Identification of differentially expressed genes in the heart precursor cells of the chick embryo

Margaret Bento[1,2,3], Elizabeth Correia[1,2], Ana T. Tavares[3], Jörg D. Becker[3] and José A. Belo[1,2,3,§]

1 Regenerative Medicine Program, Departamento de Ciências Biomédicas e Medicina, Universidade do Algarve, Portugal
2 IBB-Institute for Biotechnology and Bioengineering, Centro de Biomedicina Molecular e Estrutural, Universidade do Algarve, Campus de Gambelas, Faro, Portugal
3 Instituto Gulbenkian de Ciência, 2781-901 Oeiras, Portugal

§Corresponding author
Email addresses:
MB: mbento@ualg.pt
EC: eccorreia@ualg.pt
AT: atavares@igc.gulbenkian.pt
JDB: jbecker@igc.gulbenkian.pt
JAB: jbelo@ualg.pt

ABSTRACT

Genetic evidence has implicated several genes as being critical for heart development. However, the inducers of these genes as well as their targets and pathways they are involved with, remain largely unknown. Previous studies in the avian embryo showed that at HH4 *Cerberus* (*cCer*) transcripts are detected in the anterior endomesoderm including the heart precursor cells and later in the left lateral plate mesoderm.
We have identified a promoter element of chick *cCer* able to drive EGFP expression in a population of cells that consistently exit from the anterior primitive streak region, from as early as stage HH3+, and that later will populate the heart. Using this promoter element as a tool allowed us to identify novel genes previously not known to potentially play a role in heart development.
In order to identify and study genes expressed and involved in the correct development and differentiation of the vertebrate Heart Precursor Cell (HPC) lineages, a differential

screening using Affymetrix GeneChip system technologies was performed. Remarkably, this screening led to the identification of more than 700 transcripts differentially expressed in the Heart Forming Regions (HFR). Bioinformatic tools allowed us to filter the large amount of data generated from this approach and to select a few transcripts for *in vivo* validation.
Whole-mount *in situ* hybridization and sectioning of selected genes showed heart and vascular expression patterns for these transcripts during early chick development.
We have developed an effective strategy to specifically identify genes that are differentially expressed in the HPC lineages. Within this set we have identified several genes that are expressed in the heart, blood and vascular lineages, which are likely to play a role in their development. These genes are potential candidates for future functional studies on early embryonic patterning.

1. Introduction

The genetic control of the heart development is a very complex and still poorly understood process. During early development, the heart of chick and humans has similar morphologies (De la Cruz et al., 1989) and most cardiovascular malformations found in the avian embryo are analogous to those found in humans (Nishibatake et al., 1987).
In the chick embryo, at developmental stage HH3, cardiac progenitor cells have been mapped to the anterior two thirds of the primitive streak (PS), excluding Hensen's node (HN; Garcia-Martinez and Schoenwolf, 1993). After ingression, these cells migrate anteriorly and laterally in a semicircular trajectory with slight or no mixture of cells and at stages HH4-5 cardiac progenitors are located on either side of HN in the anterior mesoderm (Yang et al., 2002; Yue et al., 2008; Cui et al., 2009). At stage HH6, when the lateral plate mesoderm dissociates into the splanchnic and somatic mesoderm layers, myocardial and endocardial progenitors are constrained to the splanchnic mesoderm and constitute the heart fields. These bilateral heart fields fold ventrally and, in the midline, give rise to the linear heart tube between HH9 and HH11 (Stalsberg and DeHaan, 1969; Linask, 1992; DeRuiter et al., 1992; Abu-Issa and Kirby, 2007). By this stage the heart tube undergoes progressive rightward looping, giving rise to a four chambered heart (De la Cruz et al., 1989).

Studies to identify the heart forming regions (HFR) in the avian embryo have been carried out since the early 1900s, and over the years several fate map studies have been undertaken. These fate map studies have established accurate boundaries for the HFR during early development (Redkar et al., 2001; Hochgreb et al., 2003). Among these studies there is a dynamic positional fate map, undertaken by Cui and colleagues that allowed to trace precardiac cells from the moment they exit the PS until the formation of the tubular heart (HH3+ to HH10; Cui et al., 2009). Early heart formation studies have proven this process to be a very complex one. Genes such as *BMP-2* and *Nkx2.5*, although having a functional role in cardiac development (Ehrman and Yutzey, 1999), were found not to be able to demarcate the entire primary heart-forming region during early developmental stages (Jiang et al., 1998; Kostetskii et al., 1999; Redkar et al., 2001; Eisenberg, 2002; Hochgreb et al., 2003; Cui et al., 2009). Moreover, several distinct cell populations were proven to contribute to the formation of the adult heart (Martinsen, 2005) including a primary and a secondary heart field, identified by the expression of specific markers (Kelly et al., 2001; Waldo et al., 2001; Mjaatvedt et al., 2001).

Cerberus-like genes encode for cysteine-knot secreted proteins that behave as antagonists of members of the transforming growth factor-β (TGF-β) family during development (Hsu et al., 1998; Piccolo et al., 1999; Rodriguez-Esteban et al., 1999; Belo et al., 2000; Belo et al., 2009). The *Xenopus Cerberus* (*Xcer*) gene was first identified as a Spemann organizer associated transcript that encodes a secreted protein capable of inducing ectopic heads and duplicated hearts and livers, when injected into Xenopus eggs (Bouwmeester, 1996). Also, both Cerberus and its mouse homologue *Cer1* induce upregulation of *Nkx2.5* (Bouwmeester et al., 1996; Belo et al., 1997), and it has been demonstrated that the *XCer* expressing endoderm is required for heart inducing activity (Schultheiss et al., 1995; Schneider and Mercola, 1999; Marvin et al., 2001). *Xenopus Cer*, mouse *cer1* and *cCer* transcripts are detected in equivalent embryonic structures, the anterior endomesoderm, anterior visceral endoderm, and hypoblast, respectively (Bouwmeester et al., 1996; Belo et al., 1997; Rodriguez-Esteban et al., 1999; Liguori et al., 2008). Additionally, in chick and mouse embryos, the expression of these genes is also present in the anterior definitive mesendoderm (Rodriguez-Esteban et al., 1999; Belo et al., 1997).

In previous work aiming to study the transcriptional regulation of *cCer*, we set up a reporter assay for heart/hemangioblast cell populations in the chick embryo. The study uncovered a 2.5 kb fragment upstream of the ATG of chick Cer that was able to drive the expression of EGFP into cell populations that express *cCer* such as the anterior mesendoderm, lateral plate mesoderm, heart and anterior blood islands or hemangioblasts (Tavares et al., 2007; data not shown). Aiming to identify uncharacterized genes that may play a role in heart development we used the 2.5Cer EGFP reporter to isolate an early cardiac endomesodermal population. This cell population was selected as representative of the HFR in a differential screening that resulted in more than 700 differentially expressed transcripts. The outcome of the screening is a genetic profile of the HFR at stage HH4, which provides a hint for cardiogenic precursor lineage identity. Moreover, we describe the expression patterns for four previously uncharacterized genes, and show that *CHEP18*, *CHEP33*, *CHEP40* and *cAdtk1* are differentially expressed in the cardiogenic region in early chick embryos.

2. Results

2.1. Cer2.5-EGFP reporter construct is expressed in early heart precursor cells (HPC)

In previous work, we have reported that the cCer2.5-EGFP construct, a 2.5 kb DNA fragment upstream of the ATG of *cCer* subcloned into an EGFP reporter vector, was functional and able to drive EGFP expression into the anterior mesendoderm and lateral plate mesoderm (Tavares et al., 2007). In order to further characterize cCer2.5-EGFP expression, embryos at HH3 were electroporated with the cCer2.5-EGFP reporter construct and cultured (New, 1955). The pCCAGS-AFP promoter was used as control.

EGFP fluorescence could be observed in the anterior mesendoderm at early stages HH4-6 (Fig. 1A, A', B and C), in the prospective extraembryonic mesoderm and heart precursor cells at HH7 (Fig. 1D, D'), in the extraembyonic mesoderm and precardiac region at HH9 (Fig. 1F, F'), and, later on, populating the heart and the anterior blood islands (Fig 1G, G', G''). This EGFP expression is consistent with chick *Cerberus*

expression at early stage HH4, as visualized by whole-mount *in situ* hybridization (Fig. 1H). The specificity of these results was confirmed upon comparison with the ubiquitous fluorescence observed in the pCCAGs-AFP electroporated control embryo (Figure 1E).

2.2. Isolation of heart precursor cells (HPC) and RNA collection

To identify transcripts differentially expressed during early cardiac development in the cCer2.5-EGFP positive population, we microdissected this region at stage HH4, and dissociated the tissue into a single cell suspension. The EGFP positive cells from the HFR pool were sorted using FACS. The negative control population was derived from the posterior part of the primitive streak (PPPS), a non cardiac region, to minimize the risk of losing unknown cardiac genes that are expressed in the mesendoderm tissues. The subtractive hybridization scheme is outlined in Figure 2A.

