted catshark, *Scyliorhinus canicula*. Somewhat surprisingly we find few transcripts in common between the liver and pancreas, despite their relatively similar roles and shared developmental history as endodermal neighbors. The higher number of transcripts in common between brain and pancreas may provide evidence in support of the co-opting of neuronal programs by at least some pancreatic cells during vertebrate evolution [114, 122], although further comparative analyses are needed in this area. The similarity between the catshark pancreas transcriptome and those of various mammals with respect to insulin regulation, transcriptional and signaling machinery and peptide hormones and their receptors supports the single, early origin of a distinct pancreatic gland in vertebrates, although it seems likely that the peptide diversity of the early vertebrate pancreas may have been overestimated by older, immunohistochemical studies. The cartilaginous fish have a three peptide (insulin, glucagon and somatostatin) pancreas and the four peptide system seen in actinopterygian (ray-finned) and sarcopterygian (lobe-finned) fish and tetrapods is a later evolutionary innovation. The retention of the Pdx2 gene in cartilaginous fish does not apparently reflect a wider retention of duplicated members of

388 pancreas gene regulatory networks and the possible function(s) of this gene remains enigmatic. Our data,
389 together with available or in progress transcriptomic and genomic resources for this and other
390 chondrichthyan species will greatly facilitate comparative studies of elasmobranch, chondrichthyan and
391 vertebrate evolution, particularly with reference to energy metabolism and the maintenance of stable blood
392 glucose levels.
393
394 **Methods**
395 *RNA-Seq and sequence analysis*
396 Experimental methods involving animals followed institutional and national guidelines and were approved
397 by the Bangor University Ethical Review Committee. Total RNA was extracted from freshly-dissected
398 pancreas, liver and brain of two adult male and female catsharks approximately 24 hours post feeding.
399 Pancreas samples were sequenced with using 2 x 250bp paired-end reads on the Illumina MiSeq platform
400 at the Centre for Genomic Research (CGR) at the University of Liverpool. Brain and liver samples were
401 sequenced using 2x 150bp paired-end reads on the Illumina HiSeq 2000 platform at the Institute of
402 Biological, Environmental & Rural Sciences (IBERS) at Aberystwyth University. Sequencing reads from
403 the three tissues were assembled into a global tissue assembly using Trinity [123] with the jellyfish K-mer
404 counting method. Tissue distribution of transcripts was assessed by mapping sequencing reads from each
405 tissue to this global assembly, with an FPKM (fragments per kilobase per million mapped reads) value of
406 >1 taken as confirmation of expression. Transcript annotation and assignment of gene ontology (GO)
407 terms was performed using BLAST2GO [124, 125], the KEGG Automatic Annotation Server (KAAS
408 [90]) and by local BLAST using BLAST+ v2.2.27 [126].
409
410 *Immunohistochemistry*
411 Male catsharks were euthanized according to a Schedule 1 method and the pancreas removed and fixed in
412 4% paraformaldehyde/PBS overnight at 4°C. The fixed pancreas was then rinsed several times in PBS,
413 dehydrated through a graded ethanol series and stored in 100% ethanol. 5µm sections of paraffin

414 embedded catshark pancreas were cut and mounted on glass slides. Slides were microwave treated in Tris-
415 EGTA (TEG) buffer ph9.0 and allowed to cool for 30mins. Slides were rinsed in PBS and blocked in
416 normal donkey serum and TNB buffer (Perkin Elmer) for 30mins. Primary antisera were added over night
417 at room temperature and the next day the slides were rinsed 3 x 5min each in PBS and specific cross
418 absorbed donkey anti- mouse, rabbit, or guinea pig secondary antisera (Jackson Immunoresearch) were
419 added for 30mins. The slides were rinsed in PBS and mounted. Details of antisera are given in Table 1 in
420 Additional file 10. All pictures were taken on a Zeiss Meta510 confocal microscope.
421
422 *Antibody Absorption*
423 In order to test their specificity against the Pancreatic polypeptide family, the antisera were incubated
424 overnight at 4°C with 10µg of either pancreatic polypeptide (Sigma), Neuropeptide Y (Bachem) or
425 peptide YY (in-house synthesis) or no peptide. The next day the antisera were added to the slides and the
426 staining was performed as above. The staining intensity was compared to the no peptide control and given
427 a rating of 1-3 (+, ++, +++). The results are shown in Table 2 in Additional file 10.
428
429 **Competing interests**
430 The authors declare no competing interests.
431
432 **Authors' contributions**
433 JFM devised the study and drafted the manuscript; JFM, ADH MJH and MTS carried out RNA-Seq
434 experiments and data analysis, RSH carried out immunohistochemical experiments. All authors read and
435 approved the final manuscript.
436
437 **Acknowledgements**
438 This research was supported by a Diabetes UK Small Grant to JFM and donations by the Llandudno and
439 District Diabetes Group. We thank Ashley Tweedale and Gavan Cooke for assistance with collection and

maintenance of catsharks and Charity MK McGrae for assistance with immunohistochemistry and imaging. JFM, MJH and MTS are supported by the Biosciences, Environment and Agriculture Alliance (BEAA) between Bangor University and Aberystwyth University and ADH is funded by a Bangor University 125th Anniversary Studentship.

References

1. Falkmer S: **Origin of the parenchymal cells of the endocrine pancreas: Some phylogenetic and ontogenetic aspects.** Frontiers of Gastrointestinal Research 1995, **23**:2-29.

2. Youson JH, Al-Mahrouki AA: **Ontogenetic and phylogenetic development of the endocrine pancreas (islet organ) in fishes.** General and Comparative Endocrinology 1999, **116**(3):303-335.

3. Heller RS: **The comparative anatomy of islets.** The Islets of Langerhans 2010, :21-37.

4. Epple A, Brinn JE: **Pancreatic islets.** Vertebrate Endocrinology: Fundamentals and Biomedical Implications: Morphological considerations 1986, **1**:279-317.

5. Shimeld SM, Holland PWH: **Vertebrate innovations.** Proceedings of the National Academy of Sciences of the United States of America 2000, **97**(9):4449.

6. Khaner O: **Evolutionary innovations of the vertebrates.** Integrative Zoology 2007, **2**(2):60-67.

7. Ainali C, Simon M, Freilich S, Espinosa O, Hazelwood L, Tsoka S, Ouzounis CA, Hancock JM: **Protein coalitions in a core mammalian biochemical network linked by rapidly evolving proteins.** BMC Evolutionary Biology 2011, **11**(1):142.

8. Rashidi A, Kirkwood TBL, Shanley DP: **Metabolic evolution suggests an explanation for the weakness of antioxidant defences in beta-cells.** Mechanisms of Ageing and Development 2009, **130**(4):216-221.

9. Madsen OD: **Pancreas phylogeny and ontogeny in relation to a "pancreatic stem cell'.** Comptes Rendus Biologies 2007, **330**(6-7):534-537.

10. Wu Q, Brown MR: **Signaling and function of insulin-like peptides in insects.** Annual Review of Entomology 2006, **51**:1-24.

11. Anderson WG, Ali MF, Einarsdottir IE, Schaffer L, Hazon N, Conlon JM: **Purification, characterization, and biological activity of insulins from the spotted dogfish, *Scyliorhinus canicula*, and the hammerhead shark, *Sphyrna lewini*.** General and Comparative Endocrinology 2002, **126**(1):113-122.

12. Conlon JM, Hazon N, Thim L: **Primary structures of peptides derived from proglucagon isolated from the pancreas of the elasmobranch fish, Scyliorhinus canicula.** Peptides 1994, **15**(1):163-167.

13. Conlon JM, Balasubramaniam A, Hazon N: **Structural characterization and biological activity of a neuropeptide Y-related peptide from the dogfish, *Scyliorhinus canicula*.** Endocrinology 1991, **128**(5):2273-2279.

14. Conlon JM, Bjenning C, Hazon N: **Structural characterization of neuropeptide Y from the brain of the dogfish, *Scyliorhinus canicula*.** Peptides 1992, **13**(3):493-497.

15. Conlon JM, Dafgard E, Falkmer S, Thim L: **A glucagon-like peptide, structurally related to mammalian oxyntomodulin, from the pancreas of a holocephalan fish, *Hydrolagus colliei*.** Biochemical Journal 1987, **245**(3):851.

16. El-Salhy M: **Immunocytochemical investigation of the gastro-enteropancreatic (GEP) neurohormonal peptides in the pancreas and gastrointestinal tract of the dogfish *Squalus acanthias*.** Histochemistry and Cell Biology 1984, **80**(2):193-205.

17. Kobayashi K, Syed Ali S: **Cell types of the endocrine pancreas in the shark *Scyliorhinus stellaris* as revealed by correlative light and electron microscopy.** Cell and Tissue Research 1981, **215**(3):475-490.

18. Sekine Y, Yui R: **Immunohistochemical study of the pancreatic endocrine cells of the ray, *Dasyatis akajei*.** Archivum Histologicum Japonicum 1981, **44**(1):95.

19. Diamare V: **Vergleichende anatomisch-physiologische Studien fiber den Pankreasdiabetes. III Mitt.** Zentralblatt für Physiologie 1908, **21**:863-869.

