Transcriptomic landscape of the primitive streak

Cantas Alev[1,2], Yuping Wu[1], Takeya Kasukawa[3], Lars M. Jakt[4], Hiroki R. Ueda[3,5] and Guojun Sheng[1,*]

SUMMARY

In birds and mammals, all mesoderm cells are generated from the primitive streak. Nascent mesoderm cells contain unique dorsoventral (D/V) identities according to their relative ingression position along the streak. Molecular mechanisms controlling this initial phase of mesoderm diversification are not well understood. Using the chick model, we generated high-quality transcriptomic datasets of different streak regions and analyzed their molecular heterogeneity. Fifteen percent of expressed genes exhibit differential expression levels, as represented by two major groups (dorsal to ventral and ventral to dorsal). A complete set of transcription factors and many novel genes with strong and region-specific expression were uncovered. Core components of BMP, Wnt and FGF pathways showed little regional difference, whereas their positive and negative regulators exhibited both dorsal-to-ventral and ventral-to-dorsal gradients, suggesting that robust D/V positional information is generated by fine-tuned regulation of key signaling pathways at multiple levels. Overall, our study provides a comprehensive molecular resource for understanding mesoderm diversification in vivo and targeted mesoderm lineage differentiation in vitro.

KEY WORDS: Primitive streak, Gastrulation, Mesoderm, Chicken, Dorsoventral, Transcriptome

INTRODUCTION

The mesoderm germ layer gives rise to the most diverse array of cell types in the animal body. One of the earliest steps in generating this diversity is the patterning of mesoderm precursors along the dorsoventral (D/V) body axis, the molecular mechanisms of which remain poorly understood. In both vertebrates and invertebrates, mesoderm cells emerge from the ectoderm through the gastrulation process, and in amniotes this process takes place in a transitory structure called the primitive streak (referred to hereafter as the streak) (Bellairs, 1986; Mikawa et al., 2004; Nakaya and Sheng, 2008; Psychoyos and Stern, 1996a). The streak is composed mainly of mesoderm precursors, which are still an integral part of the epiblast and with many epithelial characteristics, en route to becoming true mesoderm cells with mesenchymal morphologies (Nakaya and Sheng, 2009; Nakaya et al., 2008). The streak also maintains an 'identity' gradient within constituent cells and gives these cells unique D/V positional cues after their exit from the streak. Along the D/V body axis, mesoderm cells in amniotes are patterned to contribute later on to five broadly defined tissue lineages: the axial, paraxial, intermediate, lateral plate and extraembryonic mesoderm. This patterning event starts when epiblast-located mesoderm precursors enter the streak, and their D/V identities, although still labile at this stage, are reflected in their relative positions along the length of the streak, where they undergo epithelial-to-mesenchymal transition.

At Hamburger and Hamilton stage (HH) 4 of chick development, which is also known as the full primitive streak stage, fate map analyses have shown that cells located in the node region (the anterior ~1/8th of the streak) will give rise to more dorsal structures, including the axial mesoderm and medial somites, whereas the more posteriorly located cells in the streak will give rise to progressively more ventral types, including lateral somites from the remainder of the first quarter and the second quarter, intermediate mesoderm from the second quarter, lateral plate mesoderm from the remainder of the first quarter and the second and third quarters, and extraembryonic mesoderm from the second to fourth quarters (Garcia-Martinez et al., 1993; James and Schultheiss, 2003; Lopez-Sanchez et al., 2001; Nakazawa et al., 2006; Nicolet, 1970; Psychoyos and Stern, 1996a; Selleck and Stern, 1991). Although a general fate map is delineated in this way in avian and mammalian embryos, neither the boundary nor the exact fate is defined sharply at this stage of chick development (Inagaki et al., 1993; Psychoyos and Stern, 1996a; Psychoyos and Stern, 1996b). When dissected and cultured in a neutral or ectopic environment, pieces of streak tissue taken from along its anteroposterior (A/P) length tend to differentiate into the mesoderm lineages of their respective D/V identities (Abercrombie and Waddington, 1937; Garcia-Martinez and Schoenwolf, 1992; Inagaki and Schoenwolf, 1993; Nakazawa et al., 2006). Yet, experiments in which regions of the streak were extirpated or reversed anteroposteriorly indicated a labile D/V specification among streak cells and a regulative interaction between mesoderm precursors located in the streak and those located in the surrounding epiblast region (Abercrombie, 1950; Abercrombie and Bellairs, 1954; Garcia-Martinez and Schoenwolf, 1992; Grabowski, 1956; Inagaki et al., 1993; Joubin and Stern, 1999; Psychoyos and Stern, 1996b).

The molecular mechanisms by which the streak maintains a D/V gradient during its brief existence are not well understood. Central to our current understanding of vertebrate mesoderm D/V patterning is the BMP activity gradient generated by a balancing act between pro-BMP and anti-BMP signals (Dale et al., 1992; Dosch et al., 1997; Harland, 2004; Imai et al., 2001; Inomata et al., 2008; Jones et al., 1992; Jones and Smith, 1998; Khokha et al., 2005; Munoz-Sanjuan and Brivanlou, 2004; Piccolo et al., 1996; Schmid et al., 2000; Streit and Stern, 1999). In addition, the Wnt

[1]Laboratory for Early Embryogenesis, RIKEN Center for Developmental Biology, Kobe, Hyogo 650-0047, Japan. [2]Vascular Regeneration Research Group, Institute of Biomedical Research and Innovation, Kobe, Hyogo 650-0047, Japan. [3]Functional Genomics Unit, RIKEN Center for Developmental Biology, Kobe, Hyogo 650-0047, Japan. [4]Laboratory for Stem Cell Biology, RIKEN Center for Developmental Biology, Kobe, Hyogo 650-0047, Japan. [5]Laboratory for Systems Biology, RIKEN Center for Developmental Biology, Kobe, Hyogo 650-0047, Japan.

*Author for correspondence (sheng@cdb.riken.jp)

signaling pathway, which is responsible for the initial breaking of D/V symmetry before streak formation, and the FGF signaling pathway, which is mainly implicated in an earlier step of mesoderm induction, also contribute to mesoderm D/V patterning (Furthauer et al., 1997; Furthauer et al., 2004; Keren et al., 2008; Lekven et al., 2001; Marom et al., 1999; Muraoka et al., 2006; Yabe et al., 2003). Currently, it is unclear to what extent each of these, or any other, signaling pathway contributes to the generation of this D/V gradient. Nor is the exact molecular nature of this gradient known. Given the labile nature of their fate, it is essential to identify and understand the differences in global gene expression profiles among the mesoderm precursor populations that are destined to become distinct lineages and cell types later on. The importance of addressing this in vivo is underscored by the efforts currently being made in the field of stem cell biology to direct in vitro differentiation of cultured stem/pluripotent cells towards specific mesoderm lineages (Era et al., 2008; Murry and Keller, 2008; Takeuchi and Bruneau, 2009; Vijayaragavan et al., 2009).

Here we have taken a transcriptomic approach to address this question. Streak tissues were dissected from stage-matched HH4 chick embryos and were further divided into four equal pieces along its A/P length. Chicken genome arrays were used to obtain global gene expression profiles of these four streak regions. We show that ~40% of all chicken genes are expressed in the streak, one seventh of which exhibit significant differences in expression levels among the four streak regions. Gene ontology analyses revealed that different streak regions have overall identical profiles among expressed genes and very similar ontological profiles among differentially expressed genes. Genes involved in signaling and transcriptional regulation are highly represented among differentially expressed categories. All major groups of transcription factors and kinases are present in both dorsal-to-ventral (D→V) and ventral-to-dorsal (V→D) groups, with some subfamilies exhibiting exclusive D→V or V→D expression. Most core members of the BMP/TGFβ, Wnt and FGF pathways are expressed with no regional difference, whereas both negative and positive modulators of these pathways are present in both dorsal and ventral mesoderm precursors. Finally, a large number of novel genes with region-specific expression patterns are uncovered in this study, significantly increasing the scope of molecular players currently associated with mesoderm D/V patterning.

MATERIALS AND METHODS

Data acquisition and analysis
A flow chart outlining the data acquisition and analysis process is shown in Fig. S9 in the supplementary material. Details are described in the following sections.

Streak tissue collection and RNA in situ analysis
Fertilized hens' eggs were purchased from Shiroyama Farm (Kanagawa, Japan) and fresh eggs were incubated, without prior storage, at 38.5°C for ~18 hours to reach HH4. Embryos were collected in Pannett-Compton solution (Streit and Stern, 2008). Only embryos at HH4– or HH4 were used for streak dissection. The entire streak was cut out first and then cut into four equal pieces along its A/P length. Collected pieces were frozen in liquid nitrogen after pooling ~10 pieces, and ~90 pieces were combined later on to be used for RNA isolation as one sample. The numbering of samples corresponds to the original embryo/streak pool (e.g. A1, B1, C1 and D1 were derived from the same 85 streaks). Total RNAs were isolated using the RNeasy Kit (Qiagen). RNA concentrations were measured and equalized, and RNA quality was checked on a gel. Five micrograms of total RNA were used for each array. A standard protocol was followed for

whole-mount RNA in situ hybridization analysis. Details of primer sets, probe lengths and corresponding regions of probes generated are shown in Table S4 in the supplementary material.

GeneChip analysis
cDNA synthesis and cRNA labeling reactions were performed according to the one-cycle protocol provided by Affymetrix. Affymetrix high-density oligonucleotide arrays for *Gallus gallus* (GeneChip Chicken Genome Array) were hybridized, stained and washed according to the expression analysis technical manual (Affymetrix). The expression values were summarized by the RMA method. The resulting expression values were used in all the subsequent analyses. Raw datasets have been submitted to NCBI GEO with accession number GSE22230.