Prior to sorting, we evaluated whether the cells from the targeted areas were typical embryonic chick cells (Fig. 2B). Indeed, the R3 region contains cells with a level of granularity similar to that seen in the control embryonic chick cells (data not shown). This approach yielded successful separation of the two different populations (EGFP negative versus EGFP positive cells; Fig. 2C). In the R8 gated region, approximately 6% of the total number of cells was EGFP-positive. To assess purity and viability of the HPC a preliminary sorting process of approximately 100 cells was conducted. This initial FACS analysis of the sorted EGFP-labeled cells indicated a cell viability and purity of 99% (data not shown), confirming that no EGFP negative cells had been collected and that the mechanical stress exerted on the cells during FACS had no effect on their viability. The complete cell suspension was then sorted resulting in the isolation of approximately 4000 EGFP positive cells. These cells were collected directly for RNA extraction. Despite the low number of cells a high quality RNA sample was obtained. Due to the limiting amount of total RNA extracted from the isolated cells, a two round linear amplification method was employed to obtain the required amount of fragmented biotinylated cRNA for hybridization to the GeneChip microarray.
The cRNA pool obtained from the first cycle of amplification was analyzed for the relative presence/absence of known markers in order to test the feasibility of the

differential screening (Fig. 2D). We used *cCer* as a marker of early endomesoderm, *cNodal* as a PS marker and *cNot1* as a HN marker. The EGFP positive population only expressed *cCer*. The PPPS population was positive for *cNodal* and *cCer*. *cNot1* expression was absent in both EGFP positive and PPPS populations. The whole embryo was positive for all the markers. All cell populations expressed *GAPDH* that was used as a reference gene.

2.3. Identification and functional network analysis of differentially expressed genes in heart precursor cells

The Affimetrix GeneChip® Chicken Genome array was used to compare the cRNA pools of EGFP positive and PPPS cell populations. The resultant data were classified and organized, using DAVID and IPA tools.

The differential screening resulted in the detection of 777 genes that are differentially expressed with a fold change greater than 2. The microarray data used in this analysis has been submitted to NCBI GEO database under accession number GSE23389 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE23389).
Gene ontology analysis was carried out on significantly regulated transcripts using DAVID (http://david.abcc.ncifcrf.gov/). The transcripts were organized (Figure 3A) into an upregulated population that contains a total of 437genes, which includes 301 uncharacterized (119 unknown genes and 182 annotated genes), and 136 characterized (of which 38 have GeneRIF). This gene expression screening also revealed 340 downregulated genes in the HPC enriched cell population, of these 205 were uncharacterized (80 unknown genes and 125 annotated genes), and 135 characterized (of which 44 have GeneRIF).
The detailed analysis of the obtained list of characterized upregulated genes in the EGFP positive population showed that several were previously associated with HPC populations, such as *Fgf8*, *TNNT2*, *NRP1*, *Isl-1* and *Wnt11* (Fig. 3B).
We used Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) algorithms and database to dynamically generate significant biological networks from the gene expression data sets. From the analysis of the pool of upregulated characterized genes resulted a list of associated network functions, being Cellular

Movement, Cardiovascular System Development and Function, Cellular Development the network with highest score (45; Table1).

2.4. Validation of the differentially expressed candidate genes in the heart precursor cells

From the list of upregulated uncharacterized genes, we selected a number of genes for expression pattern characterization. The selection was based on their fold change and predicted function.
We named them *CHEP33* (which stands for Chick Heart Progenitor #33), *CHEP36*, *CHEP40* and *CHEP18*. *CHEP36* was also found in a previously screening for Anterior Visceral Endoderm genes performed in our lab, and denominated as ADTK1 (Anterior Distal Tyrosine Kinase 1; Gonçalves et al, 2011). For that reason this gene will be renamed as cAdtk1 in this manuscript.

The *cAdtk1* gene (GenBank accession no. CR405754.1) is 6,8 fold upregulated in the HFR when compared with the posterior part of the Primitive streak. Whole-mount *in situ* hybridization of *cAdtk1* (Fig. 4A-H) showed that it is expressed as early as stage HH4 in the anterior mesendoderm (Fig. 4A) with an expression pattern similar to that of c*Cer* (Fig. 1H). At stage HH5-6 expression is restricted to a mesodermal area on either side of the embryo overlapping the heart fields and in a more medial ectodermal region on both sides of the head process (Fig. 4B, C and C'). At stage HH7-8 expression is seen in both in the lateral plate mesoderm and in the proamnion (extra-embryonic ectoderm and extra-embryonic endoderm) (Fig. 4D, D', D'' and E). By stage HH9, *cAdtk1* seems to be expressed in the wall of sinus venosus (Fig. 4F and F'). At stage HH10, expression appears to be restricted to the heart tube region and and lateral plate mesoderm (Fig. 4G). At later stages by HH14 expression persists in the lateral plate mesoderm and amnion, and a new expression domain can be seen in the tail bud mesoderm (Fig. 4H and data not shown).

CHEP33 (GenBank accession no. XM_421599) was obtained in the differential screening with a +4,4 fold change, and encodes for a putative protein of 239 amino acids. The predicted amino acid sequence of chick CHEP33 is identical 90,8% and

90,5% to its human and mouse orthologs respectively and contains a predicted Zinc-finger motif. Expression of *CHEP33* was initially detected at stage HH4 by whole-mount *in situ* hybridization in bilateral prospective heart fields on the left and right sides of the embryo (Fig. 4I). At stage HH5 *CHEP33* transcripts are detected in a large mesodermal area (Fig. 4J, J') and by stage HH7-8 expression can be observed in the lateral plate mesoderm where heart precursors lay, and in the forming somites (Fig. 4K). By stage HH9, expression is visible in the forming heart tube and in the somites (Fig. 4L). At stage HH11-12, the heart is intensely labeled, as well as the somites and staining in the head and lateral plate mesoderm becomes evident (Fig. 4M-N). Histological sections shown that *CHEP33* is present specifically in the myocardium, pharyngeal endoderm, neural tube, derma-myotome and somatic lateral plate mesoderm. Later in development *CHEP33* is strongly expressed in the heart (Fig. 4O).

CHEP40 gene (GenBank accession no. XM_418631.2) was detected in the screening with a +5,0 fold change, and encodes for a putative protein of 893 a.a. The predicted amino acid sequence of chick *CHEP40* has several homologues in other species, and contains a putative transmembrane domain in the medial portion of the protein. It also contains conserved domains predicted to be involved in cell adhesion, like the desintegrin and Protocadherin domains. Whole-mount *in situ* hybridization showed that by stage HH4+ *CHEP40* transcripts are detected in the HFR mesoderm (Fig. 5A, A' and A''). However, they are also present in the three germ layers of the *area pellucida* of the embryo, but absent in the primitive streak. At this stage of development, transcripts of CHEP40 are also detected in the posterior part of the *area opaca*. At stage HH6, *CHEP40* is highly expressed in the *area pellucida* except for the head process, Hensen's node and primitive streak. Some staining is also observed on the extraembryonic part of the embryo (Fig. 5B). Later in development, at HH8 *CHEP40* transcripts are present in the heart primordium (Fig. 5C) and at stages HH10 to 14 *CHEP40* transcripts were detected in the heart endocardium, intraembryonic vasculature, anterior intestinal portal and neural fold tip (Fig. 5D-F and D'-E'').

CHEP18 (GenBank accession no. XM_416761) was found to be upregulated with a +6,8 fold change, and encodes for a hypothetical protein of 388 a.a. with potential orthologs in several organisms. *CHEP18* belongs to the VHS ENTH-ANTH superfamily, more specifically to VHS_Tom1, VHS domain family, Tom1 subfamily.

Whole-mount *in situ* hybridization showed that at stage HH4 (Fig. 5G) the expression of *CHEP18* is limited to the extraembryonic mesoderm of the posterior part of the embryo, whilst at stage HH5 (Fig. 5H and H') is possible to observe staining on either side of the embryo in the mesodermal cells that border the *area pellucida* known to belong to the HFR, as well as in two mesodermal patches of labeled cells on both sides of the Hensen's node (black arrowheads). From stage HH6 to 8 (Fig. 5I-K and J'-K'') the expression remains in the extraembryonic mesoderm as well as on both sides of the Hensen's node, however, the intraembryonic mesodermal expression appears to progress medially. Later in development, at HH10 (Fig. 5L, L' and L'') *CHEP18* transcripts are present in the extraembryonic vasculature as well as in the intraembryonic vasculature and in the heart, more specifically in the endocardium. Finally at stage HH15, the expression of *CHEP18* persists in the heart as well as in the vasculature including the posterior cardinal veins and head microvessels (Fig. 5M and M').

3. Discussion

The heart is the first organ to become functional in the vertebrate embryo. It develops from two bilateral heart fields located in the anterior lateral mesoderm of the early embryo, which are formed during gastrulation. A vast array of genes and signaling pathways have been reported to be crucial in all steps of cardiogenesis, from migration of primitive-streak progenitor cells, formation of primary and secondary heart fields, fusion of cardiac crescent cells, cardiac looping, to later morphogenetic mechanisms. Developmental signaling pathways are required to act in combination with tissue-specific transcriptional cofactors to elicit inductive responses. The spatial and temporal orchestration of these processes implies a complex program of genetic control, exerted in large part through precisely controlled processes of cell-cell signaling and regulators of gene expression.