20. Diamare V: **Weitere Beobachtungen über den Experimentaldiabetes nach Pankreasextirpation bei Selachier.** Zentralblatt für Physiologie 1906, **20**:617–620.

21. deRoos R, deRoos CC, Werner CS, Werner H: **Plasma levels of glucose, alanine, lactate, and [beta]-hydroxybutyrate in the unfed spiny dogfish shark (*Squalus acanthias*) after surgery and following mammalian insulin infusion.** General and Comparative Endocrinology 1985, **58**(1):28-43.

22. deRoos R, deRoos CC: **Severe insulin-induced hypoglycemia in the spiny dogfish shark (*Squalus acanthias*).** General and Comparative Endocrinology 1979, **37**(2):186-191.

23. Patent GJ: **The Chondrichthyean Endocrine Pancreas: What are its Functions?** American Zoologist 1973, **13**(3):639-651.

24. Zammit VA, Newsholme EA: **Activities of enzymes of fat and ketone-body metabolism and effects of starvation on blood concentrations of glucose and fat fuels in teleost and elasmobranch fish.** Biochemical Journal 1979, **184**(2):313.

25. Walsh PJ, Kajimura M, Mommsen TP, Wood CM: **Metabolic organization and effects of feeding on enzyme activities of the dogfish shark (*Squalus acanthias*) rectal gland.** Journal of Experimental Biology 2006, **209**(15):2929-2938.

26. Polakof S, Mommsen TP, Soengas JL: **Glucosensing and glucose homeostasis: From fish to mammals.** Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 2011, **160**(4): 123–149.

27. Patent GJ: **Comparison of some hormonal effects on carbohydrate metabolism in an elasmobranch (*Squalus acanthias*) and a holocephalan (*Hydrolagus colliei*).** General and Comparative Endocrinology 1970, **14**(2):215-242.

28. Speers-Roesch B, Treberg JR: **The unusual energy metabolism of elasmobranch fishes.** Comparative Biochemistry and Physiology-Part A: Molecular & Integrative Physiology 2010, **155**(4):417-434.

29. Coolen M, Menuet A, Chassoux D, Compagnucci C, Henry S, Lévèque L, Da Silva C, Gavory F, Samain S, Wincker P: **The dogfish *Scyliorhinus canicula*, a reference in jawed vertebrates.** Emerging model organisms. A laboratory manual 2009, **1**:431–446.

30. Mulley JF, Holland PWH: **Parallel retention of *Pdx2* genes in cartilaginous fish and coelacanths.** Molecular Biology and Evolution 2010, **27**(10):2386-2391.

31. Richards V, Suzuki H, Stanhope M, Shivji M: **Characterization of the heart transcriptome of the white shark (*Carcharodon carcharias*).** BMC Genomics 2013, **14**(1):697.

32. Conlon JM, Reinecke M, Thorndyke MC, Falkmer S: **Insulin and other islet hormones (somatostatin, glucagon and PP) in the neuroendocrine system of some lower vertebrates and that of invertebrates: a minireview.** Hormone and Metabolic Research 1988, **20**:406-410.

33. Quan FB, Kenigfest NB, Mazan S, Tostivint H: **Molecular cloning of the cDNAs encoding three somatostatin variants in the dogfish (*Scylorhinus canicula*).** General and Comparative Endocrinology 2013, **180**:1-6.

34. Larsson TA, Tay BH, Sundstrom G, Fredriksson R, Brenner S, Larhammar D, Venkatesh B: **Neuropeptide Y-family peptides and receptors in the elephant shark, *Callorhinchus milii* confirm gene duplications before the gnathostome radiation.** Genomics 2009, **93**(3):254-260.

35. Falkmer S: **Phylogeny and ontogeny of the neuroendocrine cells of the gastrointestinal tract.** Endocrinology and Metabolism Clinics of North America 1993, **22**(4):731.

36. Larhammar D: **Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide.** Regulatory Peptides 1996, **62**(1):1-11.

37. Sundstrom G, Larsson TA, Brenner S, Venkatesh B, Larhammar D: **Evolution of the neuropeptide Y family: new genes by chromosome duplications in early vertebrates and in teleost fishes.** General and Comparative Endocrinology 2008, **155**(3):705-716.

38. Conlon JM, Bjenning C, Moon TW, Youson JH, Thim L: **Neuropeptide Y-related peptides from the pancreas of a teleostean (eel), holostean (bowfin) and elasmobranch (skate) fish.** Peptides 1991, **12**(2):221-226.

39. El-Salhy M, Grimelius L, Emson PC, Falkmer S: **Polypeptide YY-and neuropeptide Y-immunoreactive cells and nerves in the endocrine and exocrine pancreas of some vertebrates: an onto-and phylogenetic study.** The Histochemical Journal 1987, **19**(2):111-117.

543 40. Reinecke M, Weimar E, Maake C, Drakenberg K, Falkmer S, Sara VR: **IGF-2-like peptides are**
544 **present in insulin cells of the elasmobranchian endocrine pancreas: an immunohistochemical and**
545 **chromatographic study.** Histochemistry and Cell Biology 1994, **102**(5):365-371.
546 41. Moghimzadeh E, Ekman R, Hakanson R, Yanaihara N, Sundler F: **Neuronal gastrin-releasing**
547 **peptide in the mammalian gut and pancreas.** Neuroscience 1983, **10**(2):553-563.
548 42. Cornelio DB, Roesler R, Schwartsmann G: **Gastrin-releasing peptide receptor as a molecular**
549 **target in experimental anticancer therapy.** Annals of Oncology 2007, **18**(9):1457-1466.
550 43. Raddatz R, Wilson AE, Artymyshyn R, Bonini JA, Borowsky B, Boteju LW, Zhou S, Kouranova EV,
551 Nagorny R, Guevarra MS: **Identification and characterization of two neuromedin U receptors**
552 **differentially expressed in peripheral tissues and the central nervous system.** Journal of Biological
553 Chemistry 2000, **275**(42):32452-32459.
554 44. Frederickson RC: **Enkephalin pentapeptides--a review of current evidence for a physiological role**
555 **in vertebrate neurotransmission.** Life Sciences 1977, **21**(1):23.
556 45. Yoshimasa T, Nakao K, Ohtsuki H, Li S, Imura H: **Methionine-enkephalin and leucine-enkephalin**
557 **in human sympathoadrenal system and pheochromocytoma.** Journal of Clinical Investigation 1982,
558 **69**(3):643.
559 46. Leighton B, Atkinson A, Coghlan MP: **Small molecule glucokinase activators as novel anti-**
560 **diabetic agents.** Biochemical Society Transactions 2005, **33**(2):371-374.
561 47. Fajans SS, Bell GI, Polonsky KS: **Molecular mechanisms and clinical pathophysiology of**
562 **maturity-onset diabetes of the young.** New England Journal of Medicine 2001, **345**(13):971-980.
563 48. Vionnet N, Stoffel M, Takeda J, Yasuda K, Bell GI, Zouali H, Lesage S, Velho G, Iris F, Passa PH:
564 **Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes**
565 **mellitus. Nature** 1992 356(6371):721-2.
566 49. De Vos A, Heimberg H, Quartier E, Huypens P, Bouwens L, Pipeleers D, Schuit F: **Human and rat**
567 **beta cells differ in glucose transporter but not in glucokinase gene expression.** Journal of Clinical
568 Investigation 1995, **96**(5):2489.

569 50. Ferrer J, Benito C, Gomis R: **Pancreatic islet GLUT2 glucose transporter mRNA and protein**
570 **expression in humans with and without NIDDM.** Diabetes 1995, **44**(12):1369-1374.
571 51. Richardson CC, Hussain K, Jones PM, Persaud S, Lobner K, Boehm A, Clark A, Christie MR: **Low**
572 **levels of glucose transporters and channels in human pancreatic beta cells early in development.**
573 Diabetologia 2007, **50**(5):1000-1005.
574 52. Leonard J, Peers B, Johnson T, Ferreri K, Lee S, Montminy MR: **Characterization of *somatostatin***
575 ***transactivating factor-1*, a novel homeobox factor that stimulates somatostatin expression in**
576 **pancreatic islet cells.** Molecular Endocrinology 1993, **7**(10):1275.
577 53. Miller CP, McGehee Jr RE, Habener JF: ***IDX*-1: a new homeodomain transcription factor**
578 **expressed in rat pancreatic islets and duodenum that transactivates the somatostatin gene.** The
579 EMBO Journal 1994, **13**(5):1145.
580 54. Ohlsson H, Karlsson K, Edlund T: ***IPF1*, a homeodomain-containing transactivator of the insulin**
581 **gene.** The EMBO Journal 1993, **12**(11):4251.
582 55. Boam DS, Docherty K: **A tissue-specific nuclear factor binds to multiple sites in the human**
583 **insulin-gene enhancer.** Biochemical Journal 1989, **264**(1):233.
584 56. Marshak S, Benshushan E, Shoshkes M, Havin L, Cerasi E, Melloul D: **Functional conservation of**
585 **regulatory elements in the pdx-1 gene: PDX-1 and hepatocyte nuclear factor 3beta transcription**
586 **factors mediate beta-cell-specific expression.** Molecular and Cellular Biology 2000, **20**(20):7583.
587 57. Ohneda K, Mirmira RG, Wang J, Johnson JD, German MS: **The homeodomain of PDX-1 mediates**
588 **multiple protein-protein interactions in the formation of a transcriptional activation complex on the**
589 **insulin promoter.** Molecular and Cellular Biology 2000, **20**(3):900.
590 58. Docherty HM, Hay CW, Ferguson LA, Barrow J, Durward E, Docherty K: **Relative contribution of**
591 **PDX-1, MafA and E47/β2 to the regulation of the human insulin promoter.** Biochemical Journal
592 2005, **389**(3):813.
593 59. Stoffers DA, Ferrer J, Clarke WL, Habener JF: **Early-onset type-II diabetes mellitus (MODY4)**
594 **linked to IPF1.** Nature Genetics 1997, **17**:138-141.