Identification and gene ontology (GO) assignments to expressed genes
For each streak region (A, B, C or D), an 'expressed gene set' was first assembled. Each GeneChip measurement for a given probe set was assigned with a 'present', 'absent' or 'marginal' call using the Affymetrix MAS5 detection algorithm. Probe sets with a present call in all three replicated measurements were retrieved. The Bioconductor 2.5.0 suite (Gentleman et al., 2004) was used for GO term assignment. The number of genes in the GO analysis was based on unique genes assigned to each GO term.

Differentially expressed gene analysis, gene clustering and chromosomal mapping
To identify differentially expressed genes, statistical tests with one-way ANOVA for each probe set were performed. Multiple comparisons were then corrected with false discovery rates (FDRs) and an FDR of less than 0.1 was chosen as significant ('significantly differentially expressed genes'). Stringent sets of the significantly differentially expressed genes were made by filtering out probe sets with ratios between their maximum and minimum mean expressions among four streak regions of less than 1.5. Next, a supervised clustering analysis was performed for classification. Fourteen template binary patterns, with repeated permutation on (0,1) with length 4 except for (0,0,0,0) and (1,1,1,1), were prepared. For each probe set in the significantly differentially expressed genes set, Pearson's correlation coefficients between its mean expression values in the four streak regions (A, B, C, D) and every binary pattern were calculated, and the most correlated pattern was chosen as its expression pattern. Probe sets with (A,B,C,D) = (1,0,0,0), (1,1,0,0) or (1,1,1,0) were grouped into the A^{group} set, and those with (A,B,C,D) = (0,1,1,1), (0,0,1,1) or (0,0,0,1) were grouped into the D^{group} set. A stringent version of A^{group} and D^{group} sets was made by applying the same clustering algorithm to the stringent significantly differentially expressed genes. Heatmaps of A^{group} and D^{group} genes representing expression changes from the overall mean expression were drawn. For the stringent A^{group} and D^{group} sets (ratio ≥1.5), we counted unique genes to which each GO term was assigned by the same method as described above. The stringent (ratio ≥1.5) A^{group} and D^{group} sets and the rest of the expressed genes were mapped on the chicken genome and their chromosomal map was drawn with the Geneplotter package in the Bioconductor 2.5.0 suite.

Pathway analyses
We used two pathway analysis tools: Ingenuity Pathways Analysis (Ingenuity Systems, www.ingenuity.com) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa and Goto, 2000). Stringent sets of the significantly differentially expressed genes were analyzed using Ingenuity Pathways Analysis. Molecules from the dataset that met the ratio cut-off of at least 1.5 and were associated with a canonical pathway in Ingenuity's Knowledge Base were considered for the analysis. This, and the KEGG pathway database, together with additional published data from PubMed, were used for the pathway diagrams in Fig. 4.

Known gene analysis
Examples of genes reported to be expressed differentially in the streak at this stage of chicken development were manually retrieved from the stringent A^{group} and D^{group} sets. A heatmap of these genes representing

expression changes from the overall mean expression value were then drawn. We also produced a bar chart representing fold-changes between A and D expression levels for these genes.

Comparison between streak (S) and neighboring (P) tissues
To identify significantly differentially expressed genes between P and S, we performed statistical tests with eBayes (Smyth, 2004). Multiple comparisons were then corrected with FDR, and an FDR of less than 0.1 was chosen as significant. A heatmap was drawn with the significantly differentially expressed genes set. We also made stringent sets of the significantly differentially expressed genes by filtering out probe sets that had ratios between mean expressions between P and S parts that were less than 1.5 and 2.0. For two-dimensional mapping of genes that were differently expressed in both A/B/C/D and P/S, these genes were identified by choosing genes in stringent (ratio ≥1.5 or ≥2.0) differentially expressed gene sets of both A/B/C/D and P/S. To quantify degrees of expression changes from A to D and from P to S for each commonly differentially expressed gene, linear functions were fitted to data points as [(1, expression in A), (2, in B), (3, in C), (4, in D)] and [(1, in P), (2, in S)], respectively, and slopes of the fitted functions were calculated. Finally, two-dimensional graphs of the slopes in A/B/C/D and in P/S were drawn for common stringent sets with a ratio of at least 1.5 or 2.0.

Counts for various probe sets
There are 38,535 probe sets in total on the Affymetrix chicken GeneChip. Of these, 26,277 probe sets have 14,931 unique GeneIDs (unique genes) and 12,258 probe sets have no GeneIDs. In clusterable significantly differentially expressed genes (for A, B, C, D), there are 7277 probe sets in total, with 5514 probe sets having 4362 unique GeneIDs and 1763 probe sets having no GeneID. In significantly differentially expressed genes with a ratio of at least 1.5 (for A, B, C, D), there are 2250 probe sets in total, with 1733 probe sets having 1324 unique GeneIDs and 513 probe sets having no GeneID. In A^{group} genes with a ratio of at least 1.5, there are 1173 probe sets, with 887 probe sets having 671 unique GeneIDs and 286 probe sets having no GeneID. In D^{group} genes with a ratio of at least 1.5, there are 1039 probe sets, with 827 probe sets having 633 unique GeneIDs and 212 probe sets having no GeneID. In significantly differentially expressed genes (P/S), there are 4325 probe sets in total, with 3015 probe sets having 2608 unique GeneIDs and 1310 probe sets having no GeneID. In significantly differentially expressed genes with a ratio of at least 1.5 (P/S), there are 571 probe sets in total with 452 probe sets having 357 unique GeneIDs and 119 probe sets have no GeneID.

RESULTS
Streak tissue preparation and chicken genome array analysis
We performed a comparative transcriptomic analysis of HH4 tissues from different A/P levels of the streak (Fig. 1A,B), reflecting the mesoderm D/V axis in early chick embryos. Although mesoderm cells are generated from the streak continuously from HH2 to at least HH9, HH4 was chosen because at this stage the streak is fully elongated, with precursors of all mesoderm lineages present and fate map analyses most fully documented (see Introduction). Six hundred embryos, incubated to the desired stage from freshly laid eggs, were collected. Of these embryos, 263 had a well-formed streak and no head process and were selected for dissection (Fig. 1C), corresponding to HH4– and HH4, but not HH4+. For each embryo, a rectangular piece of streak tissue, roughly equivalent to the brachyury-expressing domain, but not including brachyury-positive epiblast cells lateral to the streak (Fig. 1A,B), was cut out and further dissected into four equal pieces termed A, B, C and D, with A representing most anterior and therefore most dorsal, and D the most posterior and therefore most ventral (Fig. 1B). This division was chosen for technical reasons only and does not correspond to boundaries of
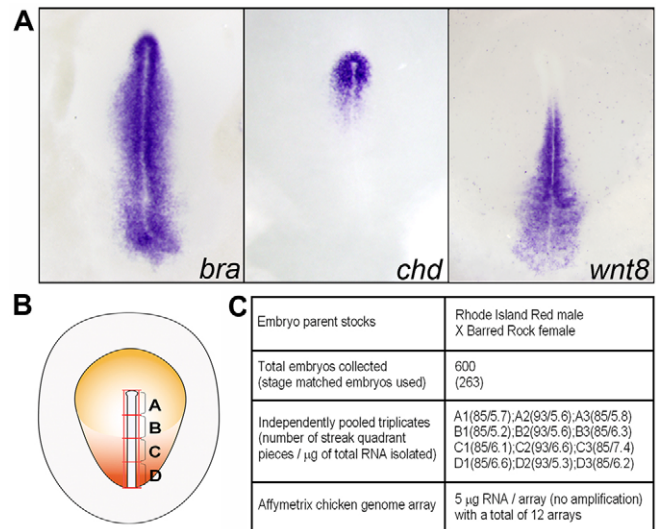
Fig. 1. Streak tissue sample preparation and sample quality analysis. (**A**) HH4 chick embryos stained for brachyury (*BRA*) (which marks the general mesoderm precursor territory) chordin (*CHD*) (dorsal mesoderm precursors) and *WNT8* (ventral mesoderm precursors) gene expression. (**B**) The four streak regions, A, B, C and D, of HH4 embryos dissected for RNA preparation. (**C**) Parental strains, number of embryos and amount of RNA obtained and used for the chicken genome array analysis.

mesoderm lineages reported from streak lineage-tracing experiments (see Materials and methods and Discussion for a comparison of these tissue pieces with known lineage territories). Each of the four regions was represented by three independent samples, with each sample containing 5-7 μg total RNA derived from 85-93 pooled pieces. Five micrograms of RNA from each sample were used to screen the Affymetrix chicken genome array without an amplification step (Fig. 1C). The overall array data analysis indicated the high quality of all RNA samples (see Fig. S1 in the supplementary material).

Global profile and gene ontology categorization
The Affymetrix chicken genome array contains 39,000 transcripts, corresponding to ~22,000 unique genes. We first analyzed the percentage of transcripts with detectable expression in each of the four regions based on present/absent calls in the analysis (see Materials and methods). Only transcripts with a present call in all three samples were considered to be expressed. Both total transcript-based and unique gene-based calculations indicated that ~40% of all genes in the entire chicken genome are expressed in each region of the streak. Genes expressed in different regions showed remarkably overlapping patterns, as more than 35% of all genes are expressed in all four regions. We next asked whether the difference in expressed transcripts (less than 4% in pair-wise comparison between different regions) represents molecular features unique to individual streak regions. To achieve this, we categorized expressed genes using Bioconductor software based on NCBI gene ontology (GO) term assignments curated for the chicken genome. We analyzed all genes that can be assigned a GO function and plotted GO terms containing more than 20 genes for any given term (see Fig. S2 in the supplementary material). In total, ~200 GO terms for molecular functions (see Fig. S2A in the supplementary material), 300 for biological processes (see Fig. S2B

in the supplementary material) and 85 for cellular components (see Fig. S2C in the supplementary material) contained more than 20 expressed genes. A uniform distribution for all GO terms using all three categorization criteria was observed for A, B, C and D regions (see Figs S2 and S3 in the supplementary material), suggesting that no particular molecular, cellular or biological aspect is unique for any one of the four streak regions.