We generated a cCer2.5-EGFP construct that effectively produces specific expression in cells leaving the PS at HH3+ and later populate the heart tube.
To confirm the identity of the isolated cell population, a RT-PCR analysis was performed. Known cardiac genes such as *NKX2.5* and *GATA4* were not used in this

analysis because they are only expressed later, by stage HH5 (Abu-Issa and Kirby, 2007; Warkman et al., 2008). For that reason we used *cCer* as a marker of early endomesoderm, *cNodal* as a PS marker, and *cNot1* as an HN marker. Our results were in agreement with the expected since the selected EGFP+ cell population expresses *cCer* and doesn't express *cNodal*, suggesting that we have isolated an enriched population of endomesoderm containing heart precursor cells. The presence of *cCer* transcripts in the PPPS region can be explained since at stage HH4 *cCer* expressing cells are leaving the primitive streak toward the cardiogenic mesoderm, and the dissected posterior region contains some of these cells (Fig. 1H).

A successful combination of methods was employed, resulting in the isolation of 777 independent transcripts potentially implicated in early heart development at stage HH4. These results were catalogued using DAVID (Figure 3A). In our study, reference genes as *Actin* and *GAPDH* showed no fold change, since they are ubiquitously expressed in all eukaryotic cells (Fig. 3B). This also indicates that the EGFP positive population and the negative control have a similar number of cells. We defined as upregulated genes, those found differentially expressed and increased in the EGFP-positive population when compared with the control population. From the pool of upregulated genes, several known cardiac genes were found to be present, some of which with great relevance during early heart development. Both *NRP1* and *Wnt11* genes were strongly expressed in the EGFP-positive cell population and have been previously implicated in HPC mechanisms. NRP1 has been described to be a pro-angiogenic receptor (Ball et al., 2010), and that in order to promote cardiomyogenesis, the Wnt canonical pathway must be blocked by Wnt inhibitors, such as Wnt11 (Marvin et al., 2001; Maye et al., 2004). Multipotent cardiovascular progenitor cells, which express *Isl1* together with *Nkx2.5* and *Flk-1,* give rise to cardiomyocytes, smooth muscle cells, and endothelial cells, representing all three major cell lineages of the heart (Moretti et al., 2006). *Flk-1* and *Isl1* expressions were found to be upregulated in the EGFP positive population. Another gene found to be enriched in this cell population was *TNNT2*, which is known to be present in the embryonic heart of the chick and is responsible for binding tropomyosin to regulate calcium binding and contractility of heart muscle (Druyan et al., 2007). Fgf8 is one of the earliest genes known to regulate myocardial progenitor cells, that when reduced in expression leads to abnormal heart development. Though WISH studies at HH4 show its expression occurs in the cranial part of the PS (Dathe et al., 2002), we also found it to be enriched in the HPC when compared to our control cell population.

Therefore, the interpretation of some results might not be immediate, as several differentially expressed genes are known to be involved in vasculogenic and angiogenic processes (*Flk1*, *VEGF-D* and *VE-cadherin*). Additionally, multiple genes associated with hemangioblast fate, including *Flk-1* and *Scl* were detected. This led us to suppose that along with the cardiac precursor cells, hemangioblasts were also isolated. The hemangioblast was proposed a long time ago to be the common precursor for endothelial and hematopoietic cells, and the hemangioblast cells of the lateral plate mesoderm can give rise to both the angioblasts of the vascular system and the pluripotential hematopoietic stem cells of the blood and lymphoid systems. The primitive endothelial and hematopoietic lineages are believed to derive from mesodermal cells that have migrated from the PS to form the blood islands in the extraembryonic yolk sac, being Flk-1 and Scl two critical molecules involved in this process (Murray, 1932; Ema et al., 2003). It has also been reported that the region in the mesoderm posteriorly adjacent to the cardiogenic region is responsible for the emergence of the blood precursors. We assume that the presence of both heart and hemangioblast precursor cells in our differential screening can be explained by the use of 2.5cCer as a reporter. This transgene also drives the expression of EGFP in the anteriormost hemangioblast population (Fig. 1F and G), which is contiguous to the cardiac precursor cells region. However the isolation of both cell precursors doesn't invalidate our aim to study cardiogenesis. In a study performed in zebrafish, a close genetic link between anterior hemangioblast and cardiac precursors was identified which supports the notion that the anterior hemangioblast population is an evolutionary ancestor of the second heart field (Peterkin et al., 2009). Moreover, cardiomyocytes have been obtained from ES cell derivatives expressing the hemangioblast marker, VEGF receptor, *Flk-1*, which is consistent with the idea of common genetic programme (Kattman et al., 2006). It is known that the cardiogenic process at such early stages needs BMP signaling from the anterior endoderm (Schultheiss et al., 1997), to induce a field with cardiogenic competence, and it is also known that in this region Wnt inhibition is necessary to promote cardiomyogenesis (Schneider and Mercola, 2001; Marvin et al., 2001). As we mentioned before, *Wnt11*, an inhibitor of the Wnt canonical pathway (Maye et al., 2004), was found to be more expressed in sorted cells, which was to be expected since it should be present to promote cardiomyogenesis. However, *Bmp2* and *Bmp4* were not found to be enriched in the HPC. This is probably due to the fact

that the region selected as negative control also expresses these genes at a similar expression level (Marvin et al., 2001; Brand, 2003).
IPA tool was used to evaluate if the characterized genes found enriched in the EGFP positive population, were together due to random change. A list of associated network functions was produced being *Cellular Movement, Cardiovascular System Development and Function, Cellular Development* the one with highest score (45; Table1), meaning that there is a 1 in 2250 chance that the Focus Genes are together due to random chance. The enrichment of cardiac specific genes in the EGFP population, together with the functional *in silico* characterization of the differentially regulated transcripts, strengthens the possibility that heart precursor cell isolation was accomplished.

Following categorization of the genes differentially expressed in the HPC at stage HH4, we focused our study on the 301 upregulated uncharacterized transcripts. Aware of the possible existence of false positives among these novel genes due to the absence of replicates for our arrays, we performed whole-mount *in situ* hybridizations to determine if they fitted HPC expression. Four genes were chosen for further characterization: *CHEP33* a novel gene with a putative zinc finger binding domain; *cAdtk1* containing a predicted dual-specificity Ser/Thr/Tyr kinase catalytic domain; *CHEP40* a novel gene possibly involved in cell adhesion and *CHEP18* that belongs to the VHS ENTH-ANTH superfamily. The whole-mount *in situ* hybridizations performed in this study showed that the expression patterns of the novel genes overlap with the HPC region. However, expression of *CHEP18* at stage HH4 is absent in the HFR. This could be due to a weak expression level that would not be detected by whole-mount *in situ* hybridization. Or, as there is a narrow window of time between stages HH4 and HH5, it is possible that some embryos collected for the screening could have been at an early HH5 stage, and already expressing CHEP18 in the HFR. As for the expression of *CHEP40* at stage HH4, expression can be observed almost all over the *area pellucida*, being however weaker in the posterior part of the embryo than in the HFR. At first this fact seems to invalidate the experimental approach used as the basis of this study, however by analyzing the expression pattern carefully, it can be seen that both the posterior part of the embryo and the primitive streak are less stained than the HFR. As this screening is a differential one, one cell population versus another, the plausible explanation of these results is that CHEP40 is expressed 5 times more in the HFR mesoderm than in the posterior part of the primitive streak. The posterior part of the *area opaca* at HH4 is usually associated to

vascular cell populations, which can explain CHEP40 expression at HH10 in the intraembryonic vessels. At early cardiac formation stages expression was found in the heart for *CHEP33*, *CHEP18* and *CHEP40*. *cAdtk1* is present in HFR during pre-heart tube stages, being as well present in structures adjacent to the heart after its formation. This gene, as previously described in mouse embryos, is also absent from the heart tube at early cardiac stages (Imuta et al., 2009). Nevertheless it has been described to be strongly expressed in the heart of adult mice. In the mouse, at E6.5 *Pkdcc* (*cAdtk1* ortholog) is expressed in the anterior viseral endoderm and anterior PS, whereas in the chick at HH4 there is no expression in the PS. In the chick at HH5 *cAdtk1* is expressed in the HFR mesoderm and in the neural plate ectoderm, whereas no ectodermal expression is seen in the mouse at E7.5. WISH studies of the four selected clones allowed us to corroborate the hypothesis that the *cCer* positive population expresses cardiogenic genes. This was proven by the detection in the heart, later in development, of transcripts of these studied genes. Furthermore, the expression of the new clones in the mesendoderm at stage HH4, validates the differential screening which involved the RNA isolated from these cells at the same stage.

In conclusion, we have identified a large set of novel genes that are differentially expressed in the H/HPC precursor lineages. By developing a procedure to isolate the heart precursor cells using the cer2.5-EGFP construct, we were able to specifically separate a population of H/HPC expressing already known cardiac markers, and a long list of still uncharacterized genes. From those, we demonstrated by WISH that a number of these *in silico*-identified transcripts are in fact expressed in these cell populations and potentially required for heart formation. More importantly, our study unveiled a large number of uncharacterized genes that can now be used by the scientific community for further developmental studies.