595 60. Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, Habener JF: **Pancreatic agenesis attributable to**
596 **a single nucleotide deletion in the human IPF1 gene coding sequence.** Nature Genetics 1997,
597 **15**(1):106-110.
598 61. Schwitzgebel VM, Mamin A, Brun T, Ritz-Laser B, Zaiko M, Maret A, Jornayvaz FR, Theintz GE,
599 Michielin O, Melloul D: **Agenesis of human pancreas due to decreased half-life of insulin promoter**
600 **factor 1.** Journal of Clinical Endocrinology & Metabolism 2003, **88**(9):4398.
601 62. Mulley JF, Holland PWH: **Genomic Organisation of the Seven ParaHox Genes of Coelacanths.**
602 Journal of Experimental Zoology Part B: Molecular and Developmental Evolution doi:
603 10.1002/jez.b.22513.
604 63. Poulin G, Turgeon B, Drouin J: **NeuroD1/beta2 contributes to cell-specific transcription of the**
605 **proopiomelanocortin gene.** Molecular and Cellular Biology 1997, **17**(11):6673-6682.
606 64. Chao CS, Loomis ZL, Lee JE, Sussel L: **Genetic identification of a novel NeuroD1 function in the**
607 **early differentiation of islet α, PP and ε cells.** Developmental Biology 2007, **312**(2):523-532.
608 65. Chae JH, Stein GH, Lee JE: **NeuroD: the predicted and the surprising.** Molecules and Cells 2004,
609 **18**(3):271.
610 66. Malecki MT, Jhala US, Antonellis A, Fields L, Doria A, Orban T, Saad M, Warram JH, Montminy M,
611 Krolewski AS: **Mutations in NEUROD1 are associated with the development of type 2 diabetes**
612 **mellitus.** Nature Genetics 1999, **23**(3):323-328.
613 67. Naya FJ, Stellrecht CM, Tsai MJ: **Tissue-specific regulation of the insulin gene by a novel basic**
614 **helix-loop-helix transcription factor.** Genes & Development 1995, **9**(8):1009-1019.
615 68. Naya FJ, Huang HP, Qiu Y, Mutoh H, DeMayo FJ, Leiter AB, Tsai MJ: **Diabetes, defective**
616 **pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-**
617 **deficient mice.** Genes & Development 1997, **11**(18):2323-2334.
618 69. Qiu Y, Guo M, Huang S, Stein R: **Insulin gene transcription is mediated by interactions between**
619 **the p300 coactivator and PDX-1, BETA2, and E47.** Molecular and Cellular Biology 2002, **22**(2):412-
620 420.

621 70. Zhang H, Wang WP, Guo T, Yang JC, Chen P, Ma KT, Guan YF, Zhou CY: **The LIM-**
622 **homeodomain protein ISL1 activates insulin gene promoter directly through synergy with BETA2.**
623 Journal of Molecular Biology 2009, **392**(3):566-577.
624 71. Peng SY, Wang WP, Meng J, Li T, Zhang H, Li Y, Chen P, Ma KT, Zhou CY: **ISL1 physically**
625 **interacts with BETA2 to promote insulin gene transcriptional synergy in non-beta cells.** Biochimica
626 et Biophysica Acta (BBA)-Gene Structure and Expression 2005, **1731**(3):154-159.
627 72. Glick E, Leshkowitz D, Walker MD: **Transcription factor BETA2 acts cooperatively with E2A**
628 **and PDX1 to activate the insulin gene promoter.** Journal of Biological Chemistry 2000, **275**(3):2199-
629 2204.
630 73. Courtois G, Morgan JG, Campbell LA, Fourel G, Crabtree GR: **Interaction of a liver-specific**
631 **nuclear factor with the fibrinogen and alpha 1-antitrypsin promoters.** Science 1987, **238**(4827):688-
632 692.
633 74. Brooks AR, Levy-Wilson B: **Hepatocyte nuclear factor 1 and C/EBP are essential for the activity**
634 **of the human apolipoprotein B gene second-intron enhancer.** Molecular and Cellular Biology 1992,
635 **12**(3):1134-1148.
636 75. Maire P, Wuarin J, Schibler U: **The role of cis-acting promoter elements in tissue-specific albumin**
637 **gene expression.** Science 1989, **244**(4902):343-346.
638 76. Emens LA, Landers DW, Moss LG: **Hepatocyte nuclear factor 1 alpha is expressed in a hamster**
639 **insulinoma line and transactivates the rat insulin I gene.** Proceedings of the National Academy of
640 Sciences of the United States of America 1992, **89**(16):7300-7304.
641 77. Pontoglio M, Sreenan S, Roe M, Pugh W, Ostrega D, Doyen A, Pick AJ, Baldwin A, Velho G,
642 Froguel P: **Defective insulin secretion in hepatocyte nuclear factor 1 alpha-deficient mice.** Journal of
643 Clinical Investigation 1998, **101**(10):2215-2221.
644 78. Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, Southam L, Cox RD, Lathrop
645 GM, Boriraj VV: **Mutations in the hepatocyte nuclear factor-1 alpha gene in maturity-onset diabetes**
646 **of the young (MODY3).** Nature 1996 **384**(6608):455-458.

79. Vaxillaire M, Rouard M, Yamagata K, Oda N, Kaisaki PJ, Boriraj VV, Chevre JC, Boccio V, Cox RD, Lathrop GM: **Identification of Nine Novel Mutations in the Hepatocyte Nuclear Factor 1 Alpha Aene Associated with Maturity-Onset Diabetes of the Young (MODY3).** Human Molecular Genetics 1997, **6**(4):583-586.

80. Bjorkhaug L, Sagen JV, Thorsby P, Sovik O, Molven A, Njolstad PR: **Hepatocyte nuclear factor-1a gene mutations and diabetes in Norway.** Journal of Clinical Endocrinology and Metabolism 2003, **88**:920-931.

81. Iype T, Taylor DG, Ziesmann SM, Garmey JC, Watada H, Mirmira RG: **The transcriptional repressor Nkx6. 1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic β-cell differentiation: evidence for feedback activation of the nkx6.1 gene by Nkx6. 1.** Molecular Endocrinology 2004, **18**(6):1363-1375.

82. Sander M, Sussel L, Conners J, Scheel D, Kalamaras J, Cruz FD, Schwitzgebel V, Hayes-Jordan A, German M: **Homeobox gene Nkx6. 1 lies downstream of Nkx2. 2 in the major pathway of beta-cell formation in the pancreas.** Development 2000, **127**(24):5533-5540.

83. Schisler JC, Jensen PB, Taylor DG, Becker TC, Knop FK, Takekawa S, German M, Weir GC, Lu D, Mirmira RG: **The Nkx6. 1 homeodomain transcription factor suppresses glucagon expression and regulates glucose-stimulated insulin secretion in islet beta cells.** Proceedings of the National Academy of Sciences of the United States of America 2005, **102**(20):7297-7302.

84. Aramata S, Han SI, Kataoka K: **Roles and regulation of transcription factor MafA in islet beta-cells.** Endocrine Journal 2007, **54**(5):659.

85. Zhang C, Moriguchi T, Kajihara M, Esaki R, Harada A, Shimohata H, Oishi H, Hamada M, Morito N, Hasegawa K: **MafA is a key regulator of glucose-stimulated insulin secretion.** Molecular and Cellular Biology 2005, **25**(12):4969-4976.

86. Dorrell C, Schug J, Lin CF, Canaday PS, Fox AJ, Smirnova O, Bonnah R, Streeter PR, Stoeckert CJ, Kaestner KH: **Transcriptomes of the major human pancreatic cell types.** Diabetologia 2011, **54**(11):2832-44.