Chromosomal location and cluster analysis of expressed transcripts

It has been well documented that D/V patterning in vertebrates is marked by regional differences in gene expression and in signaling and transcriptional regulation. We asked whether, among all expressed genes, there is any regional distinction in elevated transcript levels. Statistical analysis was performed based on expression values for each transcript in all samples (see Materials and methods). In total, 6567 transcripts (40% of expressed transcripts and 16% of all transcripts in the genome array) can be grouped with high statistical confidence into six out of 14 possible clusters (Fig. 2A), representing ~3700 unique genes. These six clusters represent either A→D (dorsal→ventral) or D→A (ventral→dorsal) asymmetric expression, and constitute over 90% of transcripts in all 14 possible clusters (Fig. 2A; see Fig. S4 in the supplementary material). With an arbitrary cut-off ratio of 1.5 (highest/lowest) (see Table S1 in the supplementary material for the full list), 2212 transcripts (14% of expressed transcripts and representing 1200 unique genes) can be grouped into these six clusters (Fig. 2B). Among them are many genes, the expression patterns of which have been described at this stage of vertebrate development (see Fig. S5 in the supplementary material). The top
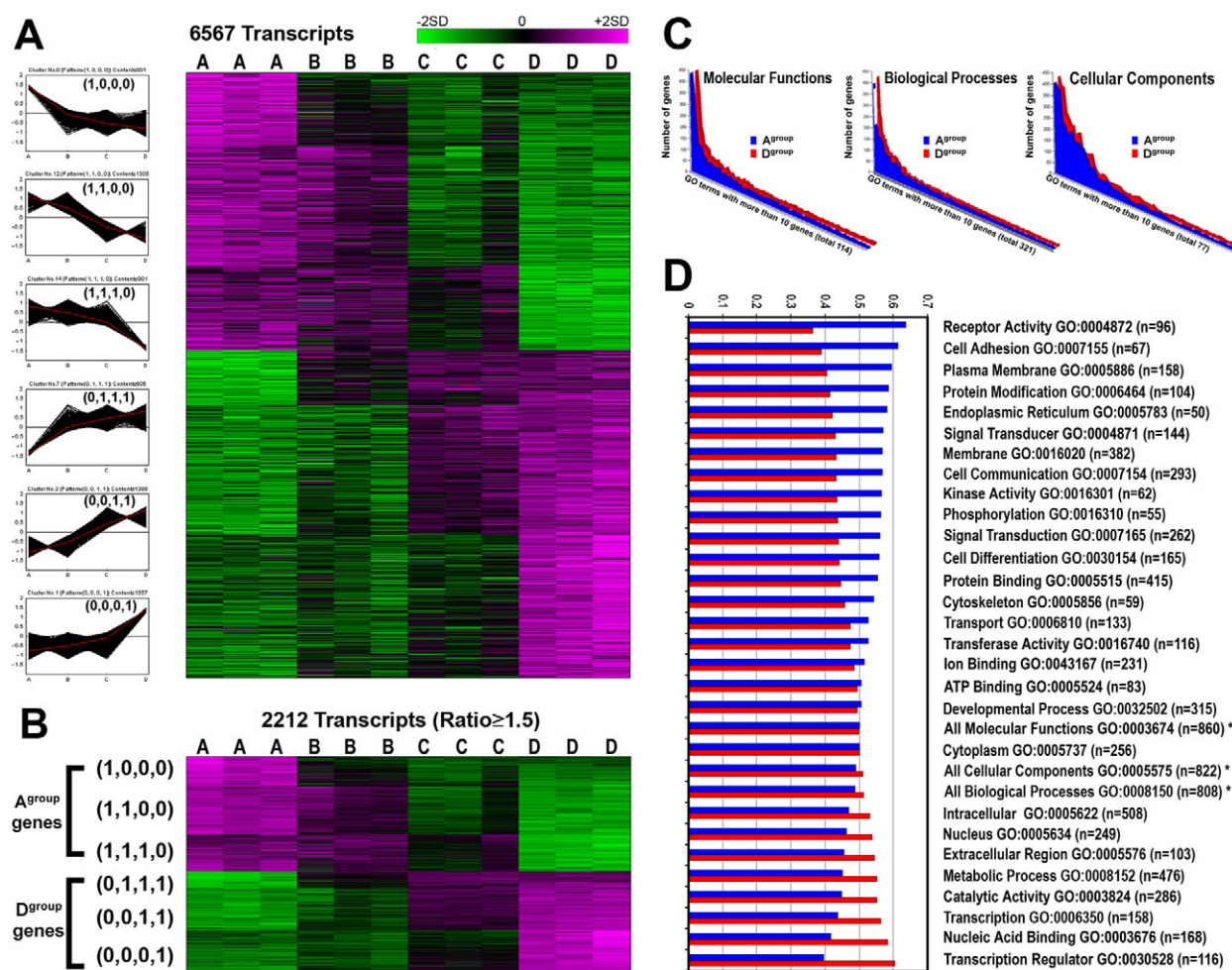
Fig. 2. Clustering of expressed and differentially expressed genes and GO analysis of differentially expressed genes. (**A**) Expressed genes are clustered into one of 14 possible clusters, with six major clusters shown here. Details of the clustering algorithm are given in the Materials and methods. 1, higher expression levels; 0, lower/absent expression levels. Purple to green in the heatmap indicates high to low expression levels. (**B**) Genes in these six clusters with at least a 1.5-fold difference between the highest and lowest expression levels. A^{group} (genes with higher levels dorsally) contains three clusters (1,0,0,0, 1,1,0,0 and 1,1,1,0), and D^{group} (genes with higher levels ventrally) contains three clusters (0,1,1,1, 0,0,1,1 and 0,0,0,1). (**C**) GO analysis of genes shown in B, comparing A^{group} and D^{group} genes. The overall distributions are similar, with some GO terms showing a difference. (**D**) Examples of GO terms with over 50 genes. The *x*-axis indicates the percentage of the total number of differentially expressed genes for a particular GO term that is present in either A^{group} (blue) or D^{group} (red). All three main GO categories (asterisks) have equal numbers in A^{group} and D^{group}, as is the case for many specific GO terms (e.g. ATP binding). Prominently, three GO terms have a higher representation of A^{group} genes (receptor activity, cell adhesion and plasma membrane) and three other terms have a higher representation of D^{group} genes (transcriptional regulator, nucleic acid binding and transcription).

20 genes in each cluster are listed in Table 1. The remaining eight clusters (see Fig. S4 in the supplementary material) contain 36 genes that exhibit differential expression with a 1.5-fold cut-off ratio (see Table S2 in the supplementary material), including *LEFTY2* (0110 cluster pattern; 4-fold difference). These clusters will not be discussed further.