4. Methods

4.1. DNA constructs and morpholinos

The pCAGGS-AFP vector (Momose et al, 1999), carrying the cDNA of the green fluorescence protein under the control of the CAGGS promoter, was used to control the extent and efficiency of electroporation.

The pCer2.5-EGFP vector contains a 2.5Kb fragment upstream the ATG from the Cerberus regulatory region.

4.2. Chick embryo isolation, culture and electroporation method

Fertilized chicken eggs (Sociedade Agrícola Quinta da Freiria, SA, Torres Vedras, Portugal) were incubated until the desired stage was reached according to Hamburger and Hamilton (1951) at 37,5°C. Embryos were explanted at HH stage 3+ together with the vitelline membrane and anchored to a metacrilate ring following the protocol of New (1955).
Embryos were microinjected and electroporated as described previously (Tavares et al., 2007) at HH stage 3 with a DNA solution (0.5-3 mg/ml; 0.1% Fast Green; Sigma). The embryos were incubated at 37 °C for the appropriate period of time (7-48h), and observed under a fluorescence stereomicroscope (Leica MZ16FA) and further dissected.

4.3. Excision of the GFP expressing HFRs and FACS analysis

The HFR reside in bilateral regions on either side of Hensen's node of a stage HH3+-4 embryo. The chick embryos (*n*=6, *n* refers to the number of embryos successfully electroporated) were collected at stage HH4 and two different pools were made: one comprising the region containing the EGFP-Positive cells, and another the posterior part of the primitive streak (PPPS). Both regions were removed and washed in cold PBS. The PPPS proceeded directly for RNA extraction.

The tissues containing the EGFP positive cells were subjected to 5 minutes digestion in 0.25% trypsin/PBS at RT and triturated 10–20 times with a 1000 µL pipette. PBS containing 10% FBS (Hyclone) was added followed by centrifugation. Cells were ressuspended in PBS/2%FBS, filtered through a sterile stainless nylon mesh and proceeded for FACS analysis.

Fluorescence-activated cell sorting was performed in a Moflo High-Speed Cell Sorter (Dako-Cytomation), using a 100-μm ceramic nozzle with 30 psi sheath pressure, a 488-nm laser line from a Coherent Sapphire 488 to 200 CDRH laser for GFP excitation. GFP was detected using a 530-/40-nm band pass filter.
An initial gate (R3, Fig.2A) is set on a FSC (Forward scatter) *vs.* SSC (Side scatter) dot plot to contain the cell population based on their cell size and intracellular complexity,

similar to that seen in the control embryonic chick cells performed before (data not shown).

The events in gate R3 are then displayed on a second dot plot (Fig. 2C) and a second gate is set to contain the GFP positive population. FL1 *vs*. FL2 are the intensity of light detected in green (FL1, GFP) and yellow light channels (FL2). Compensation was applied.

Cells in the gated region R8 were sorted with a sorting rate of ~1 cell per second and 100 cells were initially collected and then tested. FACS analysis (Becton Dickinson FacsCalibur cytometer) of these post-sorting GFP-labeled cells indicated a cell viability and purity of 99% (data not shown). The GFP positive cells were sorted directly to an Eppendorf tube containing 1ml of TRizol Reagent (Invitrogen).

4.4. RNA isolation, target synthesis and hybridization to Affymetrix GeneChips

Total RNA was extracted from cells by using TRIzol Reagent (Invitrogen) according to the manufacturer's instructions with the following modification; cells were lysed using 1mL of TRIzol. The cell lysate was ressuspended with a pipette several times, vortexed for 30s and incubated at room temperature for 5min. Five µg of linear acrylamide (Ambion, Inc.) were added to the samples and transferred to a Phase Lock Gel (PLG) Heavy 2ml tube (Eppendorf AG); 0.2vol. of chloroform:isoamyl acid (25:1) was added to the sample, and mixed vigorously for 15 s. Samples were incubated for 2-3 minutes at RT and centrifuged at 12000 ×g for 15min. at 4ºC. The upper aqueous phase was transferred to a new PLG Heavy tube and a second extraction with 0.2 vol. chloroform:isoamyl acid (25:1) was performed. The aqueous phase was transferred to a nuclease-free tube and the total RNA was precipitated O/N at -20ºC with 0.8 vol. of 2-propanol. Total RNA was collected by centrifugation at 14000 ×g for 40 min. at 4ºC, and the pellet was washed twice with 1ml of 80% ethanol at -20ºC. The pellet was air dried and ressuspended in 6µl of nuclease-free water (Promega Corp.), of which 1µl was used to determine RNA concentration and purity in a NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies), being integrity confirmed using an Agilent 2100 Bioanalyzer with a RNA 6000 Nano/ pico Assay (Agilent Technologies, Palo Alto, CA).

RNA was processed for use on Affymetrix (Santa Clara, CA, USA) GeneChip Chicken Genome Arrays, according to the manufacturer's Two-Cycle Target Labeling Assay. Briefly, 45 ng of total RNA containing spiked in Poly-A RNA controls (GeneChip Expression GeneChip Eukaryotic Poly-A RNA Control Kit; Affymetrix) was used in a reverse transcription reaction (Two-Cycle DNA synthesis kit; Affymetrix) to generate first-strand cDNA. After second-strand synthesis, double-stranded cDNA was used in an in vitro transcription (IVT) reaction to generate cRNA (MEGAscript T7 kit; Ambion, Austin, TX). The cRNA obtained was used for a second round of cDNA and cRNA synthesis, resulting in biotinylated cRNA (GeneChip Expression 3'-Amplification Reagents for IVT-Labeling; Affymetrix). Size distribution of the cRNA and fragmented cRNA, respectively, was assessed using an Agilent 2100 Bioanalyzer with a RNA 6000 Nano Assay.

15 µg of fragmented cRNA was used in a 300 µl hybridization containing added hybridization controls. 200 µl of mixture was hybridized on arrays for 16 h at 45°C. Standard post hybridization wash and double-stain protocols (EukGE-WS2v5) were used on an Affymetrix GeneChip Fluidics Station 400. Arrays were scanned on an Affymetrix GeneChip scanner 3000. All quality parameters for the arrays were confirmed to be in the recommended range.

4.5. Semi-quantitative RT-PCR

After the first cycle of amplification described above, a semi-quantitative RT-PCR method using primers for known markers was performed to check the representativity of the cRNA pools. The cDNA was prepared according to the RevertAid™ H Minus First Strand cDNA Synthesis Kit protocol (Fermentas) as suggested by the manufacturer. Semi-quantitative RT-PCR was performed (BioRad). Gene expression levels were measured by the ΔΔ Ct method, comparing the EGFP positive population, the PPPS population and whole embryos controls, with GAPDH used as the reference gene. Pre-microarray validation was performed from the same samples used for the microarray experiment. The Semi-Quantitative RT-PCR primers were the follow; GAPDH: FW - 5' ATG CAT CGT GCA CCA CCA AC 3', Rev - 5' CTC CAG ACG GCA GGT CAG GT 3' ; cNot: FW - 5' GCA GCA GTA CAT GGT GGG CAC AGA GC 3', Rev - 5' GCA TCC CCA GTA CCT CCA TCA GGC 3'; cNodal : FW - 5' GCA CCG CCG CAC CAG GAA GGA GAG 3', Rev - 5' GCA CGG CAC CTG GCT GGG

CTT GTA 3', cCer: FW - 5' GAT CTG AGA ACC TGG GCA GC 3', Rev - 5' CAC TGG GAC AGA ACT GGT GC 3'.

4.6. GeneChip data analysis

Scanned arrays were analyzed first with Affymetrix GCOS 1.3 software to obtain Absent/Present calls and for subsequent analysis with dChip 2006 (http://www.dchip.org, Wong Lab, Harvard). Four GeneChip chicken arrays used for another study were added to the two arrays in this study to ensure a robust model for expression value computation: The arrays were normalized to a baseline array with median CEL intensity by applying an Invariant Set Normalization Method (Li and Wong, 2001). Normalized CEL intensities of the six arrays were used to obtain model-based gene expression indices based on a PM (Perfect Match)-only model (Kattman et al., 2006). Only genes called Present in at least one of the two arrays in this study were kept for downstream analysis (17575 genes). All genes compared were considered to be differentially expressed if the 90% lower confidence bound of the fold change between experiment and baseline was above 2. The lower confidence bound criterion means that we can be 90% confident that the fold change is a value between the lower confidence bound and a variable upper confidence bound. It has been shown that the lower confidence bound is a conservative estimate of the fold change, and therefore more reliable as a ranking statistic for changes in gene expression (Li and Wong, 2001).

4.7. Gene list analysis using bioinformatic resources (DAVID and Ingenuity Pathway Analysis)

Genes identified by Affymetrix and with a fold change calculated greater than 2 were used for network and gene ontology analyses.
Gene accession numbers were uploaded into DAVID (http://david.abcc.ncifcrf.gov/) (Huang et al., 2009; Dennis et al., 2003) to filter the redundancy and categorize according to the degree of annotation. The set of up and downregulatedgenes was subdivided into characterized and uncharacterized. The uncharacterized genes include genes that have an ID in Entrez or Ensemble databases (annotated), and genes that do not (unknown). The characterized genes are annotated genes that have been published (PUBMED ID), some of these with functional information (GeneRIF).