673 87. Kutlu B, Burdick D, Baxter D, Rasschaert J, Flamez D, Eizirik D, Welsh N, Goodman N, Hood L:
674 **Detailed transcriptome atlas of the pancreatic beta cell.** BMC Medical Genomics 2009, **2**(1):3.
675 88. Kong YM, MacDonald RJ, Wen X, Yang P, Barbera VM, Swift GH: **A comprehensive survey of**
676 **DNA-binding transcription factor gene expression in human fetal and adult organs.** Gene Expression
677 Patterns 2006, **6**(7):678-686.
678 89. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** Nucleic Acids Research
679 2000, **28**(1):27-30.
680 90. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation**
681 **and pathway reconstruction server.** Nucleic Acids Research 2007, **35**(2):W182-W185.
682 91. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information,**
683 **knowledge and principle: back to metabolism in KEGG.** Nucleic Acids Research 2014, **42**(D1):D199-
684 D205.
685 92. Taylor JS, Van de Peer Y, Meyer A: **Genome duplication, divergent resolution and speciation.**
686 Trends in Genetics 2001, **17**(6):299-301.
687 93. Jensen J: **Gene regulatory factors in pancreatic development.** Developmental Dynamics 2003,
688 **229**(1):176-200.
689 94. Habener JF, Kemp DM, Thomas MK: **Minireview: transcriptional regulation in pancreatic**
690 **development.** Endocrinology 2005, **146**(3):1025.
691 95. Rudnick A, Ling TY, Odagiri H, Rutter WJ, German MS: **Pancreatic beta cells express a diverse set**
692 **of homeobox genes.** Proceedings of the National Academy of Sciences of the United States of America
693 1994, **91**(25):12203.
694 96. Zhong Y, Holland PWH: **HomeoDB2: functional expansion of a comparative homeobox gene**
695 **database for evolutionary developmental biology.** Evolution & Development 2011, **13**(6):567-568.
696 97. Zhong YF, Butts T, Holland PWH: **HomeoDB: a database of homeobox gene diversity.** Evolution
697 & Development 2008, **10**(5):516-518.

698 98. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall
699 M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A: **Pfam: clans, web tools and services.**
700 Nucleic Acids Research 2006, **34**(1):D247-D251.
701 99. Ballantyne JS: **Jaws: the inside story. The metabolism of elasmobranch fishes.** Comparative
702 Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 1997, **118**(4):703-742.
703 100. Moon TW, Mommsen TP: **Enzymes of intermediary metabolism in tissue of the little skate, *Raja***
704 ***erinacea*.** Journal of Experimental Zoology 2005, **244**(1):9-15.
705 101. Walsh PJ, Kajimura M, Mommsen TP, Wood CM: **Metabolic organization and effects of feeding**
706 **on enzyme activities of the dogfish shark (*Squalus acanthias*) rectal gland.** Journal of Experimental
707 Biology 2006, **209**(15):2929-2938.
708 102. Zendzian EN, Barnard EA: **Distributions of pancreatic ribonuclease, chymotrypsin, and trypsin**
709 **in vertebrates.** Archives of Biochemistry and Biophysics 1967, **122**(3):699-713.
710 103. Prahl JW, Neurath H: **Pancreatic Enzymes of the Spiny Pacific Dogfish. I. Cationic**
711 **Chymotrypsinogen and Chymotrypsin.** Biochemistry 1966, **5**(6):2131-2146.
712 104. Prahl JW, Neurath H: **Pancreatic Enzymes of the Spiny Pacific Dogfish. II. Procarboxypeptidase**
713 **B and Carboxypeptidase B.** Biochemistry 1966, **5**(12):4137-4145.
714 105. Neurath H, Lacko AG: **Procarboxypeptidase A and carboxypeptidase A of the spiny Pacific**
715 **dogfish (*Squalus acanthias*).** Biochemistry 1970, **9**(24):4680-4690.
716 106. Dong W, Fricker LD, Day R: **Carboxypeptidase D is a potential candidate to carry out**
717 **redundant processing functions of carboxypeptidase E based on comparative distribution studies in**
718 **the rat central nervous system.** Neuroscience 1999, **89**(4):1301-1317.
719 107. Fricker LD: **Carboxypeptidase E.** Annual Review of Physiology 1988, **50**(1):309-321.
720 108. Brockerhoff H, Hoyle RJ: **Hydrolysis of triglycerides by the pancreatic lipase of a skate.**
721 Biochimica et Biophysica Acta 1965, **98**:435.
722 109. Patton JS: **High levels of pancreatic nonspecific lipase in rattlesnake and leopard shark.** Lipids
723 1975, **10**(9):562-564.

724 110. Patton JS, Warner TG, Benson AA: **Partial characterization of the bile salt-dependent**
725 **triacylglycerol lipase from the leopard shark pancreas.** Biochimica et Biophysica Acta (BBA)-Lipids
726 and Lipid Metabolism 1977, **486**(2):322-330.
727 111. Sternby B, Larsson A, Borgstrom B: **Evolutionary studies on pancreatic colipase.** Biochimica et
728 Biophysica Acta (BBA)-Lipids and Lipid Metabolism 1983, **750**(2):340-345.
729 112. Sternby B, Engstrom A, Hellman U: **Purification and characterization of pancreatic colipase**
730 **from the dogfish (*Squalus acanthius*).** Biochimica et Biophysica Acta (BBA)-Protein Structure and
731 Molecular Enzymology 1984, **789**(2):159-163.
732 113. Bacha AB, Karray A, Daoud L, Bouchaala E, Ali MB, Gargouri Y, Ali YB: **Biochemical properties**
733 **of pancreatic colipase from the common stingray *Dasyatis pastinaca*.** Lipids in health and disease
734 2011, **10**(1):69.
735 114. Arntfield ME, van der Kooy D: **β-Cell evolution: How the pancreas borrowed from the brain.**
736 Bioessays 2011, **33**(8):582-587.
737 115. Upchurch BH, Aponte GW, Leiter AB: **Expression of peptide YY in all four islet cell types in the**
738 **developing mouse pancreas suggests a common peptide YY-producing progenitor.** Development
739 1994, **120**(2):245-252.
740 116. Schonhoff S, Baggio L, Ratineau C, Ray SK, Lindner J, Magnuson MA, Drucker DJ, Leiter AB:
741 **Energy homeostasis and gastrointestinal endocrine differentiation do not require the anorectic**
742 **hormone peptide YY.** Molecular and Cellular Biology 2005, **25**(10):4189-4199.
743 117. Chang C, Shen W, Rozenfeld S, Lawrence HJ, Largman C, Cleary ML: **Pbx proteins display**
744 **hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins.** Genes &
745 Development 1995, **9**(6):663-674.
746 118. Liu A, Desai BM, Stoffers DA: **Identification of PCIF1, a POZ domain protein that inhibits**
747 **PDX-1 (MODY4) transcriptional activity.** Molecular and Cellular Biology 2004, **24**(10):4372-4383.

748 119. Peshavaria M, Henderson E, Sharma A, Wright CV, Stein R: **Functional characterization of the**
749 **transactivation properties of the PDX-1 homeodomain protein.** Molecular and Cellular Biology 1997,
750 **17**(7):3987-3996.
751 120. Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts
752 T, Candiani S, Dishaw LJ: **The amphioxus genome illuminates vertebrate origins and**
753 **cephalochordate biology.** Genome Research 2008, **18**(7):1100-1111.
754 121. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M,
755 Shoguchi E, Terry A, Yu Jr K: **The amphioxus genome and the evolution of the chordate karyotype.**
756 Nature 2008, **453**(7198):1064-1072.
757 122. Le Roith D, Shiloach J, Roth J: **Is there an earlier phylogenetic precursor that is common to both**
758 **the nervous and endocrine systems?** Peptides 1982, **3**(3):211-215.
759 123. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
760 Raychowdhury R, Zeng Q: **Full-length transcriptome assembly from RNA-Seq data without a**
761 **reference genome.** Nature Biotechnology 2011, **29**(7):644-652.
762 124. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for**
763 **annotation, visualization and analysis in functional genomics research.** Bioinformatics 2005,
764 **21**(18):3674-3676.
765 125. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M,
766 Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO**
767 **suite.** Nucleic acids research 2008, **36**(10):3420-3435.
768 126. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+:**
769 **architecture and applications.** BMC bioinformatics 2009, **10**(1):421.
770
771
772
773

774 **Figure legends**

775

776 **Figure 1.** Phylogenetic tree of the major extant vertebrate groups. The relationships of the most common
777 chondrichthyan (cartilaginous fish) model species (Elephant shark, *Callorhinchus milii*; Little skate,
778 *Leucoraja erinacea*; Lesser spotted catshark, *Scyliorhinus canicula*; Spiny dogfish, *Squalus acanthias*) are
779 shown, as are representative lineages from the ray-finned (actinopterygian) and lobe-finned
780 (sarcopterygian) fish. The origin of the combined endocrine and exocrine pancreatic gland at the base of
781 the jawed vertebrates is indicated.

782

783 **Figure 2.** Tissue distribution of transcripts, as determined by mapping the sequencing reads derived from
784 each tissue to a combined, all-tissue assembly. Contig values of ≥1 FPKM (Fragments Per Kilobase of
785 exon per Million fragments mapped) were taken as evidence for expression.

786

787 **Figure 3.** Length and tissue distribution of open reading frames (ORFs) derived from assembled contigs

788

789 **Figure 4.** Proportion of transcripts assigned to each of the top 25 gene ontology (GO) slim 'Biological
790 Process' terms for catshark pancreas, brain and liver tissue-specific transcripts.

791

792 **Figure 5.** Proportion of transcripts assigned to each of the top 25 gene ontology (GO) slim 'Molecular
793 Function' terms for catshark pancreas, brain and liver tissue-specific transcripts.

