Supplementary Information - Detailed Materials and Methods

Chick embryo cultures and tissue dissection

Fertile hens' eggs (Henry Stewart & Co. and Winter Farm, UK) were incubated at 38°C in a humidified chamber to the desired stages. Embryos were staged according to Hamburger and Hamilton (HH; 1951) (1) or Eyal-Giladi and Kochav (EGK) (2) for pre-primitive streak (pre-streak) stages. Embryos were cultured using a modified New culture method (3, 4). For neural and pre-placodal region induction assays, Hensen's node was isolated from HH4^{-} chick embryos and lateral head mesoderm underlying the pre-placodal region (5, 6) was dissected from HH5-6 embryos. Hypoblast grafts were isolated from EGK XII embryos (7). All grafts were placed into the inner extraembryonic regions of HH4- stage hosts (8-10) and embryos were cultured for different times. Grafted tissues were removed before collecting the adjacent epiblast for further analysis by RNAseq (50 tissues) or NanoString (5-7 tissues). Equivalent tissues were collected from the contralateral side of each embryo to serve as non-induced controls.

For RNAseq the following tissues were dissected from normal embryos: central and anterior-lateral epiblast from EGK XII-XIII, neural plate from HH6-7 HH6, anterior and posterior pre-placodal region from HH6-7 and non-neural ectoderm from HH6-7 embryos. For ectodermal tissues from embryos older than HH5, the underlying mesoderm and endoderm were removed using mild trypsin or dispase treatment when required.

Explant cultures

For explant cultures, epiblast tissue was dissected from the middle and anterior-lateral regions of EGK XII embryos. Tissue was kept on ice until being embedded in rat tail collagen and cultured for 40 hours or 6 days as previously described (11, 12). Fisher's exact test was used to estimate statistical significance of gene expression differences between medial and lateral explants.

RNA isolation, library preparation and RNA sequencing

RNA sequencing of node induced (5 hours post-graft) and control area opaca epiblast tissue from the same embryos was conducted by Edinburgh Genomics. First, 50 samples from each condition were pooled in Trizol and total RNA extracted. RNA quality was assessed using the Agilent 2100 Bioanalyzer. All samples

had an RNA Integrity Number (RIN) between 9-10. From these, labeled RNA libraries were constructed using the Illumina® Truseq mRNA library preparation kit. RNA libraries were sequenced over 2 lanes via 100-cycle, paired-end sequencing using the Illumina® HiSeq 2000 system.

Pre-processing of RNA-Seq samples, Quality Control and alignment to the Chicken genome

All raw files were first converted into Sanger FASTQ format, using FASTQ Groomer (Galaxy Version 1.0.4). To ensure that only paired-end reads were retained for downstream analyses, FASTQ joiner (Galaxy Version 2.0.1) was used to join together only those reads that were tagged with the same sequence identifiers in both paired-end FASTQ files. Next, FASTQ splitter (Galaxy Version 1.0.0) was used to split the joined paired-end files back into two separate files. Quality control of FASTQ files was performed using FastQC version 0.10.1 (13); bases with a Phred score of less than 33 were removed using the tool, FASTX trimmer, from version 0.0.13 of the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Paired-end reads passing quality control were subsequently aligned to the Galgal4.71 assembly of the Chicken genome, guided by the Ensembl annotation (Galgal4.71.gtf), using TopHat2 version-2.0.7 (14). The following settings were used to facilitate genome-wide alignment of reads: -N 4, --read-edit dist 6, -m 1 --max-insertion-length 5, --max-deletion-length 5, -p 8, --no-discordant, --no-mixed --microexon-search, --library-type fr-unstranded, --mate-std-dev 51.