Gene accession numbers were imported into the Ingenuity Pathway Analysis version 8.0 (IPA) software (Ingenuity Systems ®, Mountain View, CA, USA), and are the starting point for generating biological networks. IPA then computed a score for each network according to the fit of the user's set of significant genes. A score of 2 indicated a 1 in 100 chance that the focus genes were together in a network because of random chance. Since the size of the created network could potentially be enormous, the IPA software limited the number of molecules in the network to 35, leaving only the most important ones based on the number of connections for each focus gene (focus genes = a subset of uploaded significant genes having direct interactions with other genes in the database) to other significant genes. Networks are scored based on the number of focus molecules in the network, its size, the total number of focus molecules analyzed, and the total number of molecules in the knowledge database that could potentially be included in the networks.

4.8. Clones selection and whole-mount *in situ* hybridization

Chicken EST's for the selected clones were obtained from the BBSRC chick EST database (Hubbard et al., 2004). Clones used were ChEST884h13 for *cAdtk1* (seq. identifier 603865952F1), ChEST514i19 for *CHEP33* (seq. identifier 603542577F1), ChEST739g15 for *CHEP40* (seq. identifier 603784860F1) and ChEST746d11 for *CHEP18* (seq. identifier 603787942F1).
The clones were sequenced using T7 and T3 primers and whole-mount *in situ* hybridization was performed as previously described (Wilknson, 1993). DIG labeled RNA probes were generated by digesting plasmids with Not I, and transcribing with T7 RNA polymerase.

Acknowledgements

We thank J.C. Izpisúa-Belmonte for the cNKx2.5 probe and S. Marques and J.M. Inácio for critical reading of this manuscript. M. Bento and E. Correia are recipients of F.C.T. PhD and BI fellowships, respectively. A.T. Tavares was a recipient of a F.C.T. PD fellowship. This work was supported by research grants from F.C.T., from IGC/FCG and from IBB/CBME, LA to J.A. Belo.

References

Abu-Issa, R., Kirby, M.L., 2007. Heart field: from mesoderm to heart tube. Annu Rev Cell Dev Biol. 23:45–68.

Ball, S.G., Bayley, C., Shuttleworth, C.A., Kielty, C.M., 2010. Neuropilin-1 regulates platelet-derived growth factor receptor signalling in mesenchymal stem cells. Biochem J. 427(1):29-40.

Belo, J.A., Bachiller, D., Agius, E., Kemp, C., Borges, A.C., Marques, S., Piccolo, S., De Robertis, E.M., 2000. Cerberus-like is a secreted BMP and nodal antagonist not essential for mouse development. Genesis. 26(4):265-70.

Belo, J.A., Bouwmeester, T., Leyns, L., Kertesz, N., Gallo, M., Follettie, M., De Robertis, E.M., 1997. Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. Mech Dev. 68:45-57.

Belo, J.A., Silva, A.C., Borges, A.C., Filipe, M., Bento, M., Gonçalves, L., Vitorino, M., Salgueiro, A.M., Teixeira, V., Tavares, A.T., Marques, S., 2009.. Generating asymmetries in the early vertebrate embryo: the role of the Cerberus-like family. Int. J. Dev. Biol. 53:1399-407.

Bouwmeester, T., Kim, S., Sasai, Y., Lu, B., De Robertis, E.M., 1996. Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. Nature 382:595-601.

Brand, T., 2003. Heart development: molecular insights into cardiac specification and early morphogenesis. Dev. Biol. 258:1–19.

Chapman, S.C., Schubert, F.R., Schoenwolf, G.C., Lumsden, A., 2002. Analysis of spatial and temporal gene expression patterns in blastula and gastrula stage chick embryos. Dev Biol 245:187–199.

Cui, C., Cheuvront, T.J., Lansford, R.D., Moreno-Rodriguez, R.A., Schultheiss, T.M., Rongish, B.J. 2009. Dynamic positional fate map of the primary heart-forming region. Dev. Biol. 332(2):212-22.

Dathe, V., Gamel, A., Männer, J., Brand-Saberi, B., Christ, B., 2002. Morphological left-right asymmetry of Hensen's node precedes the asymmetric expression of Shh and Fgf8 in the chick embryo. Anat Embryol 205:343–354.

De la Cruz, M.V., Sánchez-Gómez, C., Palomino, M.A., 1989. The primitive cardiac regions in the straight tube heart (Stage 9-) and their anatomical expression in the mature heart: an experimental study in the chick embryo. J Anat. 165:121–131.

Dennis, G.Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4(5):P3.

DeRuiter, M.C., Poelmann, R.E., VanderPlas-de Vries, I., Mentink, M.M.T., Gittenberger-de Groot, A.C., 1992. The development of the myoeardium and endocardium in mouse embryos Fusion of two heart tubes? Anat. Embryo1. 185:461-473

Druyan, S., Cahaner, A., Ashwell, C.M., 2007. The Expression Patterns of Hypoxia-Inducing Factor Subunit α-1, Heme Oxygenase, Hypoxia Upregulated Protein 1, and Cardiac Troponin T During Development of the Chicken Heart. Poultry Science 86:2384–2389.

Eisenberg, L.M., 2002. Belief vs. scientific observation: the curious story of the precardiac mesoderm. Anat. Rec. 266:194–197.

Ema, M., Faloon, P., Zhang, W.J., Hirashima, M., Reid, T., Stanford, W.L., Orkin, S., Choi, K., Rossant, J., 2003. Combinatorial effects of Flk1 and Tal1 on vascular and hematopoietic development in the mouse. Genes Dev. 17:380-393.

Erhman, L.A., Yutzey, K.E., 1999. Lack of regulation in the heart forming region of avian embryos. Dev. Biol. 207:163-175.

Garcia-Martinez, V., Schoenwolf, G.C., 1993. Primitive-streak origin of the cardiovascular system in avian embryos. Dev. Biol. 159:706-719

Gonçalves, L., Filipe, M., Marques, S., Salgueiro, A.M., Becker, J., Belo, J.A., 2011. Identification and functional analysis of novel genes expressed in the Anterior Visceral Endoderm. Int. J. Dev. Biol. doi: 10.1387/ijdb.103273lg

Hamburger, V., Hamilton, H.L., 1951. A series of normal stages in the development of the chick embryo. Journal of Morphology 88:49-92.

Hochgreb, T., Linhares, V.L., Menezes, D.C., Sampaio, A.C., Yan, C.Y.I., Cardoso, W.V., Rosenthal, N., Xavier-Neto, J., 2003. A caudorostral wave of RALDH2 conveys anteroposterior information to the cardiac field. Development 130, 5363-5374

Hsu, D.R., Economides, A.N., Wang, X., Eimon, P.M., Harland, R.M., 1998. The Xenopus dorsalizing factor Gremlin identifies a novel family of secreted proteins that antagonize BMP activities. Mol. Cell 1:673–683.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 4(1):44-57.

Hubbard, S.J., Grafham, D.V., Beattie, K.J., Overton, I.M., McLaren, S.R., Croning, M.D.R., Boardman, P.E., Bonfield, J.K., Burnside, J., Davies, R.M., Rarrell, E.R., Francis, M.D., Griffiths-Jones, S., Humphray, S.J., Hyland, C., Scott, C.E., Tang, H., Taylor, R.G., Tickle, C., Brown, W.R.A., Birney, E., Rogers, J., Wilson, S.A., 2004. Transcriptome analysis for the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags. Genome Res. 15:174-183.

Imuta, Y., Nishioka, N., Kiyonari, H., Sasaki, H., 2009. Short limbs, cleft palate, and delayed formation of flat proliferative chondrocytes in mice with targeted disruption of a putative protein kinase gene, Pkdcc (AW548124). Dev. Dyn. 238:210–222.

Jiang, Y., Tarzami, S., Burch, J.B., Evans, T., 1998. Common role for each of the cGATA-4/5/6 genes in the regulation of cardiac morphogenesis. Dev. Genetic. 22:263-277

Kattman, S.J., Huber, T.L., Keller, G.M., 2006. Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. Dev.Cell 11:723-732.

Kelly, R.G., Brown, N.A., Buckingham, M.E., 2001. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Dev. Cell 1:435–40.
Kostetskii, I., Jiang, Y., Kostetskaia, E., Yuan, S., Evans, T., Zile, M., 1999. Retinoid Signaling Required for Normal Heart Development Regulates GATA-4 in a Pathway Distinct from Cardiomyocyte Differentiation. Dev. Biol. 206, 206–218

Li, C., Wong, W.H., 2001. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol. 2(8):RESEARCH0032.

Liguori, G.L., Borges, A.C., D'Andrea, D., Liguoro, A., Gonçalves, L., Salgueiro, A.M., Pérsico, M.G., Belo, J.A., 2008. Cripto independent nodal signalling promotes positioning of the A-P axis in the early mouse embryo. Dev Biol. 315:280-9.