794

795 **Figure 6.** Annotation of the precursor peptides of catshark preproinsulin and preproglucagon. Signal
796 peptides (amino acids 1-24 and 1-20 respectively) are underlined and basic amino acid cleavage sites are
797 lowercase. Glucagon-like peptides (GLP) 1a and 1b are annotated based on similarity to the duplicated
798 GLP1 peptides in the unpublished *Squalus acanthias* and *Hydrolagus colliei* proglucagon sequences

799 available on Genbank (accession numbers AAS57653 and AAS57654). An oxyntomodulin-like peptide
800 has been purified from *H. colliei* and corresponds to amino acids 47-82 in the Catshark preproglucagon.
801
802 **Figure 7.** Amino acid alignment of vertebrate Peptide YY (PYY), Neuropeptide Y (NPY) and Pancreatic
803 polypeptide (PP) sequences. Genbank accession numbers are given in square brackets. Sca, Scyliorhinus
804 canicula (lesser spotted catshark); Sac, Squalus acanthias (spiny dogfish); Ler, Leucoraja erinacea (little
805 skate); Cmi, Callorhinchus milii (elephant shark); Hsa, Homo sapiens (human); Lfl, Lampetra planeri
806 (brook lamprey); Loc, Leucoraja ocellata (winter skate).
807
808 **Figure 8.** Immunolocalization of pancreatic hormones and pancreatic polypeptide and PYY in catshark
809 pancreas. (A) The distribution of the pancreatic hormones insulin (blue), glucagon (green) and
810 somatostatin (Red) in uniquely shaped islet structures. (B) Pancreatic polypeptide (red) specific antisera
811 fail to stain a specific subset of endocrine cells in the pancreas, while insulin (blue) and glucagon show a
812 normal distribution. (C-D) PYY shows colocalization with most of the insulin immunoreactive cells but
813 not glucagon or somatostatin. All images are 250x magnification
814
815 **Figure 9.** Protein domains in vertebrate PDX1 and PDX2. Transactivation domains A-E [119], PCIF1-
816 interaction domains [118], homeodomains, DNA-binding domains (i) and nuclear localisation signals (ii)
817 are highlighted. Domain E contains the PBX-interacting hexapeptide motif [117]. There is very little
818 conservation of amino acid sequence between the paralogous PDX1 and PDX2 suggesting that they carry
819 out distinct functions within the pancreas, although clearly both are localised to the nucleus, bind DNA
820 and interact with PBX proteins. Hsa, human (*Homo sapiens*); Mmu, mouse (*Mus musculus*); Rno, rat
821 (*Rattus norvegicus*); Gga, chicken (*Gallus gallus*); Acar, Anole lizard (*Anolis carolinensis*); Xla, *Xenopus*
822 *laevis*; Xtr, *Xenopus tropicalis*; Lme, Indonesian coelacanth (*Latimeria menadoensis*); Lch, African
823 coelacanth (*Latimeria chalumnae*); Acal, Bowfin (*Amia calva*); Loc, Spotted gar (*Lepisosteus oculatus*);
824 Dre, zebrafish (*Danio rerio*); Tru, fugu (*Takifugu rubripes*); Ola, medaka (*Oryzias latipes*); Gac,

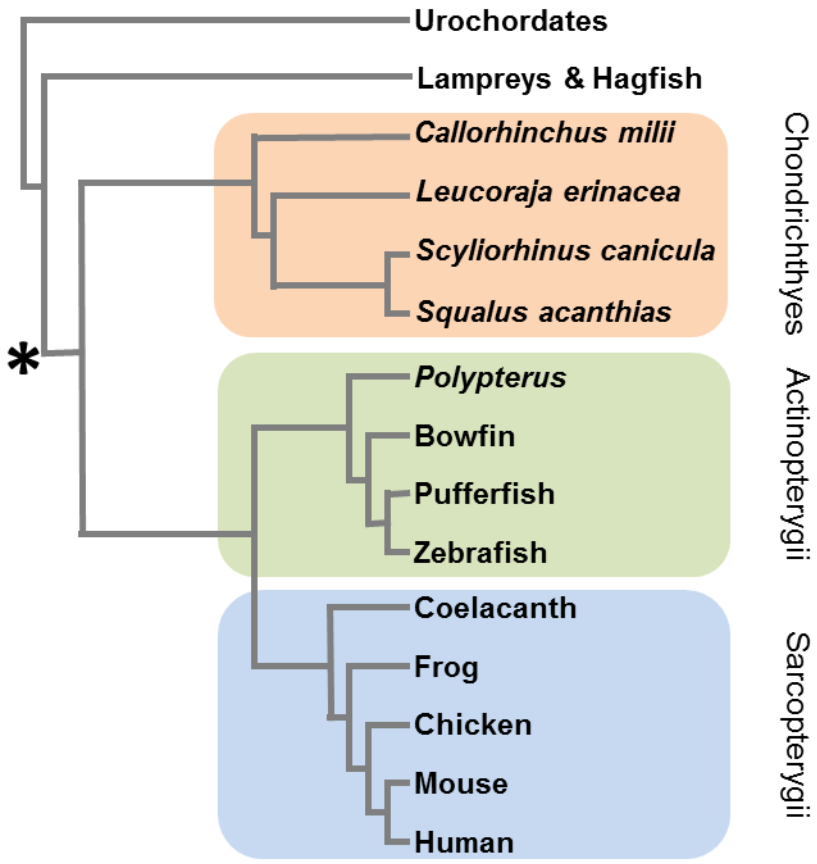
825 stickleback; Tni, Green spotted puffer (*Tetraodon nigroviridis*); Sca, lesser spotted catshark (*Scyliorhinus*

826 *canicula*); Ler, little skate (*Leucoraja erinacea*); Cmi, elephant shark (*Callorhinchus milii*); Bfl,

827 amphioxus (*Branchiostoma floridae*)