Table 1. Top 20 differentially expressed genes in each of the six clusters

Cluster pattern	Ratio	Gene	Cluster pattern	Ratio	Gene	Cluster pattern	Ratio	Gene
A[group] genes								
(1,0,0,0)	5.79	FZD8	(1,1,0,0)	5.53	ADMP	(1,1,1,0)	3.86	EDNRB
(1,0,0,0)	4.01	FOXD3	(1,1,0,0)	5.03	Unknown (LOC421390)	(1,1,1,0)	3.80	Unknown (BU272039)
(1,0,0,0)	3.45	ARHGAP15	(1,1,0,0)	5.01	OTX2	(1,1,1,0)	3.08	FGF8
(1,0,0,0)	3.19	HHEX	(1,1,0,0)	4.94	GSC	(1,1,1,0)	3.06	EOMES
(1,0,0,0)	2.96	LBH	(1,1,0,0)	4.21	CHRD	(1,1,1,0)	2.83	CER1
(1,0,0,0)	2.84	CRHR1	(1,1,0,0)	4.09	FZD8	(1,1,1,0)	2.76	FGF18
(1,0,0,0)	2.80	SST	(1,1,0,0)	4.00	ZIC2	(1,1,1,0)	2.59	TLR2
(1,0,0,0)	2.59	FOXA2	(1,1,0,0)	3.52	RUNX1T1	(1,1,1,0)	2.40	RASGEF1B
(1,0,0,0)	2.58	GRM7	(1,1,0,0)	3.50	Unknown	(1,1,1,0)	2.25	Unknown (ChEST845i18)
(1,0,0,0)	2.58	CA2	(1,1,0,0)	3.49	SFRP1	(1,1,1,0)	2.02	PODXL
(1,0,0,0)	2.56	SOX2	(1,1,0,0)	3.49	LHX1	(1,1,1,0)	1.98	Unknown (ChEST228d14)
(1,0,0,0)	2.52	B3GAT2	(1,1,0,0)	3.45	TMEFF2	(1,1,1,0)	1.93	USP44
(1,0,0,0)	2.50	Unknown (LOC425015)	(1,1,0,0)	3.42	Unknown (LOC424623)	(1,1,1,0)	1.91	BTBD6
(1,0,0,0)	2.42	HDAC9	(1,1,0,0)	3.34	GNOT1	(1,1,1,0)	1.85	LRIG3
(1,0,0,0)	2.29	HFE2	(1,1,0,0)	3.17	CNOT2	(1,1,1,0)	1.82	Unknown (ChEST401i9)
(1,0,0,0)	2.25	STMN2	(1,1,0,0)	3.16	PCSK5	(1,1,1,0)	1.77	Unknown
(1,0,0,0)	2.13	Unknown (LOC768462)	(1,1,0,0)	3.15	Unknown (ChEST397p8)	(1,1,1,0)	1.73	CNTN4
(1,0,0,0)	2.02	ISM1	(1,1,0,0)	3.11	Unknown (ChEST286i13)	(1,1,1,0)	1.71	NFASC
(1,0,0,0)	1.99	DMRT3	(1,1,0,0)	3.04	THPO	(1,1,1,0)	1.60	KIF1A
(1,0,0,0)	1.98	TNFSF15	(1,1,0,0)	2.94	RASD1	(1,1,1,0)	1.60	PHACTR3
D[group] genes								
(0,0,0,1)	3.14	MSX2	(0,0,1,1)	5.52	HOXA2	(0,1,1,1)	4.73	CDX4
(0,0,0,1)	3.07	HOXA9	(0,0,1,1)	4.86	SZL	(0,1,1,1)	3.65	CDX2
(0,0,0,1)	2.90	FLI1	(0,0,1,1)	4.06	KDR	(0,1,1,1)	3.41	Unknown (CK614519)
(0,0,0,1)	2.77	BETA3	(0,0,1,1)	3.71	SOCS2	(0,1,1,1)	3.17	HOXB1
(0,0,0,1)	2.61	FET1	(0,0,1,1)	3.36	HOXA3	(0,1,1,1)	3.03	WNT5A
(0,0,0,1)	2.58	Unknown (ChEST457l4)	(0,0,1,1)	3.00	HOXB3	(0,1,1,1)	2.95	MYC
(0,0,0,1)	2.58	ALDH1A2	(0,0,1,1)	2.96	APCDD1	(0,1,1,1)	2.92	MESP2
(0,0,0,1)	2.51	CORO2A	(0,0,1,1)	2.95	NOXO1	(0,1,1,1)	2.73	WNT8A
(0,0,0,1)	2.48	Unknown (ChEST264i23)	(0,0,1,1)	2.91	BMP7	(0,1,1,1)	2.69	HAPLN1
(0,0,0,1)	2.41	RBP4	(0,0,1,1)	2.80	Unknown (LOC524813)	(0,1,1,1)	2.61	CXCL14
(0,0,0,1)	2.38	GPD1	(0,0,1,1)	2.79	MT3	(0,1,1,1)	2.56	KLF1
(0,0,0,1)	2.37	KITLG	(0,0,1,1)	2.78	Unknown (LOC425711)	(0,1,1,1)	2.42	Unknown (ChEST847k15)
(0,0,0,1)	2.34	ALDH1A1	(0,0,1,1)	2.71	KLF15	(0,1,1,1)	2.34	PIWIL1
(0,0,0,1)	2.26	VEGFA	(0,0,1,1)	2.61	Unknown (LOC421682)	(0,1,1,1)	2.24	WNT3A
(0,0,0,1)	2.22	Unknown (LOC771069)	(0,0,1,1)	2.30	ADA	(0,1,1,1)	2.17	FZD10
(0,0,0,1)	2.12	SLC12A3	(0,0,1,1)	2.30	PLXNA2	(0,1,1,1)	2.15	MSX1
(0,0,0,1)	2.12	UGT1A1	(0,0,1,1)	2.30	IRF2	(0,1,1,1)	2.14	SPAM1
(0,0,0,1)	2.10	BPIL2	(0,0,1,1)	2.17	Unknown (ChEST594o6)	(0,1,1,1)	2.09	COL13A1
(0,0,0,1)	2.09	HAND2	(0,0,1,1)	2.16	Unknown (ChEST539d8)	(0,1,1,1)	2.07	ADAMTS3
(0,0,0,1)	2.07	RGN	(0,0,1,1)	2.13	EVX1	(0,1,1,1)	2.00	DLX5

Top 20 genes in each of the six clusters. Three of the A[group] (1,0,0,0, 1,1,0,0 and 1,1,1,0) and three of the D[group] (0,0,0,1, 0,0,1,1 and 0,1,1,1) are listed, indicating their cluster pattern, fold difference (ratio in log2 scale) and gene symbols. Details of each gene can be found in Table S1 in the supplementary material. Prominent among them are many well-described genes, including *FZD8* and *ADMP* in A[group] and *MSX2* in D[group]. The list also includes many unknown genes or genes previously not associated with mesoderm D/V patterning.

DEVELOPMENT

Developmental progression and tissue fate specification often involve coordinated transcriptional regulation. In mice, it has been reported that clustered transcriptional hotspots in the genome accompany the onset of organogenesis (Mitiku and Baker, 2007). We asked whether a similar phenomenon occurs in streak tissues fated for different mesoderm lineages. All 6600 transcripts that can be clustered by our statistical analysis were plotted onto the chicken chromosomes. Transcripts exhibiting a greater than 1.5-fold difference are shown in red (ventral high) or blue (dorsal high), whereas those with a less than 1.5-fold difference are shown in gray (see Fig. S6 in the supplementary material). Overall, these genes are scattered throughout the genome and no hotspot for either expressed or differentially expressed genes was detected.

These data indicated that, on the one hand, the majority of expressed genes (60% by statistical clustering and 85% by an arbitrary 1.5-fold cut-off) show no regional difference, whereas, on the other hand, a large number of genes did show a statistically significant difference in their expression levels. We therefore performed further analysis on genes exhibiting greater than 1.5-fold A→D (1000, 1100 and 1110 clusters) or D→A (0001, 0011 and 0111 clusters) asymmetric expression. We asked whether genes in these six clusters represent specific GO categories that highlight differential biological activities involved in D/V patterning. These genes were grouped as either A^{group} (for A→D clusters and representing the dorsal-high group) or D^{group} (for D→A clusters and representing the ventral-high group) genes (Fig. 2B), and a GO analysis similar to that described above was performed. Among 2212 transcripts, ~40% can be assigned with a GO term in all three categories. The majority of GO terms still exhibit little or no difference between A^{group} and D^{group} genes (Fig. 2C), and an approximately equal number of differentially expressed genes are represented in both the A^{group} and D^{group}. Although the number of genes assignable to a particular GO term does not necessarily indicate its importance or bias of usage in regulation, this result did suggest that even among differentially expressed genes, no particular aspect of molecular, cellular or biological activities is unique to, or represented in, dorsal or ventral streak tissue.

Genes involved in transcriptional and signaling regulation

Patterning events during development are often regulated by graded signaling and transcriptional activities. We investigated these two aspects in more detail. Interestingly, two classes of GO terms, one representing transcriptional regulation and the other receptor-mediated signaling regulation, exhibited their strongest uneven distribution between A^{group} and D^{group} genes. With all three GO categorization methods, we observed that the D^{group} has a higher number of differentially expressed genes involved in transcriptional regulation, whereas the A^{group} has more differentially expressed genes involved in signal transduction (Fig. 2D). To further investigate these differences, we manually curated all A^{group} and D^{group} genes. Among them, a total of 114 transcription factors with definable DNA-binding domains based on the transcription factor classification program TRANSFAC (BIOBASE, Germany) were found (Fig. 3). Forty-six of them belong to the A^{group} and 68 to the D^{group} (Fig. 3, blue and red, respectively). All four major superclasses (basic domain, zinc-coordinating DNA-binding domain, helix-turn-helix and beta-scaffold) and several unclassifiable transcription factors are represented. About half of them belong to the helix-turn-helix superclass, of which four classes (homeodomain, Forkhead/Winged helix, tryptophan clusters and paired box) are represented, the

homeodomain class being the largest with 39 genes. A bias was seen at the subfamily level, with all genes in the Labial, Antp, Abd-B, Caudal, Msh and Dll subfamilies of the homeodomain class belonging to the D^{group}, as were genes of the Cys4 zinc-finger class, nuclear receptor class, Teashirt subfamily and Grainyhead class. All genes in the Emx homeodomain subfamily, Sox family and all non-homeodomain-only homeodomain-containing genes belong to the A^{group}. Several large groups (Krüppel subfamily, Prd subfamily of homeodomain, Forkhead/Winged helix class, Ets family, Tcf family and Hairy family) have genes in both the A^{group} and D^{group}.

In addition to transcriptional regulation, genes involved in signal transduction represent the other major category revealed by the GO analysis. We carried out a systematic analysis for all known signaling pathways using Ingenuity Pathways Analysis software (see Materials and methods). Among 2212 transcripts with a 1.5-fold cut-off, 88 pathways are represented with at least four component genes, the top three signaling pathways being the Wnt, TGFβ/BMP and FGF pathways (see Table S3 in the supplementary material). This is in good agreement with these three pathways being documented as the major players in early D/V patterning (see Introduction). Examples of other pathways revealed by our analysis are (numbers refer to A^{group} + D^{group} throughout): keratan sulfate biosynthesis (5+5; five upregulated in each), reflecting extensive sugar chain modification; axonal guidance signaling (30+19), possibly reflecting early neuronal specification of ectoderm cells mixed with mesoderm-fated cells; ephrin receptor signaling (16+7), reflecting differential regulation in the migratory behavior of nascent mesoderm cells; G protein-coupled receptor signaling (15+6); cAMP-mediated signaling (9+6); calcium signaling (8+10); and integrin signaling (11+5) (for a complete list of pathways and their component genes, see Table S3 in the supplementary material). Interestingly, retinoic acid signaling pathway members were confined to the D^{group}. The actual number of genes involved in these respective pathways is likely to be higher than that recorded here, as our manual curation for three major pathways indicated that some genes were missed owing to minor differences in the nomenclature of orthologous genes and to a possibly incomplete dataset in the Ingenuity Pathways database.