Linask, K.K., 1992. N-Cadherin localization in early heart development and polar expression of Na+, K+-ATPase, and integrin during pericardial coelom formation and epithelialization of the differentiating myocardium. Dev Biol. 151:213–224.

Marvin, M.J., Rocco, G.D., Gardiner, A., Bush, S.M., Lassar, A.B., 2001. Inhibition of Wnt activity induces heart formation from posterior mesoderm. Genes Dev. 15: 316-327.

Martinsen, B.J., 2005. Reference Guide to the Stages of Chick Heart Embryology. Dev. Dyn. 233:1217–1237

Mjaatvedt, C.H., Nakaoka, T., Moreno-Rodriguez, R., Norris, R.A., Kern, M.J., Eisenberg, C.A., Turner, D., Markwald, R.R., 2001. The outflow tract of the heart is recruited from a novel heart-forming field. Dev Biol. 238:97–109.

Momose, T., Tonegawa, A., Takeuchi, J., Ogawa, H., Umesono, K., Yasuda, K., 1999. Efficient targeting of gene expression in chick embryos by microelectroporation. Dev Growth Differ 41:335–344.

Moretti, A., Caron, L., Nakano, A., Lam, J.T., Bernshausen, A., Chen, Y., Qyang, Y., Bu, L., Sasaki, M., Martin-Puig, S., 2006. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. Cell 127:1151–1165.

Murray, P.D.F., 1932. The development in vitro of the blood of the early chick embryo. Proc. Roy. Soc. Lond. 11:497-521.

New, D.T., 1955. A new technique for the cultivation of the chick embryo in vitro. J. Embryol. Exp. Morph. 3:326-331.

Nishibatake, M., Kirby, M.L., van Mierop, L.H., 1987. Pathogenesis of persistent truncus arteriosus and dextroposed aorta in the chick embryo after neural crest ablation. Circulation 75:255-264

Peterkin, T., Gibson, A., Patient, R., 2009. Common genetic control of haemangioblast and cardiac development in zebrafish. Development 136:1465-1474.

Piccolo, S., Agius, E., Leyns, L., Bhattacharyya, S., Grunz, H., Bouwmeester, T., Robertis, E.D., 1999. The head inducer Cerberus is a multifunctional antagonis of Nodal, BMP and Wnt signals. Nature 397: 707-710.

Pisano, M.M., Mukhopadhyay, P., Greene, R.M., 2003. Molecular fingerprinting of TGFbeta-treated embryonic maxillary mesenchymal cells. Orthod Craniofac Res. 6(4):194-209.

Qiu, Z., Cang, Y., Goff, S.P., 2009. c-Abl tyrosine kinase regulates cardiac growth and development. PNAS 103: 1136–1141.

Rawles, M.E., 1936. A study in the localization of organ-forming areas in the chick blastoderm of the head-process stage. J. Exp. Zool. 72:271-315.

Redkar, A., Montgomery, M., Litvin, J., 2001. Fate map of early avian cardiac progenitor cells. Development 128: 2269-2279.

Rodriguez-Esteban, M.C., Capdevila, J., Economides, A.N., Pascual, J., Ortiz, A., Izpisúa-Belmonte, J.C., 1999. The novel Cer-like protein Caronte mediates the establishment of embryonic left-right asymmetry. Nature 401:243-251.

Rosenquist, G.C., 1970. Location and movements of cardiogenic cells in the chick embryo: the heart-forming portion of the primitive streak. Dev. Biol. 22: 461-475.

Rudnick, D., 1948. Prospective areas and differentiation potencies in the chick blastoderm. Ann. New York Acad. Sci. 49:761-772.

Schneider, V.A., Mercola, M., 1999. Spatially distinct head and heart inducers within the Xenopus organizer region. Curr. Biol. 9(15):800-9

Schneider, V.A., Mercola, M., 2001. Wnt antagonism initiates cardiogenesis in Xenopus laevis. Genes Dev. 15(3):304-15

Schultheiss, T.M., Xydas, S., Lassar, A.B., 1995. Induction of avian cardiac myogenesis by anterior endoderm. Development 121:4203–4214.

Stalsberg, H., DeHaan, R.L., 1969. The precardiac areas and formation of the tubular heart in chick embryo. Dev Biol. 19:128–159.

Tavares, A.T., Andrade, S., Silva, A.C., Belo, J.A., 2007. Cerberus is a feedback inhibitor of Nodal asymmetric signaling in the chick embryo. Development 134(11):2051-60.

Waldo, K.L., Kumiski, D.H., Wallis, K.T., Stadt, H.A., Hutson, M.R., Platt, D.H., Kirby, M.L., 2001. Conotruncal myocardium arises from a secondary heart field. Development 128:3179 –3188.

Warkman, A.S., Yatskievych, T.A., Hardy, K.M., Krieg, P.A., Antin, P.B., 2008. Myocardin expression during avian embryonic heart development requires the endoderm but is independent of BMP signaling. Dev.Dyn. 237:216–221.

Wilknson, D.G., 1993. Whole-mount in situ hybridrzatton of vertebrate embryos, in: In situ Hybrization A Practical Approach (Wilkinson, D G , ed.), IRL, Oxford, UK, pp. 75–83.

Yang, X., Dormann, D., Munsterberg, A.E., Weijer, C.J., 2002. Cell movement patterns during gastrulation in the chick are controlled by positive and negative chemotaxis mediated by FGF4 and FGF8. Dev. Cell 3, 425-437.

Yue Q., Wagstaff, L., Yang, X., Weijer, C., Munsterberg, A., 2008. Wnt3a-mediated chemorepulsion controls movement patterns of cardiac progenitors and requires RhoA function. Development 135:1029-1037.

Figures

Figure 1 Expression of Cer2.5-EGFP reporter construct in the chick developing embryo. Embryos were targeted at HH3 and electroporated with Cer2.5-EGFP reporter construct by microinjection and electroporation in New culture (A,B,C,D,F,G). In panel E embryos were electroporated with pCAGGS-AFP (control). A', D', F', G', and G'' show sections of embryos A, D, F, and G, respectively. In D', G' and G'' cell nuclei are labelled with DAPI (blue).
(A,A',B,C): EGFP is detected specifically in the lateral anterior mesendoderm. (D,D'): EGFP fluorescence is observed specifically in the future extraembryonic mesoderm and

heart precursor cells (arrow in D'). (F-G): Expression is localised in the extraembryonic mesoderm and cardiogenic region (F,F'), and later in the heart and blood islands (G). In G', fluorescence is seen in the heart (arrow) and in a few cells of the pharynx. In G'' labelled cells are localised both in the endothelium and inside the hematopoietic region of the blood islands (arrows). (H): cCer transcripts were detected by whole-mount in situ hybridization in the anterior mesendoderm.
(ames anterior mesendoderm, hf heart fields, mes mesoderm, hp heart precursors, ht heart tube, pcar pre cardiac region, endo endothelium, ht heart, pha pharynx, bi blood islands)

Figure 2 Strategy for enrichment of transcripts expressed in heart precursor cells. (A): Schematic representation of the experimental procedure used for the differential screening. The characterized 2.5cCer promoter drives expression of EGFP reporter construct in the HPC population (green), but not in other cells (red).
(B-C): Sorting of GFP positive cells from electroporated embryos. The parameters forward scatter (FSC) and green fluorescence intensity (FL1) were used to set sorting gates. FACS analysis of dissociated cells in terms of FSC and SSC (B). Cells falling in the R8 region were sorted and collected (C). The number of cells at each intensity is shown by the number of dots, where each dot represents a single cell. The x-axis, FSC, represents the cell size, whereas the y-axis, SSC, represents cytoplasmic granular intensity. (D): Semi-quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) was performed on equivalent amounts of RNA isolated from whole embryo, PPPS and the sorted EGFP positive cells at HH4 using primers specific for cNot, cNodal, cCer and GAPDH. PCR reactions were terminated after various cycles of amplification (20, 25, 30 and 35) and the reaction products were separated by electrophoresis and visualized using ethidium bromide.

Figure 3 Analysis of the differentially expressed genes in the HPC. (A): Genes identified by Affymetrix with a fold change calculated greater than 2 were uploaded into DAVID (http://david.abcc.ncifcrf.gov/) to filter the redundancy and categorize according to the degree of annotation. This resulted in the identification of 777 genes differentially expressed in the HPC population, which included 437 upregulated and 340 downregulated. (B): Several of the found upregulated genes have been previously associated to heart as well as to hemangioblast precursor cell populations, bars in pink

and red respectively. Reference genes, GAPDH and Actin, were found at 1 fold change (black bars). The horizontal line indicates the significance threshold.
Unknown genes don't have an ID in Entrez or Ensemble databases; **Annotated** genes only have an ID in Entrez or Ensemble databases; **PUBMED ID** annotated genes that have been published; **GeneRIF** published genes with known functional information.