828

829



830

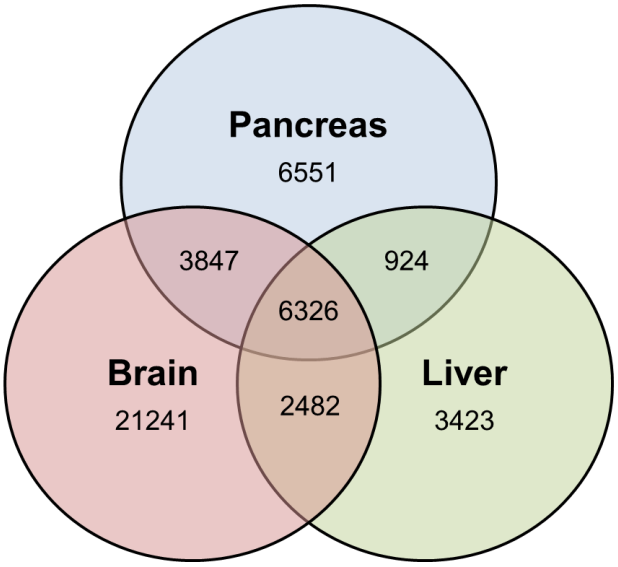
831

832

833

834

835



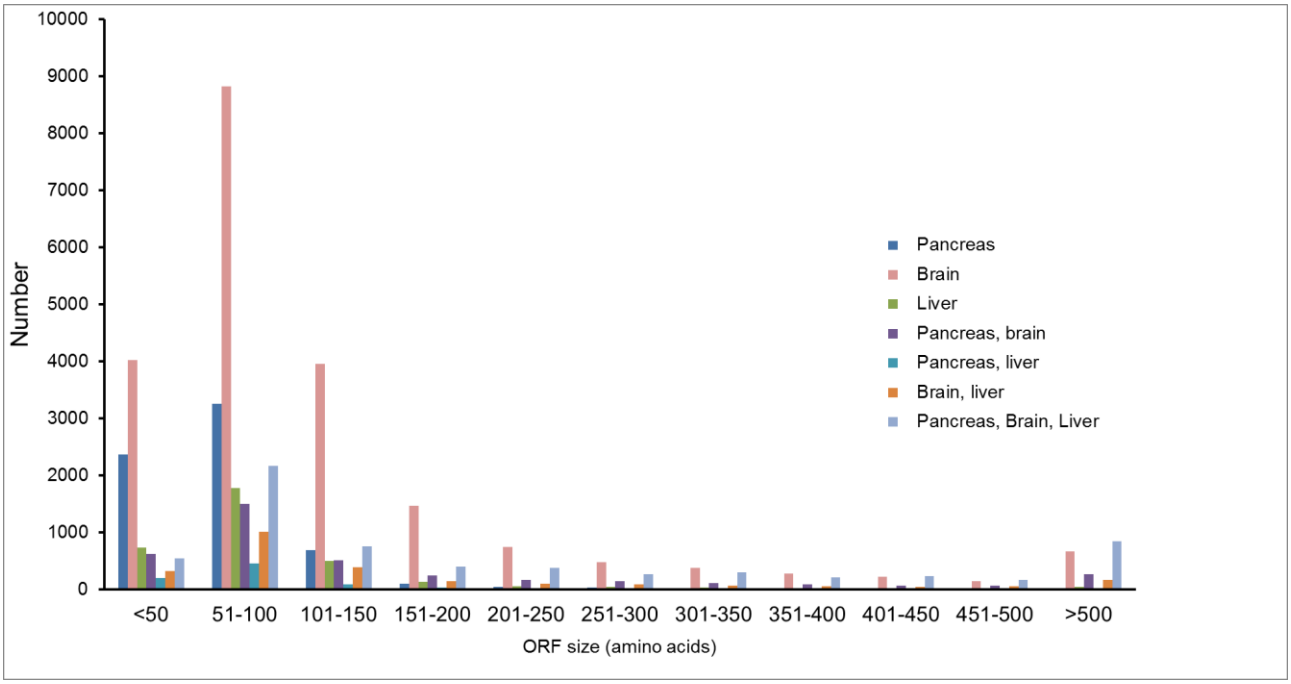
836

837

838

839

840



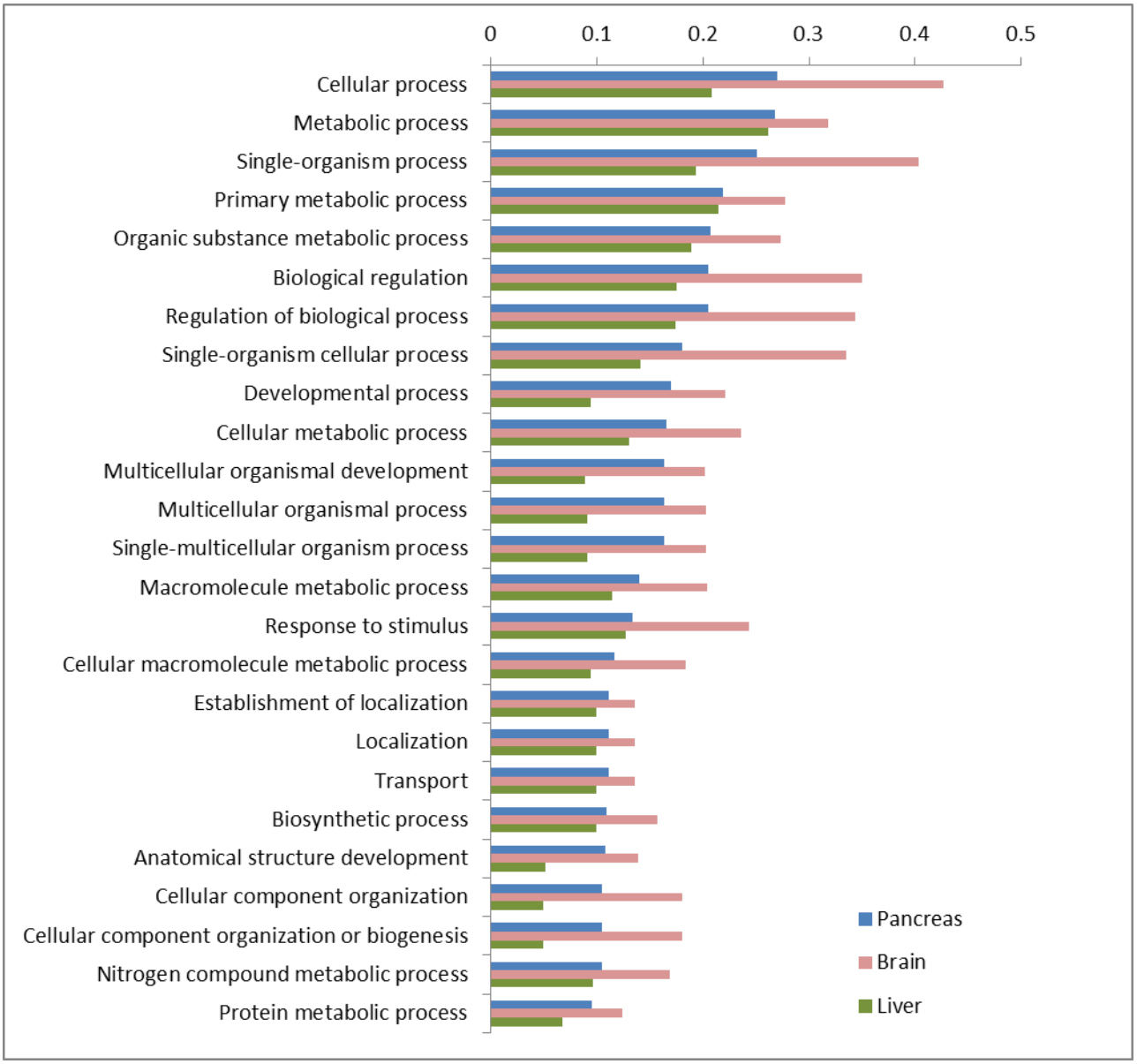
841

842

843

844

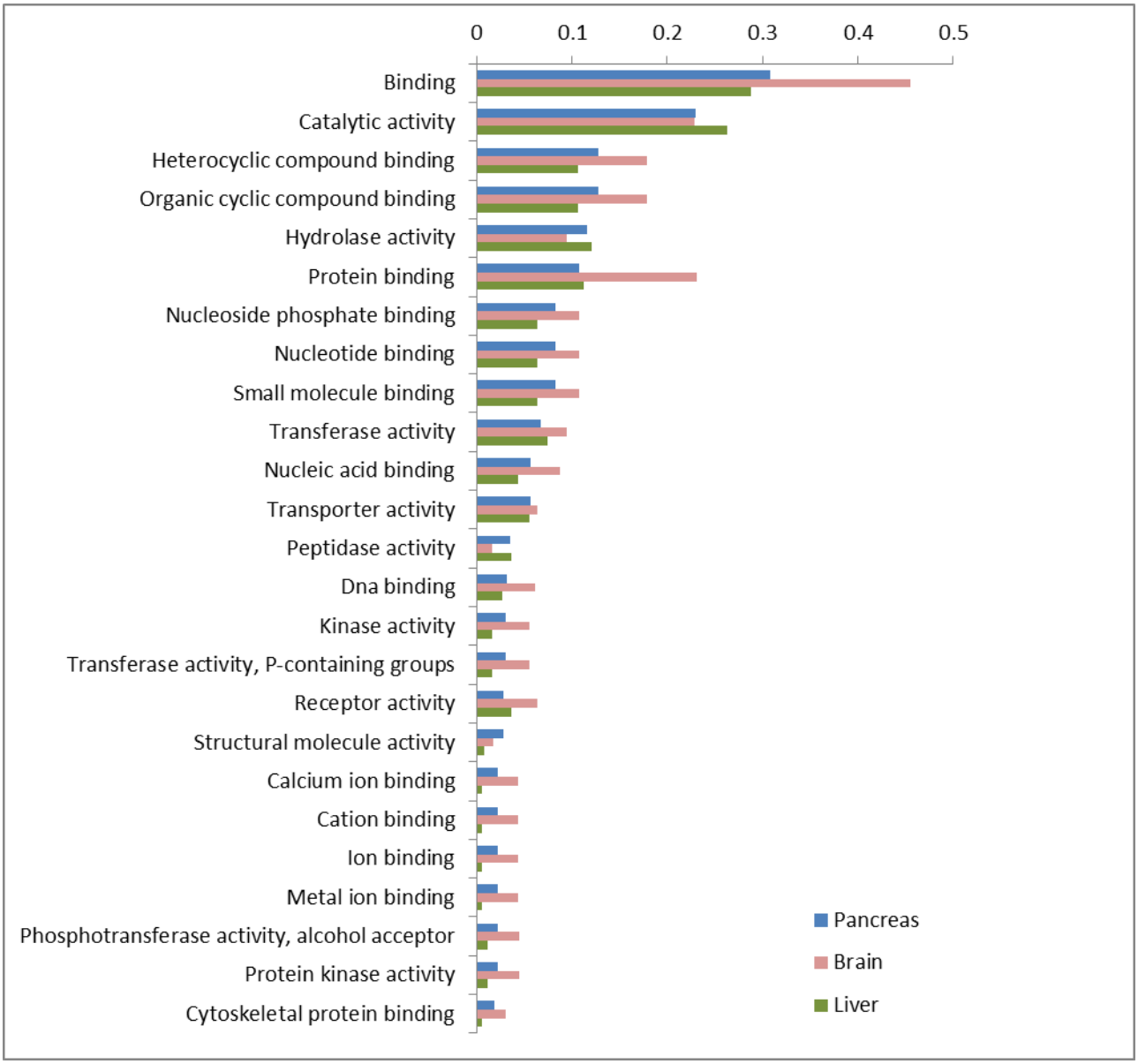
845



846

847

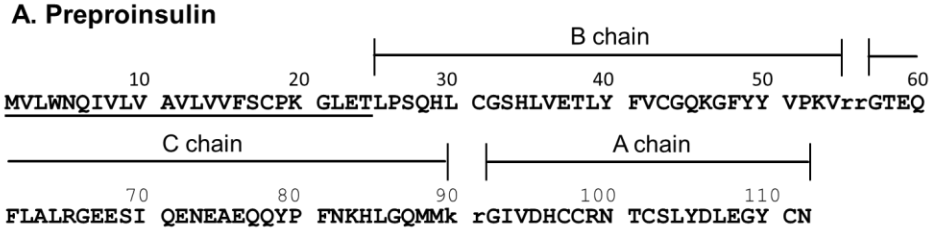
848



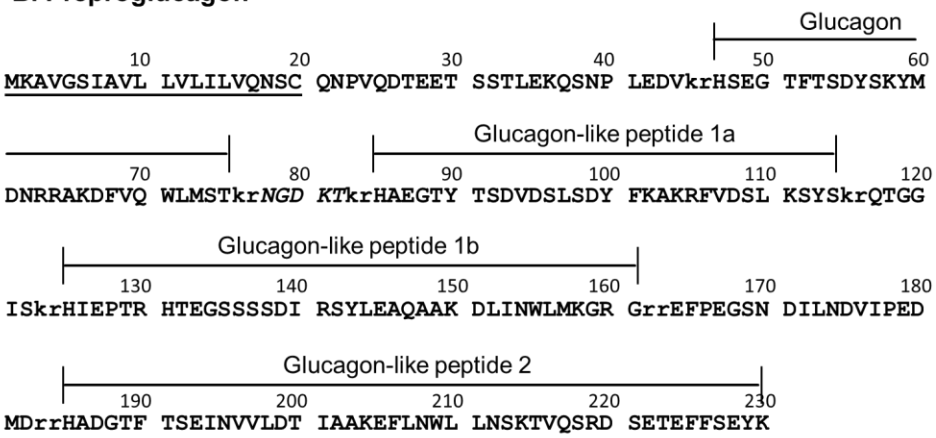
849

850

A. Preproinsulin