## Components of three major pathways: Wnt, TGFβ/BMP and FGF

The three best-represented pathways in our analysis, i.e. the Wnt, TGFβ/BMP and FGF pathways, are known to be involved in mesoderm induction and its D/V patterning. We asked whether our global analysis could reveal new features of these signaling pathways in these events. Our pathway analysis identified 28 differentially expressed genes related to the Wnt pathway (16+12), 13 for the TGFβ/BMP pathway (4+9) and 12 for the FGF pathway (8+4). In addition to these genes, we manually analyzed the expressed profiles of all genes shown to be involved in these three pathways by including genes that are expressed but that show a less than 1.5-fold difference. These genes are in gray in Fig. 4 as those with no significant change, whereas differentially expressed genes (at least 1.5-fold) are marked in blue (A^{group}) or red (D^{group}). In addition to confirming pathway genes that have been shown previously in experimental settings, we observed three prominent features in our analysis. First, for all three pathways, most of the core components are expressed without a D/V asymmetry (gray). This includes *LRP5*, *DVL1*, *CTNNB1*, *GSK3B*, *AXIN1*, *PPP2CA* and *APC1* for the Wnt pathway; *BMPR1A/B*, *BMPR2*, *TGFBR1*, *TGFBR2*, *ACVR1*, *ACVR2A/B*, *INHBA*, *INHBB*, and *SMAD1-3* for the TGFβ/pathway; and *FRS2*, *GRB2*, *SOS1*, *RAS*, *RAF1*, *MEK*
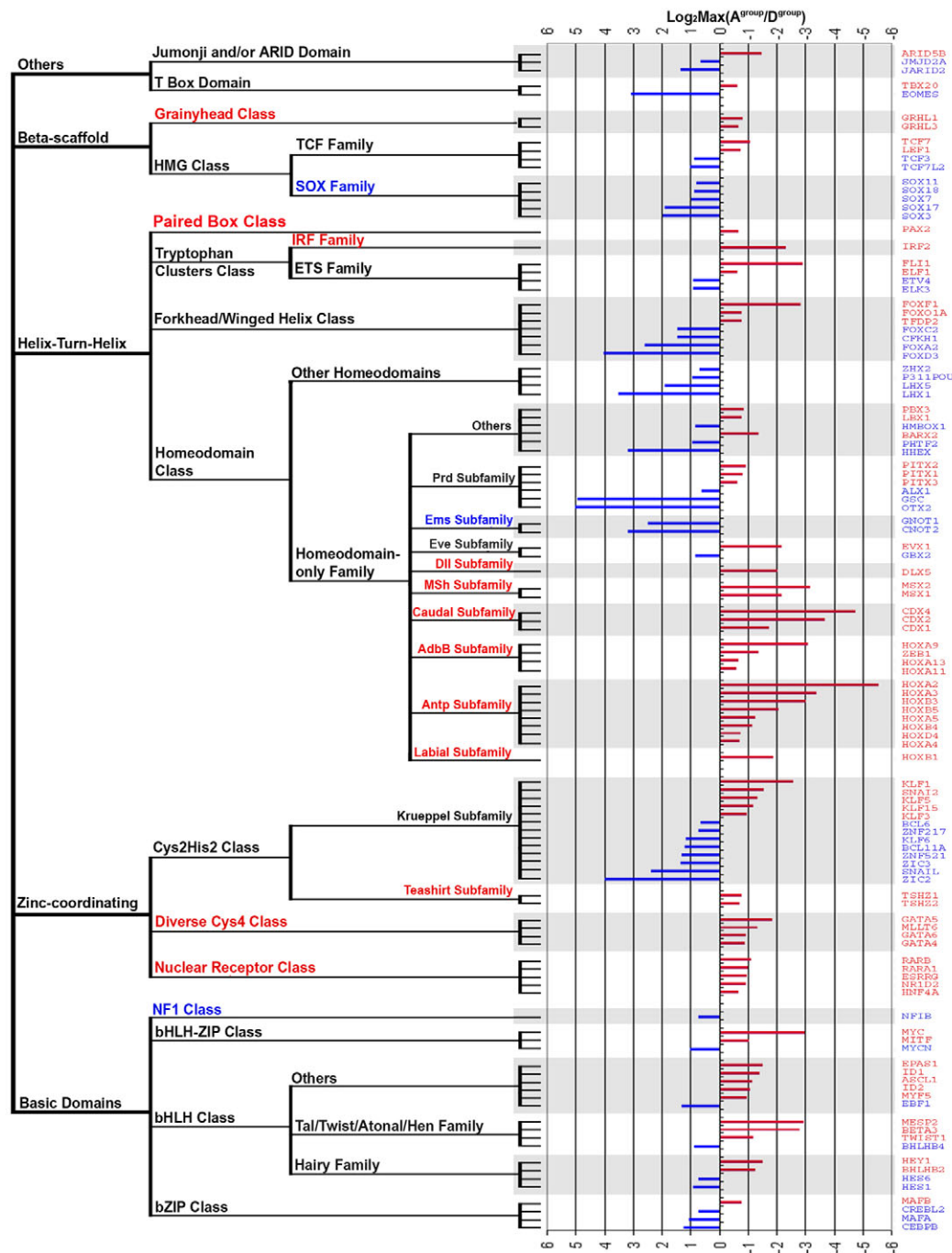
Fig. 3. Categorization of all transcription factors differentially expressed in the streak. Differentially expressed genes that encode a transcription factor with a definable DNA-biding domain are grouped based on their DNA-binding domain categories, with a total of 114 transcription factors and with 46 belonging to the A^{group} (blue bars) and 68 belonging to the D^{group} (red bars). x-axis indicates the fold difference. All transcription factor classes are represented here, with some families being exclusive in either A^{group} or D^{group}.

and *ERK2* for the FGF pathway. Second, regulatory genes are differentially expressed and regulation takes place at multiple levels for each pathway, including extracellular, receptor, cytoplasmic and nuclear. Third, both positive and negative regulators are present in both A^{group} and D^{group} genes. For the Wnt pathway, Wnt genes (*WNT2*, *WNT3*, *WNT5A*, *WNT5B* and *WNT8A*) are present with a V→D gradient. *FZD5/7/8* and *FRZB* are dorsally upregulated, whereas *FZD10* is ventrally upregulated. Both positive and negative regulators for the canonical Wnt pathway have both dorsally (*CER1*, *DKK1*, *SFRP1*, *RSPO1/3*, *SHISA*, *MARK3*, *DACT2*, *TCF3*, *TCF4*, *GBX2*, *HHEX*, *SOX3/7/17/18* and *TLE4*) and ventrally (*NDP*, *GIPC2*, *SRC*, *TCF7*, *LEF1*, *NOTUM*, *GPC3*, *SDC3*, *NKD1*, *DACT1* and *RARA/B*) upregulated inputs. For the

Wnt planar cell polarity (PCP) pathway, however, most regulators, both positive (*DKK1*, *ROR1*, *GPC4*, *SDC2*, *MAPK8IP1*, *DAAM1* and *DACT2*) and negative (*SFRP1*, *PRICKLE2*, *SHISA* and *MARK3*), are present only in the A^{group}. For the TGFβ superfamily pathway, major regulation is observed in the BMP branch of the pathway. All BMPs (*BMP2/4/7*) are ventrally upregulated, whereas negative regulators for BMP/BMPR binding are present both dorsally (*ADMP*, *CER1* and *CHRD*) and ventrally (*SIZZLED*, *SDC3*, *GPC3* and *BAMBI*). *SMAD6/7/9* show ventral upregulation and *SMAD5* shows dorsal upregulation. For the FGF pathway, a dorsal upregulation is seen for *FGF1/8/13/18*, *FGFR1/2*, *GPC4*, *SHISA*, *SEF*, *SPRY1*, *RASGEF1A/B*, *DUSP6/14* and *ETV4*, whereas a ventral upregulation is seen for *FGF4/19*, *SULF1*,

DEVELOPMENT
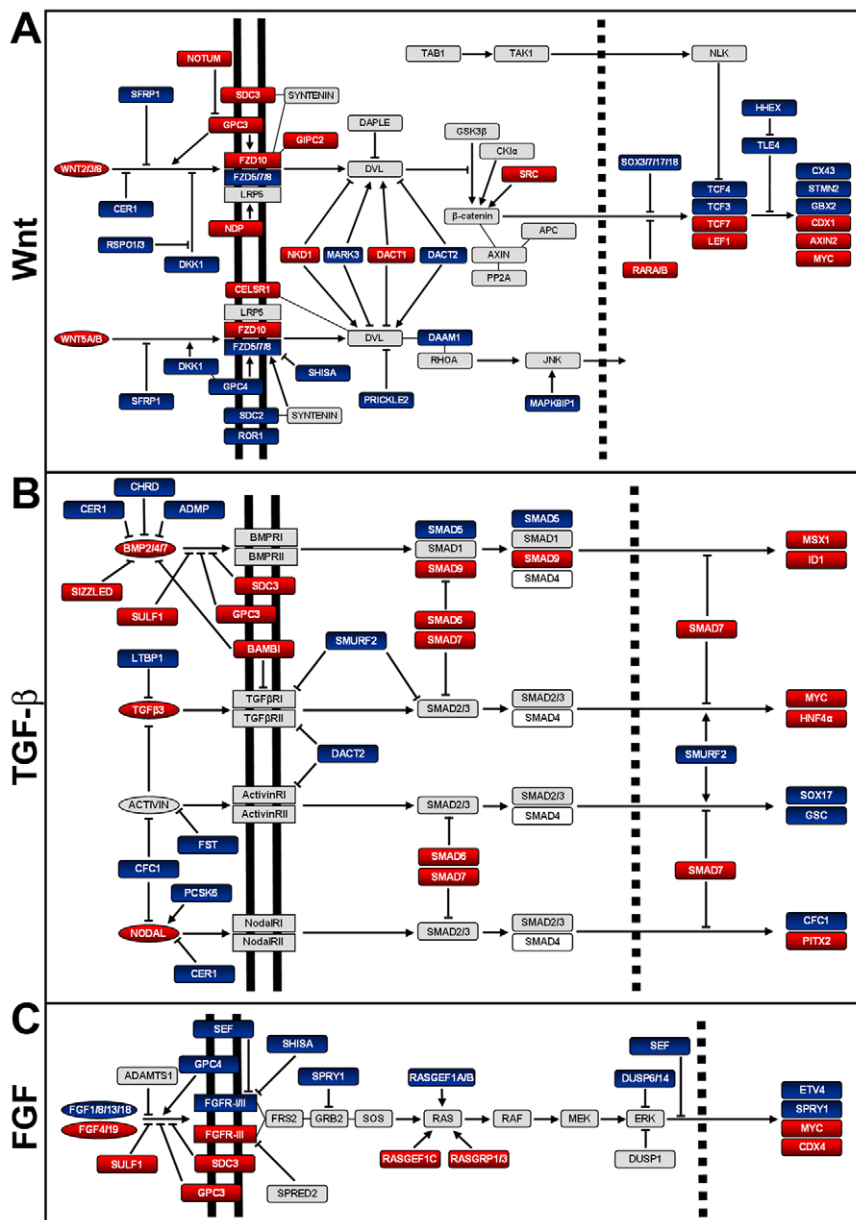
Fig. 4. Regulators of the Wnt, TGFβ/BMP and FGF pathways among differentially expressed genes. (**A-C**) The Wnt (A), TGFβ (B) and FGF (C) pathways are shown. Gray indicates expressed genes with less than a 1.5-fold difference. Blue indicates A^{group} genes (dorsal high) and red indicates D^{group} genes (ventral high). Most of the core members of these pathways are expressed uniformly. Negative and positive regulators of each pathway are present among both A^{group} and D^{group} genes, representing multiple points of regulation (extracellular, membrane, intracellular and nuclear). *SMAD4* is colored white as it is expected to be present, but no Affymetrix probe set can be clearly associated with this gene. Double line, cell membrane. Dotted line, nuclear membrane.