Figure 4 *In vivo* validation of the differential screening (*cADTK* and *CHEP33*). Whole-mount *in situ* hybridization and paraffin sectioning for *cADTK* (A-H) and *CHEP33* (I-O), in the chick embryo: (A) *cADTK* expression is seen in the anterior mesendoderm of the embryo, including the bilateral heart fields; (B, C and C') staining can be observed in the mesodermal HFRs as well as on either side of the neural folds; (D-E) in early somite stages expression is seen in the lateral plate mesoderm and in the proamnion; (F-H) later in development, staining is visible in the heart (wall of sinus venosus) and adjacent endoderm, lateral plate mesoderm and tail bud; (I, J and J') *CHEP33* expression is present in the prospective HFRs mesoderm; (K) in early somite stage, the lateral plate mesoderm and the somites appear stained; (L-N) expression is seen in the heart tube, somites; (M-N) head and lateral plate mesoderm; (N' and N'') expression can be specifically seen in the myocardium, pharyngeal endoderm, neural tube, derma-myotome and somatic lateral plate mesoderm; (O) Later in development, an intense expression is observed in the heart.
(ames anterior mesendoderm, hf heart fields, np neural plate, nfp neural fold patches, mes mesoderm, lpm lateral plate mesoderm, pa pro-amnion, hp heart primordium, ht heart, wsv wall of sinus venosus, tbm tail bud mesoderm, s somites, nt neural tube, phe pharyngeal endoderm, myo myocardium, dmyo derma-myotome, sm somatic lateral plate mesoderm).

Figure 5 *In vivo* validation of the differential screening (*CHEP40* and *CHEP18*). Whole-mount *in situ* hybridization and paraffin sectioning for *CHEP40* (A-F) and *CHEP18* (G-M), in the chick embryo: (A) *CHEP40* is present with greater intensity in the anterior part of the *area pellucida*, except for the primitive streak and Hensen's node, and also in the posterior part of the *area opaca*; (A'and A'') the expression is visible in the three embryonic layers, endoderm, mesoderm and ectoderm; (B) staining is observed heavily in the entire *area pellucida* except for the primitive streak, Hensen's node and head process; (C) at early somite stage *CHEP40* expression is seen in the heart

primordium; (D-F and D'-E'') after heart tube formation expression is seen in the heart endocardium (D', E' and E''), intraembryonic vasculature (D-F and E'), endoderm adjacent to the sinu-atrial region (D') and neural fold tip (D' and E'); (G) CHEP18 expression is present in the posterior and lateral part of the *area opaca*; (H-K and H'-J'') staining can be observed in the mesodermal layer on either side of Hensen's node (black arrowhead); (J-L and J'', K'-L'') intraembryonic and extraembryonic vasculature appear stained; (L and M) expression is seen between the somites (white arrowheads); (M) in the posterior cardinal veins and in the head microvessels; (L, L', M and M') and in the heart.

(PS Primitive Streak; ec ectoderm; mes mesoderm; en endoderm; HN Hensen´s node HP head process; hp heart primordium; ht heart; iev intraembryonic vasculature; nf neural fold; He Heart endocardium; eem extraembryonic mesoderm; eev extraembryonic vasculature; hm head mesenchyme; sm somatic lateral plate mesoderm; sp splanchnic lateral plate mesoderm; mv microvessels; pCV posterior cardinal veins;)

Tables

Table 1. Upregulated Top Network Associated Functions

ID	Molecules in Network	Score	Focus Molecules	Top Functions
1	Cadherin, **CADM1***, **CDH5**, **CDH11**, **CDH20**, **CHEK2**, **COL1A2***, Collagen type I, **FHL2**, **FIGF**, **FLT1***, **GAD1**, Gata, **HHEX**, **ISL1**, **ITGB1BP1**, **KDR***, **LIPG**, **LMO2**, **MPP3**, NFkB (complex), **NRP1**, **PBX3**, **PCSK5***, PLC gamma, **PLK2**, **PLXNA4***, **PTN**, **RBM16**, RNA polymerase II, **TAF1D***, **TAL1**, **TLL1**, Vegf, Vegf Receptor	45	27	Cellular Movement, Cardiovascular System Development and Function, Cellular Development

Genes in boldface were identified by microarray analysis to be expressed differentially more than 2 fold (Focus Genes). Other genes were either not on the expression array or not significantly regulated. (*) duplicated gene in the data set.