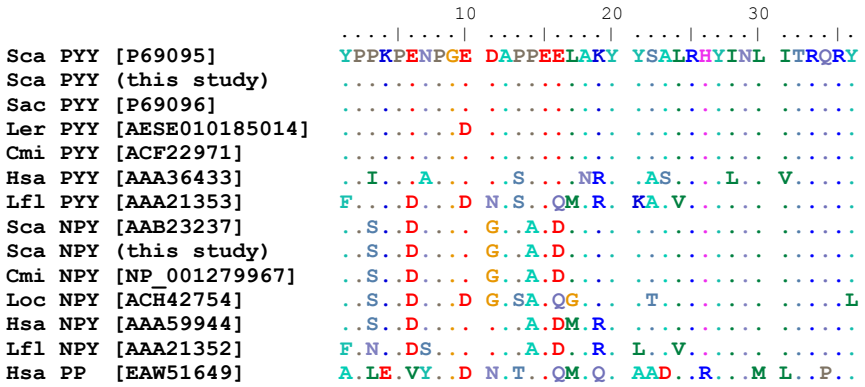
B. Preproglucagon



851

852

853

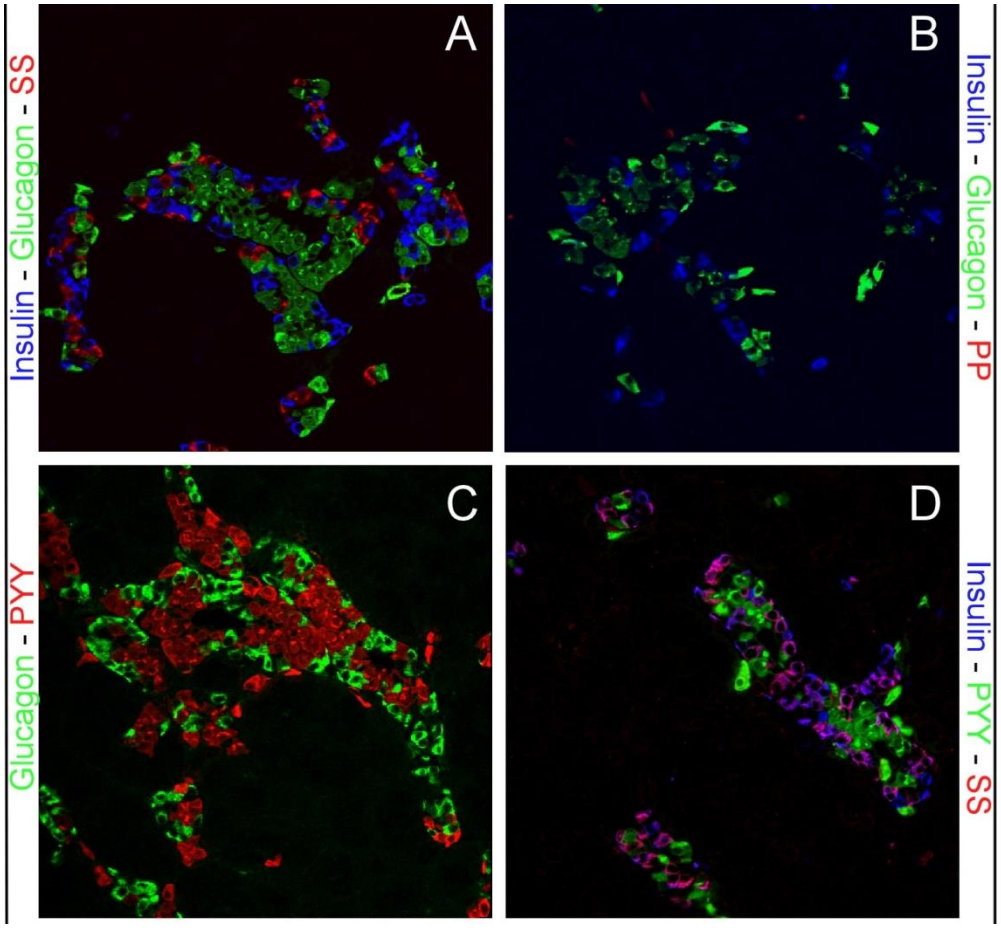


854

855

856

857

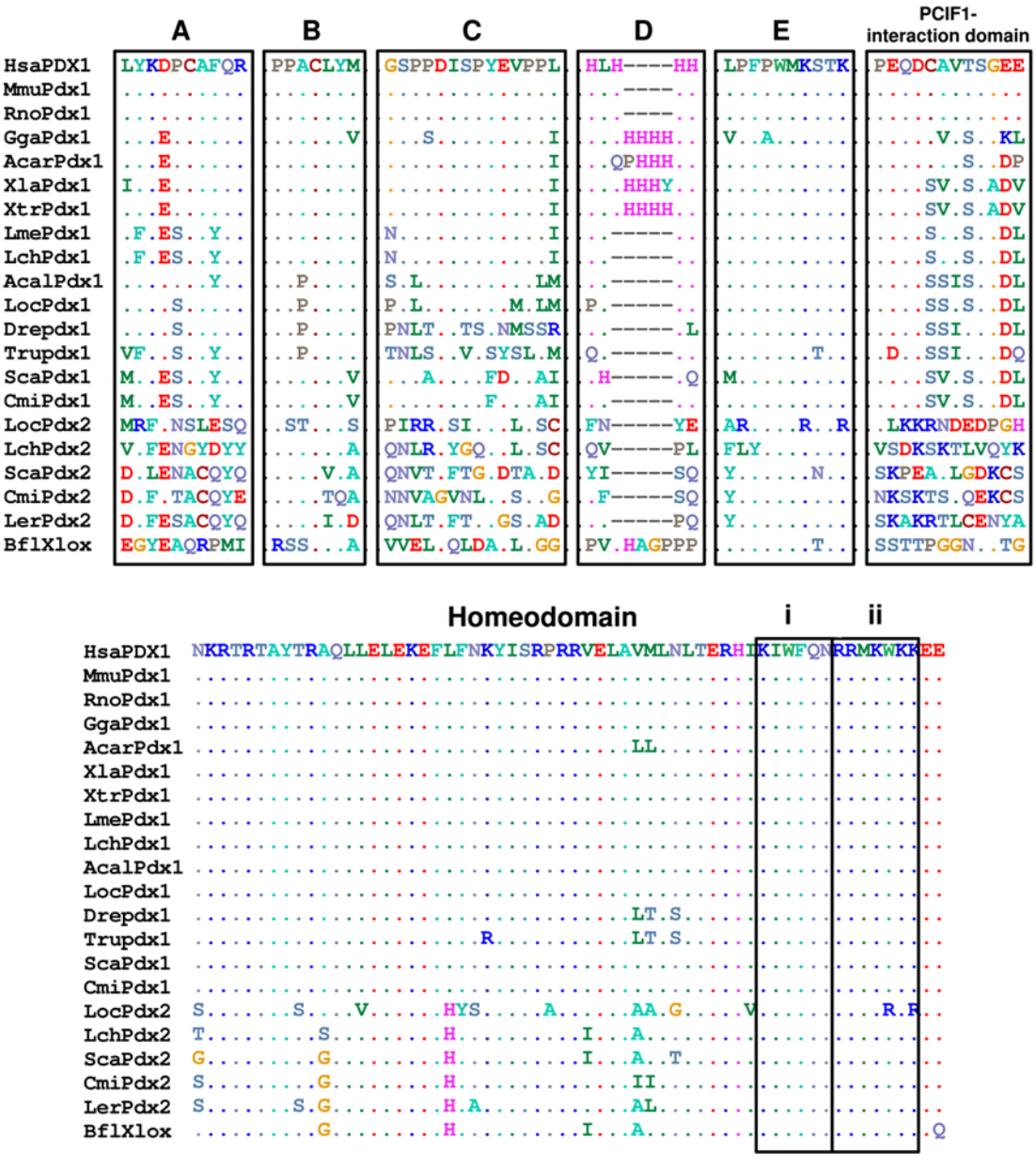


858

859

860

861



862

863 **Table 1.** Assembly characteristics. Tissue distribution of transcripts was assigned by read mapping, taking

864 a value of ≥1 fragments per kilobase per million mapped reads (FPKM) as evidence of expression.