Differential Gene Expression Analysis

To facilitate quantification of aligned reads and subsequent differential gene expression analysis across conditions, the R Bioconductor package, easyRNAseq (version 2.1.0), was used to transform aligned reads into count data. The total number of supporting reads for each sample were formatted into matrices where the value in the i-th row and the j-th column of each matrix represented the total number of reads successfully mapped to the junctions of the host gene, i, in sample, j. Differential gene expression analysis across conditions was then performed using version 1.12.1 of the R Bioconductor package, DESeq (15). Genes with an absolute normalized read count of > 150 were considered to be expressed. Of these genes, those passing a fold-change cut-off of +/- 1.5 were used to select candidates for experimental validation.

In Situ Hybridization

In situ hybridization using DIG-labeled riboprobes was performed as previously described, on whole-mount embryos or cultured explants (16, 17). When probes were not already available, they were generated from the chick EST collection (18) (Supplementary table 1). A 1.5KB probe template for PHF15 was cloned from chick cDNA using the following primers: Forward 5'-ACGGAAGTACTCGATCAGCAGTG-3' and Reverse 5'-CAGGTGGAAGATCTGCTCCTGG-3'.

NanoString nCounter

Node-, hypoblast or and head mesoderm induced tissue together with time-matched, non-induced controls were analysed for changes in gene expression using the NanoString nCounter® Analysis System (Life Sciences) using a custom made probe set. Each experiment was performed three times on independent occasions; 6-8 tissues per sample were lysed in lysis buffer (Ambion) and total RNA was hybridized with capture and reporter probes according to the nCounter Gene Expression Assay Manual. Data were normalized according to the manufacturer's instructions. Differential expression was calculated by comparing transcript counts between experimental and control conditions. Transcripts with a raw count of less than 5 were excluded from further analysis. Fold change thresholds of ≥1.2 or ≤0.75 were used to define transcripts as up- or downregulated, respectively, with a p-value of 0.05 or less (two-tailed Type 2 T-Test).

Network inference

Transcripts for individual samples were assembled with Cufflinks (v2.1.1) (19) for both Ensembl (Galgal4.71.gtf) and Refseq annotations, and passed to Cuffdiff (v2.1.1) to obtain normalized FPKM. All transcription factors with an FPKM $\geq$10 were considered expressed and used for predicting a gene regulatory network using Genie3. This method is based on a random forest algorithm (20) and ranks predicted interactions using Importance Measure (IM). Interactions with IM $\geq$0.005 were extracted and visualized with Cytoscape version 3.2.0 (21). Clusters and sub-clusters within the large network were identified using the GLay plugin in Cytoscape (22) based on the community clustering algorithm (23).

References for Supplementary Information

1. Hamburger V & Hamilton HL (1951) A series of normal stages in the development of the chick embryo. *Journal of Morphology* 88(1):49-92.
2. Eyal-Giladi H & Kochav S (1976) From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Developmental Biology* 49(2):321-337.
3. New DAT (1955) A new technique for the cultivation of the chick embryo in vitro. *J. Embryol. Exp. Morphol.* 3:326-331.
4. Stern CD & Ireland GW (1981) An integrated experimental study of endoderm formation in avian embryos. *Anat. Embryol* 163(3):245-263.
5. Litsiou A, Hanson S, & Streit A (2005) A balance of FGF, BMP and WNT signalling positions the future placode territory in the head. *Development* 132(18):4051-4062.
6. Hintze M (2017) TBD. *TBD* TBD:xx-xx.
7. Albazerchi A & Stern CD (2007) A role for the hypoblast (AVE) in the initiation of neural induction, independent of its ability to position the primitive streak. *Dev Biol* 301(2):489-503.
8. Stern CD (2008) Grafting Hensen's node. *Methods Mol. Biol* 461:265-276.
9. Streit A & Stern CD (2008) Operations on primitive streak stage avian embryos. *Methods Cell Biol* 87:3-17.
10. Storey KG, Crossley JM, De Robertis EM, Norris WE, & Stern CD (1992) Neural induction and regionalisation in the chick embryo. *Development* 114(3):729-741.
11. Streit A & Stern CD (1999) Establishment and maintenance of the border of the neural plate in the chick: involvement of FGF and BMP activity. *Mech Dev* 82(1-2):51-66.
12. Bailey AP, Bhattacharyya S, Bronner-Fraser M, & Streit A (2006) Lens specification is the ground state of all sensory placodes, from which FGF promotes olfactory identity. *Dev Cell* 11(4):505-517.
13. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
14. Kim D*, et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
15. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
16. Stern CD (1998) Detection of multiple gene products simultaneously by in situ hybridization and immunohistochemistry in whole mounts of avian embryos. *Curr. Top. Dev. Biol* 36:223-243.
17. Streit A & Stern CD (2001) Combined whole-mount in situ hybridization and immunohistochemistry in avian embryos. *Methods* 23(4):339-344.
18. Boardman PE*, et al.* (2002) A comprehensive collection of chicken cDNAs. *Curr Biol* 12(22):1965-1969.
19. Trapnell C*, et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28(5):511-515.
20. Huynh-Thu VA, Irrthum A, Wehenkel L, & Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5(9).
21. Shannon P*, et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.
22. Su G, Kuchinsky A, Morris JH, States DJ, & Meng F (2010) GLay: community structure analysis of biological networks. *Bioinformatics* 26(24):3135-3137.
23. Newman MEJ & Girvan M (2004) Finding and evaluating community structure in networks. *Physical Review E* 69:026113.