FGFR3, *SDC3*, *PGC3*, *RASGEF1C*, *RASGRP1/3*, *CDX4* and *MYC*. Overall, in addition to revealing novel members of these pathways that are potentially involved in mesoderm D/V patterning, our data point to a complex network of positive and negative regulation for both dorsal and ventral tissues, which possibly ensures the robustness of the D/V signaling gradient.
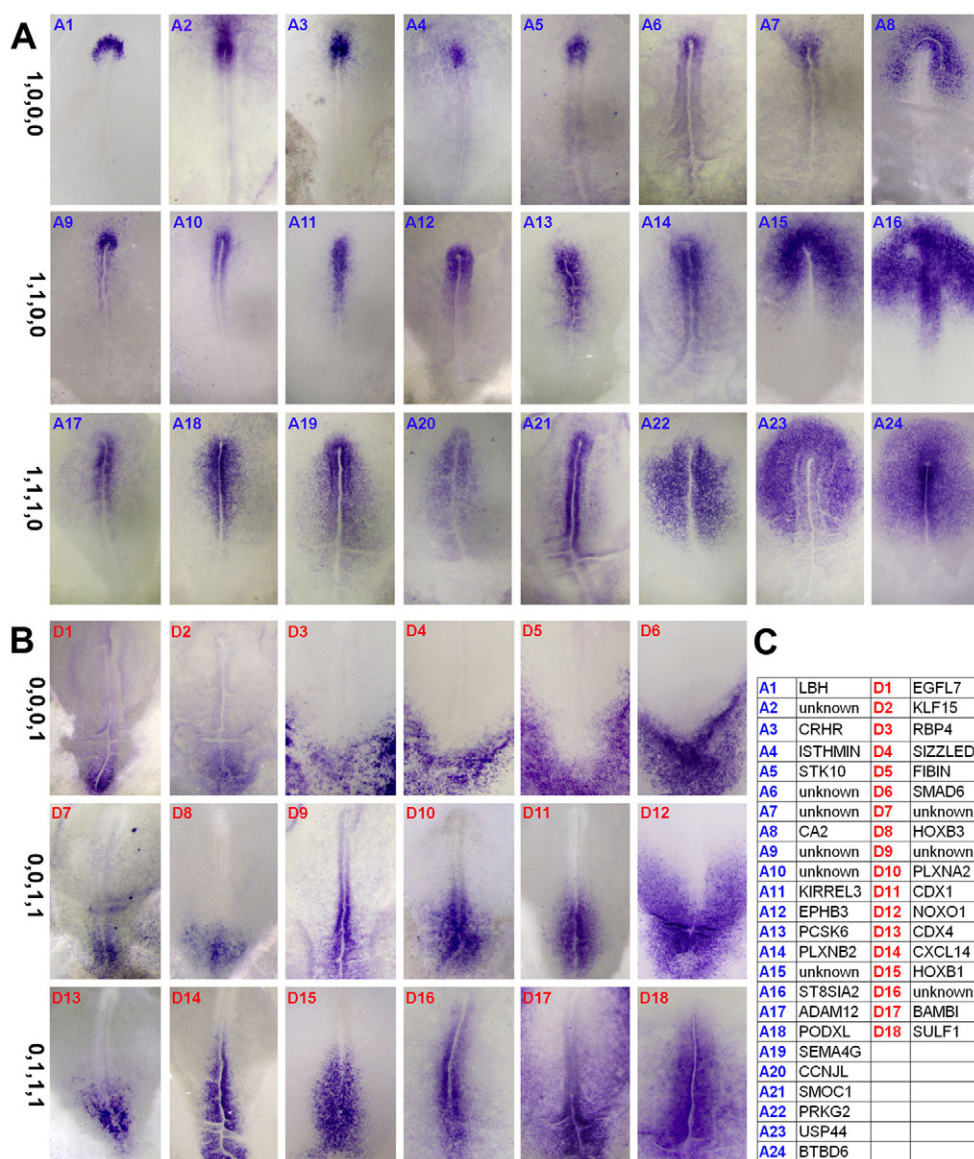
Verification by in situ hybridization

Expression data from genome array analysis represent reliable readouts and are less biased than traditional in situ-based analyses. Nevertheless, as an alternative verification of our data quality, we randomly selected genes from our array analysis representing each of the six clusters for in situ verification. Probes were prepared based on sequences in the genome array dataset for genes showing A→D (24 genes) or D→A (18 genes) asymmetric expression (see Materials and methods for details). Only genes that have not been reported in the chick system were selected, representing those of both known and unknown molecular function (see Table S4 in the

supplementary material). Overall, a near-perfect correlation was seen between the in situ expression pattern of a gene and the cluster category to which it belongs (Fig. 5). Among these are many genes of unknown function that are conserved among vertebrate species and with strong and specific expression, as are many genes of known function unrelated to D/V patterning, including *LBH*, *CRHR*, *STK10*, *PCSK6*, *EPHB3*, *ST8SIA2*, *PODXL*, *SEMA4G*, *EGFL7*, *FIBIN*, *PLXNA2*, *NOXO1*, *CXCL14* and *SULF1*. In both groups, most of the genes tested showed restricted expression in the streak and a small percentage had expression domains that extend into the neural plate territory (for the A^{group}) or the non-neural ectoderm territory (for the D^{group}).

DISCUSSION

In this work, we describe a transcriptomic profiling of positional differences of streak mesoderm precursors in chick embryos, the first such report in any vertebrate species. The quality of such an analysis depends on the amount of developmental stage- and

Fig. 5. In situ validation of A^group and D^group genes. (**A-C**) RNA in situ hybridization analyses of 24 genes in the A^group (A) and 18 in the D^group (B), randomly selected and with expression patterns not previously described for streak development. Each of the three A^group clusters (1,0,0,0; 1,1,0,0 and 1,1,1,0) is represented by eight genes, and each of the three D^group clusters (0,0,0,1; 0,0,1,1 and 0,1,1,1) is represented by six genes. Gene symbols, when known, are shown in C. Details of corresponding gene or expressed sequence tag names and regions used for generating probes are described in Table S4 in the supplementary material.

region-matched tissue samples that can be collected, and we were able to achieve this using chick embryos. Our analyses showed that the global transcriptomic makeup is highly uniform throughout the streak. This is likely to be the underlying mechanism for the lability in fate specification of streak mesoderm precursors.

Despite this overall uniformity, prominent regional differences in gene expression were observed in our analysis. These differences offer a 'bird's eye view' of the molecular preparation involved in initial mesoderm D/V patterning in all vertebrates. Although the streak is an embryological structure unique to amniotes, molecular mechanisms regulating mesoderm formation and D/V patterning are considered to be conserved in vertebrates. In birds and mammals, the majority of mesoderm cells emerge from the ectoderm by ingressing through the streak, which is positioned parallel to the future A/P axis of the embryo. D/V patterning of mesoderm precursors, which in fish and amphibians takes place perpendicular to the embryonic A/P axis, thus occurs in birds and mammals along the A/P axis of the streak. The developmental cause for this evolutionary shift in mesoderm internalization in amniotes is unclear. Despite these differences, several key events

in early vertebrate development are found to be regulated by conserved molecular mechanisms, including roles of the Wnt/β-catenin pathway in initial embryonic D/V symmetry breaking, of the FGF and TGFβ pathways in mesoderm precursor fate induction, and of the BMP pathway in ectoderm and mesoderm D/V patterning, the last of which is the focus of this work. The BMP pathway does not act alone in mesoderm D/V patterning, and it has been suggested that Wnt, FGF and TGFβ pathways act together with the BMP pathway in this process (see Introduction). Our data support these experimental observations. Moreover, we found that although core members exhibit little difference at the transcript level, regulatory molecules of each of these three major pathways clearly show positional differences, with both D→V and V→D gradients. This suggests that the activities of these pathways along the streak length are regulated subtly and dynamically, possibly to ensure a robust gradient that is resistant to perturbation. The existence of such a dynamic regulation is further supported by the fact that members of seven out of eight major kinase groups (TK, TKL, CAMK, AGC, STE, CMGC and atypical) known to be involved in the signaling of these three and other pathways,

DEVELOPMENT

exhibited graded expression in the streak (see Fig. S7 and Table S5 in the supplementary material). The exact nature of such a gradient is unclear. One possible scenario is that a signaling 'tug-of-war' converges via pathway cross-talk on intracellular BMP signaling activity (for instance, on the nuclear translocation and activity of SMAD1/5/9). Alternatively, the gradient could be a combinatorial signaling status in each of these pathways, conferring unique, yet still labile, identities to cells at different positions of the streak.