Tissue(s)	Number of contigs >300bp	Contig N50 (bp)	Longest contig (bp)	Mean ORF length (bp)	Max ORF length (bp)	Number ORFs with signal peptide
Pancreas only	6551	429	5386	177	3005	257 (3.92%)
Brain only	21241	798	16085	338	13751	1029 (4.84%)
Liver only	3423	676	13906	271	13532	220 (6.43%)
Pancreas, brain	3847	1479	10710	482	10394	210 (5.46%)
Pancreas, liver	924	1039	5907	329	4505	62 (6.71%)
Brain, liver	2482	1329	12080	481	10403	149 (6.00%)
Pancreas, liver, brain	6326	2257	11135	705	9329	356 (5.63%)

865

866

867 **Table 2.** Number of contigs with a BLAST hit in the RefSeq database and Gene Ontology (GO)
868 annotation assigned by BLAST2GO [124, 125]. The number of transcription factors and signaling
869 pathway components in each tissue or tissue combination is also shown.

	Number contigs with BLAST hit	Number contigs with GO annotation	Number of transcription factors	Number of signaling pathway components
Pancreas only	998 (15%)	766 (12%)	13	38
Brain only	8093 (38%)	6740 (32%)	145	359
Liver only	830 (24%)	661 (19%)	12	37
Pancreas, brain	1640 (43%)	1412 (37%)	33	104
Pancreas, liver	249 (27%)	200 (22%)	14	11
Brain, liver	1019 (41%)	909 (37%)	10	83
Pancreas, liver, brain	3380 (53%)	3141 (49%)	51	187

870

871

872

873

874 **Table 3.** Transcription factors and signaling pathway components unique to the lesser spotted catshark

875 pancreas based on available data.

Pancreas-specific transcription factors	Pancreas-specific signaling components
Forkhead box protein A1	1D-myo-inositol-triphosphate 3-kinase
GATA-binding protein 4	5-hydroxytryptamine receptor 2
Histone-lysine N-methyltransferase MLL1	Adenylate cyclase 2
Homeobox protein hoxB5	Bone morphogenetic protein 2
Homeobox protein hoxB7	Bone morphogenetic protein 7
Homeobox protein MOX	Cholecystokinin A receptor
Pancreas and duodenal homeobox 1	Collagen, type I/II/III/V/XI/XXIV/XXVII, alpha
Pancreas and duodenal homeobox 2	Collagen, type IV, alpha
Krueppel-like factor 5	Collagen, type VI, alpha
Nuclear receptor subfamily 0 group B 1	Cysteinyl leukotriene receptor 1
Nuclear receptor subfamily 0 group B 1	Epidermal growth factor receptor
Pancreas-specific transcription factor 1a	Fibroblast growth factor
Transcriptional enhancer factor	FMS-like tyrosine kinase 1
Zinc finger protein GLI3	Frizzled 9/10
	Glutamine synthetase
	Inositol 1,4,5-triphosphate receptor type 1
	Inositol 1,4,5-triphosphate receptor type 2
	Inositol 1,4,5-triphosphate receptor type 3
	Insulin
	Integrin alpha 1
	Integrin alpha 2
	Integrin beta 6

	Interleukin 10 receptor beta Janus kinase 2 Laminin, alpha 3/5 Laminin, beta 4 Leucine-rich repeats and death domain-containing protein Mitogen-activated protein kinase kinase kinase kinase 4 Nucleoprotein TPR Protein crumbs Receptor-interacting serine/threonine-protein kinase 1 Secretory phospholipase A2 Suppressor of cytokine signaling 1 TGF-beta receptor type-2 Transcriptional enhancer factor Transferrin Vascular endothelial growth factor C/D Zinc finger protein GLI3

876

877

878

879

880

881

882 **Table 4.** Predicted di-, tri- and tetranucleotide simple sequence repeats (microsatellites) in each catshark
883 tissue or combination or tissues. Results are shown for both full transcripts and predicted open reading
884 frames (ORFs) and in both cases dinucleotide repeats are the most common.

	Transcript			Predicted ORFs		
	Di-	Tri-	Tetra-	Di-	Tri-	Tetra-
Pancreas only	456 (6.96%)	23 (0.35%)	3 (0.05%)	60 (0.92%)	6 (0.09%)	0 (0%)
Brain only	2808 (13.22%)	268 (1.26%)	7 (0.03%)	829 (3.90%)	147 (0.69%)	1 (<0.01%)
Liver only	446 (13.03%)	24 (0.70%)	3 (0.09%)	168 (4.90%)	12 (0.35%)	1 (0.03%)
Pancreas + Brain	481 (12.50%)	43 (1.12%)	2 (0.05%)	72 (1.87%)	15 (0.39%)	2 (0.05%)
Pancreas + Liver	105 (11.36%)	5 (0.54%)	1 (0.11%)	44 (4.76%)	3 (0.32%)	1 (0.11%)
Brain + Liver	702 (28.28%)	61 (2.46%)	3 (0.12%)	106 (11.47%)	21 (2.27%)	0 (0%)
Pancreas + Liver + Brain	1240 (19.60%)	158 (2.50)	4 (0.06%)	308 (4.87%)	108 (1.71%)	2 (0.03%)

885

886

887

888

889

890 **Table 5.** Peptide diversity of the catshark pancreas, brain and liver. Our comprehensive transcriptomic
891 survey of the lesser spotted catshark pancreas highlights the disparity in the estimation of peptide diversity
892 in early vertebrates as previously suggested by immunohistochemical (IHC) studies and highlights the
893 similarity of pancreas and brain peptide complements.

Peptide	Pancreas IHC	Pancreas transcriptome	Brain transcriptome	Liver transcriptome
Insulin	+	+	-	-
Glucagon	+	+	-	-
Somatostatin	+	+	+	-
Pancreatic polypeptide	+	-	-	-
Peptide YY	+	+	+	-
Gastrin-releasing peptide	+	+	+	-
Neuromedin U	+	+	+	+
Enkephalin	+	+	+	-
Cholecystokinin	+	-	+	-
Gastrin	+	-	-	-
Vasoactive intestinal polypeptide	+	+	+	-
Gastric inhibitory polypeptide	+	-	-	-
Secretin	+	-	-	-

894

895

896

897

898 **Additional files**

899 Additional file 1: Combined tissue assembly, trimmed to remove contigs <300bp

900 Additional file 2: Pancreas-specific transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

901 Additional file 3: Brain-specific transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

902 Additional file 4: Liver-specific transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

903 Additional file 5: Pancreas/brain transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

904 Additional file 6: Pancreas/liver transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

905 Additional file 7: Brain/liver transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

906 Additional file 8: Pancreas/brain/liver transcripts present at ≥1 FPKM, trimmed to remove contigs <300bp

907 Additional file 9: Gene ontology (GO) enrichment results for pairwise tissue comparisons

908 Additional file 10: Antibody table and peptide absorption results

909

910

911 **Additional file 10.**

912 **Additional table 1.** Table of antisera used in immunohistochemical surveys of the catshark pancreas.

913 PYY, peptide YY; NPY, neuropeptide Y: PP, pancreatic polypeptide.

Antigen	Species	Dilution	Source
PYY	Rabbit	1:500	Sigma
PYY	Rabbit	1:500	AbCam, UK
NPY	Rabbit	1:1000	AbCam,UK
Insulin	Mouse	1:75	Novo Nordisk A/S, Denmark
Insulin	Guinea Pig	1:200	DAKO, Glostrup, Denmark
Glucagon	Mouse	1:50	Novo Nordisk A/S, Denmark
Glucagon	Rabbit	1:500	DAKO, Glostrup, Denmark
PP	Rabbit	1:500	Sigma
PP	Guinea Pig	1:500	Linco/Millipore,,
Somatostain	Mouse	1:100	Novo Nordisk A/S, Denmark

914

915

916

917

918

919

920

921

922 **Additional table 2.** Absorption of Pancreatic Polypeptide Family antisera. Staining is characterised from
923 strong (+++) to weak (+) or absent (-). PYY, peptide YY; NPY, neuropeptide Y: PP, pancreatic
924 polypeptide.

Antibody	Peptide	Staining Dogfish	Staining Mouse
PP	None	+++	-
PP	NPY	-	-
PP	PYY	-	-
PP	PP	-	-
PYY abcam	none	++	++
PYY abcam	PP	++	++
PYY abcam	NPY	+	++
PPY abcam	PYY	-	-
PYY Sigma	none	+++	+
PYY Sigma	PP	+++	+
PYY Sigma	NPY	+++	+
PYY sigma	PYY	-	-
NPY	none	+++	+++
NPY	PP	+++	+++
NPY	PYY	-	-
NPY	NPY	-	-

925