Supplementary Figure Legends

Figure S1. lHM- and node-regulated genes as determined by NanoString. Volcano plots of $\log_{10}$ P-values against $\log_{10}$ fold-change (induced/uninduced) for lHM (left) and node (right) grafts at 3h. Black dots indicate significant genes, <0.05. Blue dots indicate genes in common for lHM and node grafts, vertical dashed line indicates 0.05 P-value, horizontal dashed lines represent 1.5-fold change.

Figure S2: Expression patterns of selected transcription factors at pre-primitive streak (EGK XII-XIII), primitive streak (HH3-4) and neural plate stages (HH5-7).

Figure S3: Defining neural plate, anterior and posterior PPR enriched genes. Pairwise comparison of different RNAseq data sets was performed using Dseq; results are displayed as Volcano plots (A-C, F-I) with vertical lines labelling fold change -1.5 and 1.5, and horizontal line indicating p-value -0.05. Diagrams show neural plate stage embryos with relevant tissues colour coded, and the reference tissue in grey. A. Comparison of neural plate (NP, orange) and non-neural, non-placodal ectoderm (NNE, grey). B. Comparison of neural plate (NP, green) and anterior PPR (aPPR, grey). C. Comparison of neural plate (NP, yellow) and posterior PPR (pPPR, grey). D, E. Venn diagrams showing neural plate enriched genes (D) and transcription factors (E) as overlap between the three comparisons in A-C.

F. Comparison of anterior PPR genes (green) to non-neural, non-placodal ectoderm (NNE, grey). G. Comparison of anterior PPR genes (yellow) to neural plate (NP, grey). J, K. Venn diagrams showing anterior PPR enriched genes as overlap between the comparison in F and G. H. Comparison of posterior PPR genes (purple) to non-neural, non-placodal ectoderm (NNE, grey). I. Comparison of posterior PPR genes (red) to neural plate (NP, grey). L, M. Venn diagrams showing anterior PPR enriched gene s(L) and transcription factors (M) as overlap between the comparison in H and I.

Figure S4: Defining *Sox3*-like genes. Pairwise comparison of different RNAseq data sets was performed using Dseq; results are displayed as Volcano plots (A-C, F-G) with vertical lines labelling fold change -1.5 and 1.5, and horizontal line indicating p-value -0.05. A-C. Diagrams show neural plate stage embryos with

relevant tissues colour coded, and the reference tissue in grey. Anterior PPR (A; aPPR turquois), posterior PPR (B; pPPR purple) and neural plate (C; NP pink) were compared to non-neural, non-placodal ectoderm (NNE, grey). D, E. Venn diagrams showing genes (D) and transcription factors (E) in common between neural plate, aPPR and pPPR (NAP) as overlap between the three comparisons in A-C. F. To identify factors present in the entire pre-streak epiblast (purple), genes enriched exclusively in the anterior pre-streak epiblast were removed. G. Comparison between epiblast exposed to Hensen's node for 5 hours and untreated contra-lateral epiblast reveals node-induced genes (pink). H. I. Most node-induced genes (H) and transcription factors (I) are also expressed in pre-streak epiblast as shown by comparing results from F and G (PSI). J, K. Sox3 like genes (J) and transcription factors (K) are defined as those in common between NAP and PSI.