Transcription factors are integral components of receptor-mediated signaling, often as a readout of receptor signaling or as a convergence point of multiple signaling pathways. Although targets of transcriptional regulation may engage in feedback modulation of signaling pathways, it is generally considered that achieving a unique combinatorial code of transcription factors is an essential early step in cell-lineage/fate specification. It is therefore not surprising that a large number of transcription factors exhibit graded levels of expression along the streak. The 114 transcription factors we found are likely to be an underestimate, as many others with uniform transcript levels might exhibit protein or activity gradients that cannot be discerned in our current analysis. With the obvious limitation of the genome array coverage, our analysis presents a near complete list of differentially expressed transcription factors in the streak, making it possible to carry out computational modeling of transcriptional regulatory networks in the future for this well-conserved and well-defined developmental event. It also offers a valuable resource for using combinatorial transcription factor codes to direct and analyze embryonic stem/induced pluripotent stem (ES/iPS) cell differentiation toward specific mesoderm lineages in vitro.

As mentioned in the Introduction, the dissection of the streak into four equal regions was intended for quality and quantity controls of our analysis, and did not correlate with the known fate map of the streak. The most anterior A piece can be roughly termed the organizer region, although chordin, an organizer marker, is expressed in a narrower domain in chick embryos at this stage (Fig. 1A). According to fate map analyses, this piece will give rise to both axial and paraxial mesoderm tissues (Selleck and Stern, 1991) and therefore contains both what is conventionally viewed as the dorsal signaling center (axial mesoderm precursors) and what is considered as part of the non-axial mesoderm precursors (paraxial). It has been suggested that the non-axial mesoderm precursors do not contain much positional information. Our analysis, however, clearly indicates that there is fine-graded positional information along the entire length of the streak from both directions, separating the streak not only into axial and non-axial lineages, but also into multiple subregions of non-axial mesoderm. The complexity of these molecular gradients indicates that there could be more than five mesoderm lineages as traditionally defined. Some experimental evidence supports such a possibility. For instance, medial and lateral parts of the paraxial mesoderm have distinct origins from the streak (Iimura et al., 2007; Psychoyos and Stern, 1996a; Selleck and Stern, 1991), and the extraembryonic mesoderm contains hemogenic and non-hemogenic subregions (Nakazawa et al., 2006).

Genes exhibiting graded levels of expression along the streak may be involved in developmental regulation other than mesoderm D/V patterning. The ectoderm at HH4 is being patterned into neural and non-neural territories and it has been well documented that dorsal mesoderm precursors contain neural differentiation-inducing signals (Stern, 2005). The A region in our study also contains definitive endoderm and neural precursors (Chapman et al., 2007; Kimura et al., 2006; Lawson and Pedersen, 1987; Matsushita et al.,

2008; Robb and Tam, 2004; Selleck and Stern, 1991). Genes with a D→V or V→D gradient may not be restricted to the streak and in some cases this might reflect more the graded expression in neighboring ectoderm cells, as revealed by in situ analysis (Fig. 5). These aspects, although not the focus of this analysis, are nevertheless reflected in our transcriptomic datasets. For instance, some of the genes exhibiting the (1,0,0,0) type expression profile can be associated with endoderm (SOX17) or neural (SOX3) differentiation. These biases can in principle be eliminated by combining our current datasets with data from additional transcriptomic profiling of definitive endoderm or neural ectoderm. Preliminary results from one such experiment, by comparing streak tissue with its immediate lateral tissue, suggest that this can indeed be achieved (see Fig. S8 in the supplementary material; our unpublished data).

Aside from the issue of mixed germ layer lineages in the streak, streak mesoderm precursors are under an additional level of molecular regulation to control the timing of precursor cell ingression, which is manifested later on in development in the A/P positional difference of individual mesoderm lineages. A large number of Hox genes are expressed with a V→D profile. We think that this is likely to emphasize a role of Hox genes in mesoderm D/V patterning, as suggested, for instance, from studies of Hox gene control of early differentiation of hematopoietic cells (as derived from the ventral-most mesoderm in the streak) (Davidson et al., 2003; Davidson and Zon, 2006; Yue et al., 2009). However, it might also reflect a tight control of ingression timing among ventral mesoderm precursors by some of the Hox genes (Iimura and Pourquie, 2006). Taken together, it is very likely that in addition to their key roles in controlling mesoderm D/V patterning, the observed signaling and transcriptional regulatory gradients are also involved in other important developmental events occurring within and surrounding the streak.

Our analyses further revealed many novel genes with strong and differential levels of expression in the streak. They include genes of unknown molecular function, and those for which a function has been described in other systems. These genes might represent novel regulatory steps of signaling pathways known to be active in the streak or of novel pathways. For instance, our transcriptomic (see Table S1 in the supplementary material) and in situ (Fig. 5; data not shown) analyses revealed that a group of at least 24 glycosyltransferases (see Table S6 in the supplementary material) exhibits either a D→V (15) or V→D (9) profile. Sugar epitope modification of receptor and extracellular molecules has been reported to modulate signaling activities in many other systems (Haecker et al., 2008; Machingo et al., 2006; Tonoyama et al., 2009). Given that several major signaling pathways are key components in the regulation of positional identities in the streak, it is not surprising that we found these 24 enzymes exhibiting differential levels of expression in the streak, representing both early and late steps of O-linked and N-linked glycosyl modification. Other examples include large numbers of solute carrier family proteins (5 D→V and 15 V→D) and ion channel proteins (10 D→V and 3 V→D), and at least ten different cadherins (6 D→V and 4 V→D) that show graded expression along the streak (see Table S6 in the supplementary material). Solute carrier family proteins have so far not been associated with mesoderm D/V patterning, and some potassium channels, albeit different from those found in our analysis, have been reported to be involved in early left-right patterning (Levin et al., 2002). It should also be noted that the list of A region-specific genes (cluster pattern 1,0,0,0) not only covers all known organizer-specific genes, but

represents to our knowledge the most comprehensive and complete list of organizer genes, including many of unknown function (Fig. 5). Functional studies of these genes might provide novel insight into the molecular regulation of mesoderm D/V patterning.

In summary, we report in this work the first transcriptomic analysis of regional differences among vertebrate mesoderm precursors. Experimental observations of fate lability can be explained by the global ontological uniformity in expressed genes among mesoderm precursors fated for different mesoderm lineages, whereas multi-pronged signaling and transcriptional gradients suggest complex molecular mechanisms for generating and maintaining a robust mesoderm precursor identity gradient. Finally, our datasets provide a valuable resource for in vitro mesoderm lineage differentiation from ES/iPS cells in translational research fields.

Competing interests statement
The authors declare no competing financial interests.

Supplementary material
Supplementary material for this article is available at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.053462/-/DC1

References
Abercrombie, M. (1950). The effects of antero-posterior reversal of lengths of the primitive streak in the chick. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **234**, 317-338.

Abercrombie, M. and Waddington, C. H. (1937). The behaviour of grafts of primitive streak beneath the primitive streak of the chick. *J. Exp. Biol.* **14**, 319-334.

Abercrombie, M. and Bellairs, R. (1954). The effects in chick blastoderms of replacing the primitive node by a graft of posterior primitive streak. *J. Embryol. Exp. Morphol.* **2**, 55-72.

Bellairs, R. (1986). The primitive streak. *Anat. Embryol. (Berl.)* **174**, 1-14.

Chapman, S. C., Matsumoto, K., Cai, Q. and Schoenwolf, G. C. (2007). Specification of germ layer identity in the chick gastrula. *BMC Dev. Biol.* **7**, 91.

Dale, L., Howes, G., Price, B. M. and Smith, J. C. (1992). Bone morphogenetic protein 4, a ventralizing factor in early Xenopus development. *Development* **115**, 573-585.

Davidson, A. J. and Zon, L. I. (2006). The caudal-related homeobox genes cdx1a and cdx4 act redundantly to regulate hox gene expression and the formation of putative hematopoietic stem cells during zebrafish embryogenesis. *Dev. Biol.* **292**, 506-518.

Davidson, A. J., Ernst, P., Wang, Y., Dekens, M. P., Kingsley, P. D., Palis, J., Korsmeyer, S. J., Daley, G. Q. and Zon, L. I. (2003). cdx4 mutants fail to specify blood progenitors and can be rescued by multiple hox genes. *Nature* **425**, 300-306.

Dosch, R., Gawantka, V., Delius, H., Blumenstock, C. and Niehrs, C. (1997). Bmp-4 acts as a morphogen in dorsoventral mesoderm patterning in Xenopus. *Development* **124**, 2325-2334.

Era, T., Izumi, N., Hayashi, M., Tada, S., Nishikawa, S. and Nishikawa, S. (2008). Multiple mesoderm subsets give rise to endothelial cells, whereas hematopoietic cells are differentiated only from a restricted subset in embryonic stem cell differentiation culture. *Stem Cells* **26**, 401-411.

Furthauer, M., Thisse, C. and Thisse, B. (1997). A role for FGF-8 in the dorsoventral patterning of the zebrafish gastrula. *Development* **124**, 4253-4264.

Furthauer, M., Van Celst, J., Thisse, C. and Thisse, B. (2004). Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. *Development* **131**, 2853-2864.

Garcia-Martinez, V. and Schoenwolf, G. C. (1992). Positional control of mesoderm movement and fate during avian gastrulation and neurulation. *Dev. Dyn.* **193**, 249-256.