Figure 1

Click here to download high resolution image

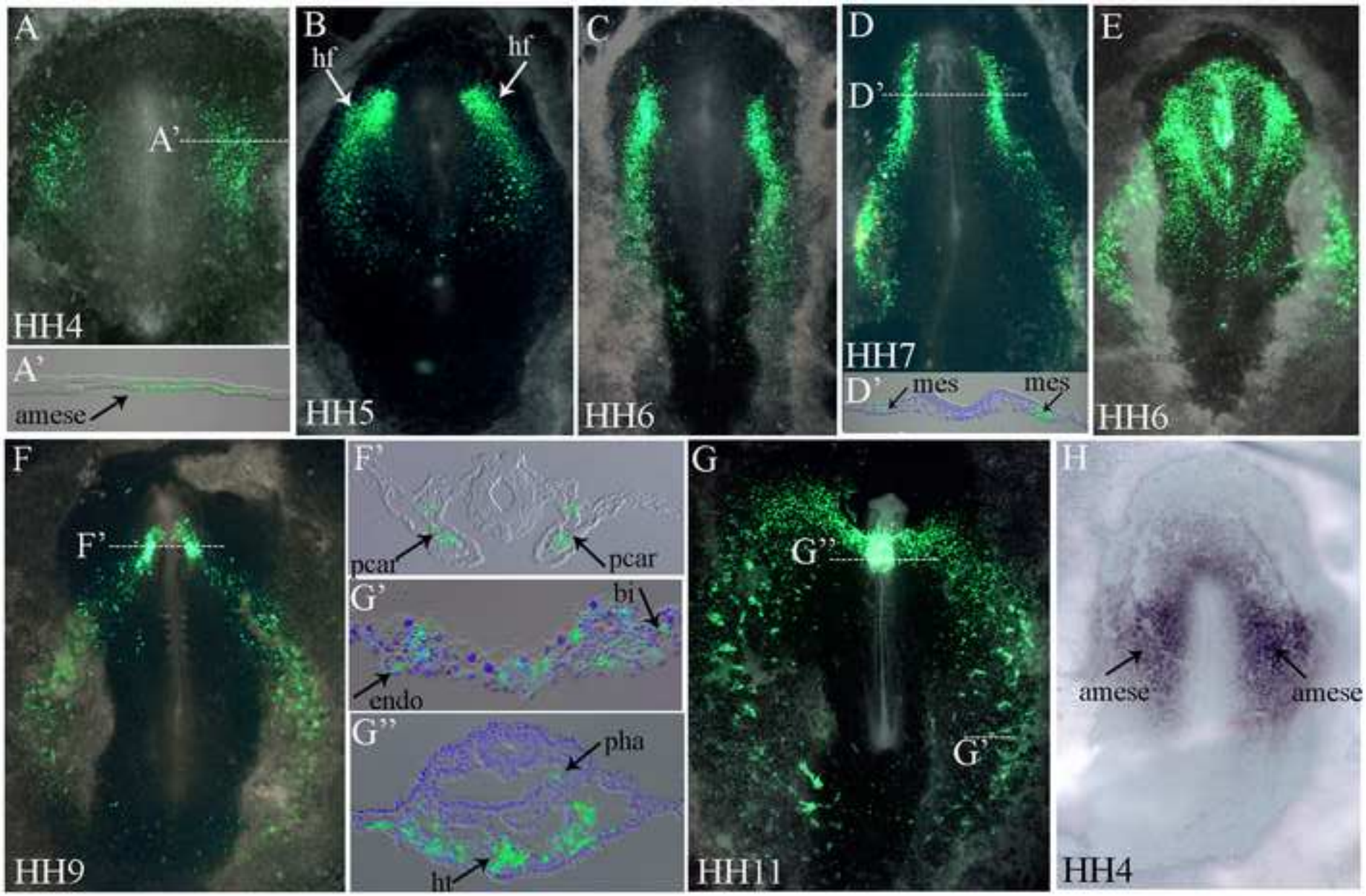

Figure 2

Click here to download high resolution image

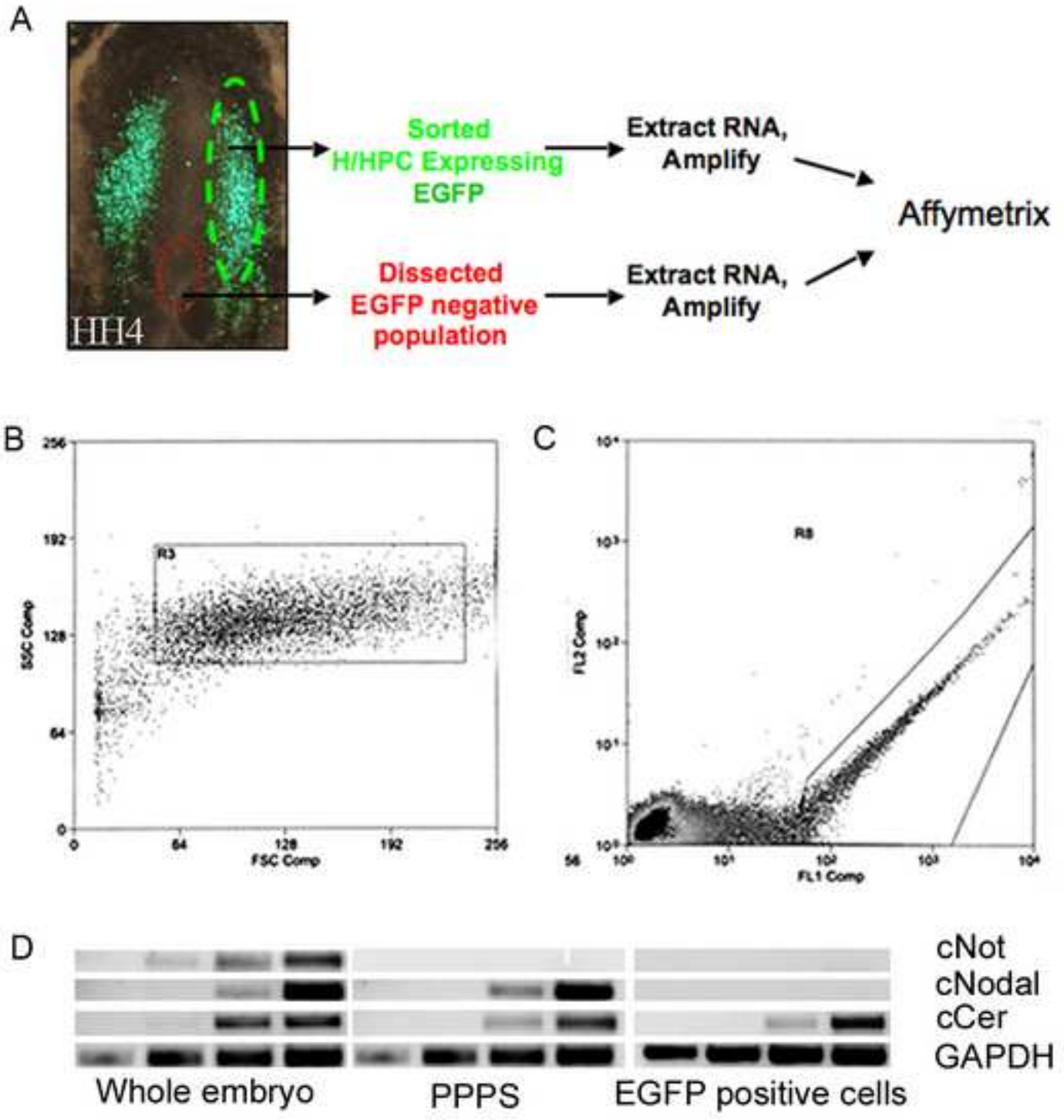

Figure 3

Click here to download high resolution image

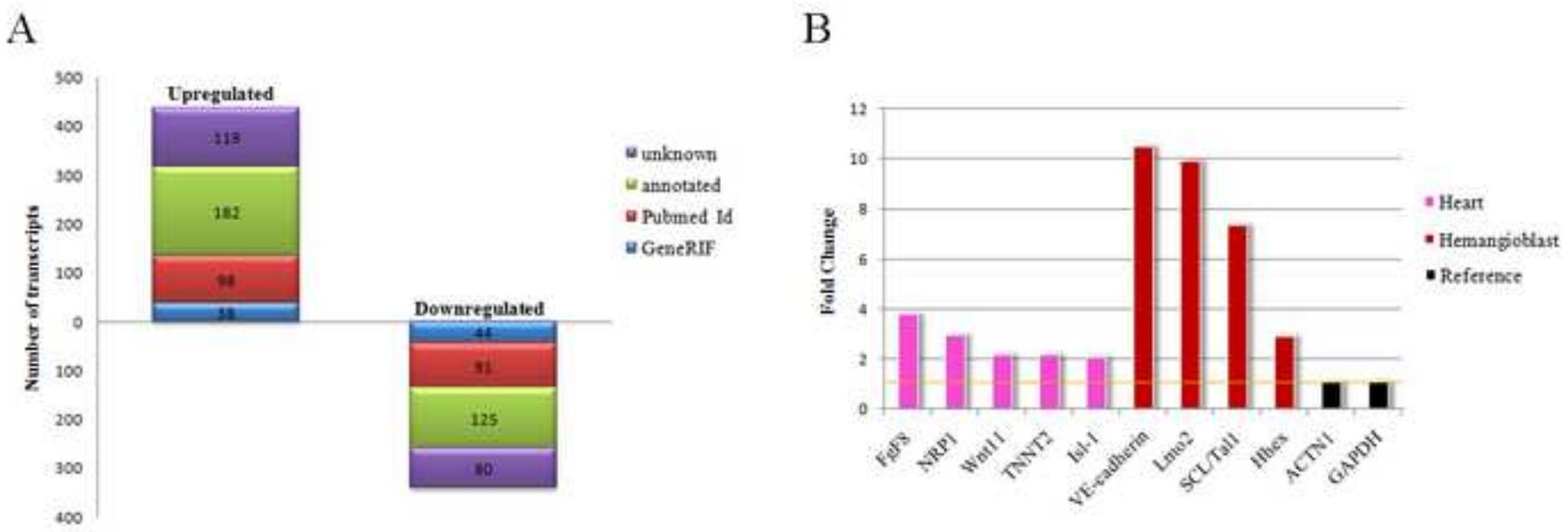

Figure 4

Click here to download high resolution image

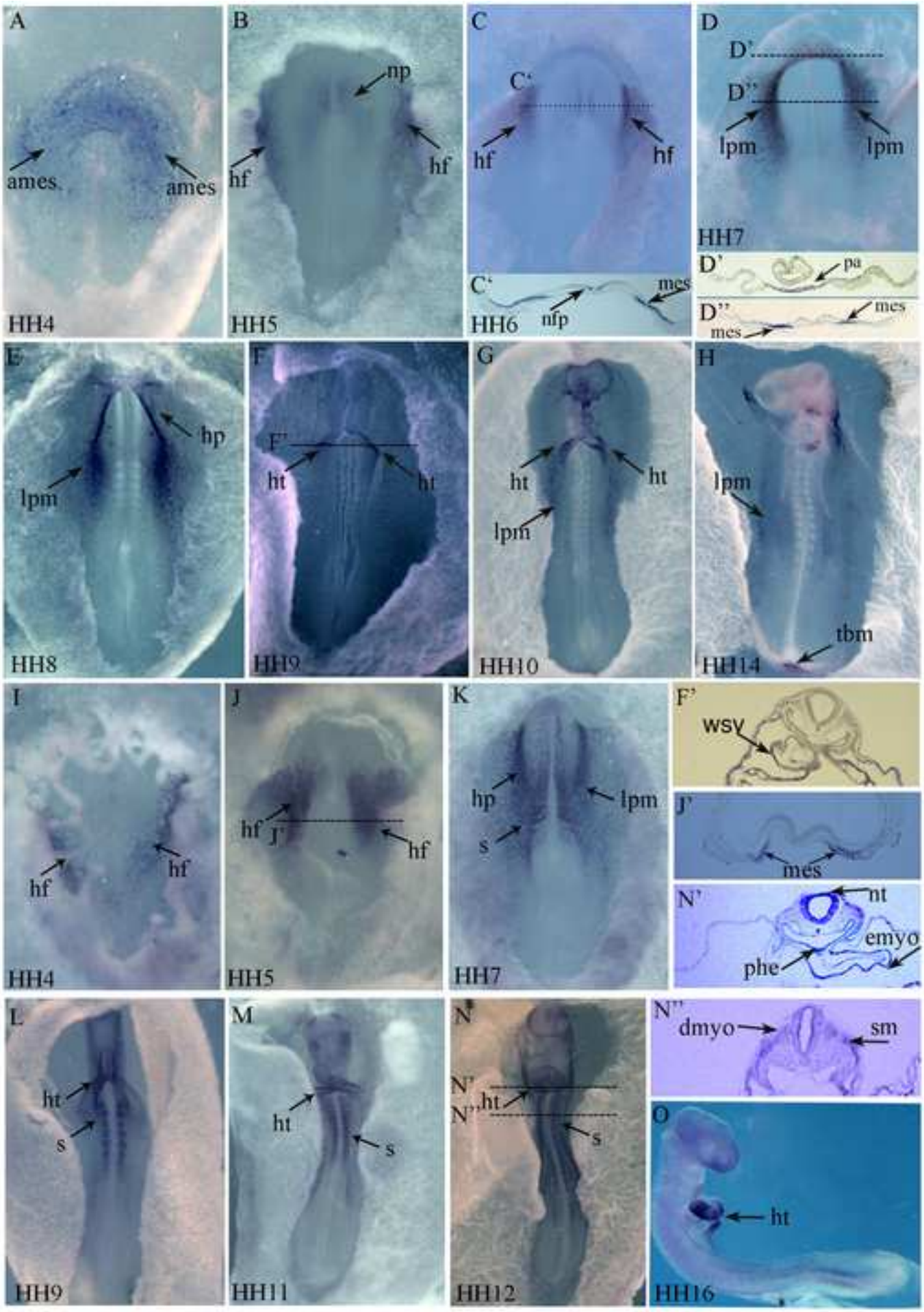

Figure 5

Click here to download high resolution image