Figure S5: A network of co-expressed transcription factors defines a pre-neural state. **A.** GENIE3-generated network of 805 transcription factors (TFs) identified from RNAseq visualized with Cytoscape. The size and colour (red to green) of the nodes reflects predicted out-degree, i.e. the number of putative target genes regulated by each transcription factor. Nodes outlined in blue correspond to *Sox3*-like TFs, in green to *Six1*-like TFs, and in red to *Sox2*-like TFs. **B-D.** The network was further dissected by community clustering. This analysis reveals three large subnetworks: cluster1 (C1) enriched with *Sox3*- and *Six1*-like factors; cluster2 (C2) with no enrichment of a specific TF category; cluster3 (C3) enriched with *Sox2*-like TFs. Further sub-clustering of C1 separates *Sox3*-like genes (C1B) from *Six1*-like TFs (C1C); C2 can be further divided into C2A, C2B, C2C and C2D; sub-clustering C3 segregates *Sox2*-like TFs (C3A) from others.

Figure S6: Analysis of network sub-clusters. A. Diagram showing the workflow to subcluster the TF network shown in Fig. 2 using community clustering. B. Hypergeometric analysis of the genes included in each cluster C1-C3 reveals the distribution of Sox3- (blue), Six1- (green) and Sox2-like (red) TFs in each cluster. P-value was calculated using GenePro and –log10(P-value) > 2 indicates significant enrichment of each TF group (horizontal line). C. Percentage of Sox3-, Six1- and Sox2-like TFs in each cluster. Initially, Sox2-like genes (C3) segregate from Sox3/Six1-like genes (C1); the latter separate into two clusters (C1B: Sox3; C1C: Six1) after further dissection. This suggests a higher degree of communality between the PPR and pre-streak epiblast than between each tissue and the neural plate.

Figure S7: Specification assay for the early epiblast. Explants were obtained from medial (M) and lateral (L) regions of pre-primitive streak epiblast and placed in culture. Expression of various markers was assessed after 40h. At HH5-6, *Dlx5*, *ERNI* and *Gata3* label the neural plate border, *Msx1*, *Pax7* and *Snail-2* (HH8) label the neural crest, *Eya2* and *Six4* Sox1 are expressed in the PPR, while *Sox3* is present in both the neural plate and PPR. *Sox2* and *Sox1* are neural markers and *Gata2* and *Tbx6* label mesoderm. After 40h explants express neural plate border markers. The number in each panel indicates the proportion of explants expressing the marker.

Figure S8. Comparison of genes induced by the hypoblast, LHM and node. A. Hypoblast was grafted next to area opaca epiblast; after 5 hours the epiblast was dissected together with the contralateral control side and both were processed for NanoString analysis. The Volcano plot shows $\log_{10}$ p-values versus $\log_{10}$ fold change (induced/non-induced). Blue dots indicate significant genes with a p-value < 0.05. B. Venn diagram showing the number of genes commonly upregulated in response to the lateral head mesoderm (LHM; yellow), node (pink) and hypoblast (green).

Figure S9. Cell identity prediction using ESCAPE database. Comparison of the transcription factor profiles that characterise different tissues or co-expression clusters as determined by RNAseq with the profiles of various cell lines. The profile of the *Sox3*-like TF cluster (blue) and of TFs present in pre-streak epiblast and induced by the node (PSI, red) is similar to the gene expression profile of ES cell lines.