Garcia-Martinez, V., Alvarez, I. S. and Schoenwolf, G. C. (1993). Locations of the ectodermal and nonectodermal subdivisions of the epiblast at stages 3 and 4 of avian gastrulation and neurulation. *J. Exp. Zool.* **267**, 431-446.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80.

Grabowski, C. T. (1956). The effects of the excision of Hensen's node on the early development of the chick. *J. Exp. Zool.* **133**, 301-343.

Haecker, A., Bergman, M., Neupert, C., Moussian, B., Luschnig, S., Aebi, M. and Mannervik, M. (2008). Wollknaeul is required for embryo patterning and encodes the Drosophila ALG5 UDP-glucose:dolichyl-phosphate glucosyltransferase. *Development* **135**, 1745-1749.

Harland, R. (2004). Dorsoventral patterning of the mesoderm. In *Gastrulation* (ed. C. D. Stern), pp. 373-388. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Iimura, T. and Pourquie, O. (2006). Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature* **442**, 568-571.

Iimura, T., Yang, X., Weijer, C. J. and Pourquie, O. (2007). Dual mode of paraxial mesoderm formation during chick gastrulation. *Proc. Natl. Acad. Sci. USA* **104**, 2744-2749.

Imai, Y., Gates, M. A., Melby, A. E., Kimelman, D., Schier, A. F. and Talbot, W. S. (2001). The homeobox genes vox and vent are redundant repressors of dorsal fates in zebrafish. *Development* **128**, 2407-2420.

Inagaki, T. and Schoenwolf, G. C. (1993). Axis development in avian embryos: the ability of Hensen's node to self-differentiate, as analyzed with heterochronic grafting experiments. *Anat. Embryol.* **188**, 1-11.

Inagaki, T., Garcia-Martinez, V. and Schoenwolf, G. C. (1993). Regulative ability of the prospective cardiogenic and vasculogenic areas of the primitive streak during avian gastrulation. *Dev. Dyn.* **197**, 57-68.

Inomata, H., Haraguchi, T. and Sasai, Y. (2008). Robust stability of the embryonic axial pattern requires a secreted scaffold for chordin degradation. *Cell* **134**, 854-865.

James, R. G. and Schultheiss, T. M. (2003). Patterning of the avian intermediate mesoderm by lateral plate and axial tissues. *Dev. Biol.* **253**, 109-124.

Jones, C. M. and Smith, J. C. (1998). Establishment of a BMP-4 morphogen gradient by long-range inhibition. *Dev. Biol.* **194**, 12-17.

Jones, C. M., Lyons, K. M., Lapan, P. M., Wright, C. V. and Hogan, B. L. (1992). DVR-4 (bone morphogenetic protein-4) as a posterior-ventralizing factor in Xenopus mesoderm induction. *Development* **115**, 639-647.

Joubin, K. and Stern, C. D. (1999). Molecular interactions continuously define the organizer during the cell movements of gastrulation. *Cell* **98**, 559-571.

Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30.

Keren, A., Keren-Politansky, A. and Bengal, E. (2008). A p38 MAPK-CREB pathway functions to pattern mesoderm in Xenopus. *Dev. Biol.* **322**, 86-94.

Khokha, M. K., Yeh, J., Grammer, T. C. and Harland, R. M. (2005). Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Dev. Cell* **8**, 401-411.

Kimura, W., Yasugi, S., Stern, C. D. and Fukuda, K. (2006). Fate and plasticity of the endoderm in the early chick embryo. *Dev. Biol.* **289**, 283-295.

Lawson, K. A. and Pedersen, R. A. (1987). Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. *Development* **101**, 627-652.

Lekven, A. C., Thorpe, C. J., Waxman, J. S. and Moon, R. T. (2001). Zebrafish wnt8 encodes two wnt8 proteins on a bicistronic transcript and is required for mesoderm and neurectoderm patterning. *Dev. Cell* **1**, 103-114.

Levin, M., Thorlin, T., Robinson, K. R., Nogi, T. and Mercola, M. (2002). Asymmetries in H$^+$/K$^+$-ATPase and cell membrane potentials comprise a very early step in left-right patterning. *Cell* **111**, 77-89.

Lopez-Sanchez, C., Garcia-Martinez, V. and Schoenwolf, G. C. (2001). Localization of cells of the prospective neural plate, heart and somites within the primitive streak and epiblast of avian embryos at intermediate primitive-streak stages. *Cells Tissues Organs* **169**, 334-346.

Machingo, Q. J., Fritz, A. and Shur, B. D. (2006). A beta1,4-galactosyltransferase is required for Bmp2-dependent patterning of the dorsoventral axis during zebrafish embryogenesis. *Development* **133**, 2233-2241.

Marom, K., Fainsod, A. and Steinbeisser, H. (1999). Patterning of the mesoderm involves several threshold responses to BMP-4 and Xwnt-8. *Mech. Dev.* **87**, 33-44.

Matsushita, S., Urase, K., Komatsu, A., Scotting, P. J., Kuroiwa, A. and Yasugi, S. (2008). Foregut endoderm is specified early in avian development through signal(s) emanating from Hensen's node or its derivatives. *Mech. Dev.* **125**, 377-395.

Mikawa, T., Poh, A. M., Kelly, K. A., Ishii, Y. and Reese, D. E. (2004). Induction and patterning of the primitive streak, an organizing center of gastrulation in the amniote. *Dev. Dyn.* **229**, 422-432.

Mitiku, N. and Baker, J. C. (2007). Genomic analysis of gastrulation and organogenesis in the mouse. *Dev. Cell* **13**, 897-907.

Munoz-Sanjuan, I. and Brivanlou, A. H. (2004). Modulation of BMP signaling during vertebrate gastrulation. In *Gastrulation* (ed. C. D. Stern), pp. 475-498. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Muraoka, O., Shimizu, T., Yabe, T., Nojima, H., Bae, Y. K., Hashimoto, H. and Hibi, M. (2006). Sizzled controls dorso-ventral polarity by repressing cleavage of the Chordin protein. *Nat. Cell Biol.* **8**, 329-338.

Murry, C. E. and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661-680.

Nakaya, Y. and Sheng, G. (2008). Epithelial to mesenchymal transition during gastrulation: an embryological view. *Dev. Growth Differ.* **50**, 755-766.

Nakaya, Y. and Sheng, G. (2009). An amicable separation: chick's way of doing EMT. *Cell Adh. Migr.* **3**, 160-163.

Nakaya, Y., Sukowati, E. W., Wu, Y. and Sheng, G. (2008). RhoA and microtubule dynamics control cell-basement membrane interaction in EMT during gastrulation. *Nat. Cell Biol.* **10**, 765-775.

Nakazawa, F., Nagai, H., Shin, M. and Sheng, G. (2006). Negative regulation of primitive hematopoiesis by the FGF signaling pathway. *Blood* **108**, 3335-3343.

Nicolet, G. (1970). An autoradiographic study of the presumptive fate of the primitive streak in chick embryos. *J. Embryol. Exp. Morphol.* **23**, 70-108.

Piccolo, S., Sasai, Y., Lu, B. and De Robertis, E. M. (1996). Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell* **86**, 589-598.

Psychoyos, D. and Stern, C. D. (1996a). Fates and migratory routes of primitive streak cells in the chick embryo. *Development* **122**, 1523-1534.

Psychoyos, D. and Stern, C. D. (1996b). Restoration of the organizer after radical ablation of Hensen's node and the anterior primitive streak in the chick embryo. *Development* **122**, 3263-3273.

Robb, L. and Tam, P. P. (2004). Gastrula organiser and embryonic patterning in the mouse. *Semin. Cell Dev. Biol.* **15**, 543-554.

Schmid, B., Furthauer, M., Connors, S. A., Trout, J., Thisse, B., Thisse, C. and Mullins, M. C. (2000). Equivalent genetic roles for bmp7/snailhouse and bmp2b/swirl in dorsoventral pattern formation. *Development* **127**, 957-967.

Selleck, M. A. and Stern, C. D. (1991). Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development* **112**, 615-626.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3.

Stern, C. D. (2005). Neural induction: old problem, new findings, yet more questions. *Development* **132**, 2007-2021.

Streit, A. and Stern, C. D. (1999). Mesoderm patterning and somite formation during node regression: differential effects of chordin and noggin. *Mech. Dev.* **85**, 85-96.

Streit, A. and Stern, C. D. (2008). Operations on primitive streak stage avian embryos. *Methods Cell Biol.* **87**, 3-17.

Takeuchi, J. K. and Bruneau, B. G. (2009). Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459**, 708-711.

Tonoyama, Y., Anzai, D., Ikeda, A., Kakuda, S., Kinoshita, M., Kawasaki, T. and Oka, S. (2009). Essential role of beta-1,4-galactosyltransferase 2 during medaka (Oryzias latipes) gastrulation. *Mech. Dev.* **126**, 580-594.

Vijayaragavan, K., Szabo, E., Bosse, M., Ramos-Mejia, V., Moon, R. T. and Bhatia, M. (2009). Noncanonical Wnt signaling orchestrates early developmental events toward hematopoietic cell fate from human embryonic stem cells. *Cell Stem Cell* **4**, 248-262.

Yabe, T., Shimizu, T., Muraoka, O., Bae, Y. K., Hirata, T., Nojima, H., Kawakami, A., Hirano, T. and Hibi, M. (2003). Ogon/Secreted Frizzled functions as a negative feedback regulator of Bmp signaling. *Development* **130**, 2705-2716.

Yue, R., Kang, J., Zhao, C., Hu, W., Tang, Y., Liu, X. and Pei, G. (2009). Beta-arrestin1 regulates zebrafish hematopoiesis through binding to YY1 and relieving polycomb group repression. *Cell* **139**, 535-546.