University of South Bohemia, Faculty of Science
Department of Molecular Biology

BACHELOR THESIS

Cloning Candidate Novel Cell-Fate Genes (Pre-Implantation Mouse Development)

Joan Elorm Abla Ahiable

Biological Chemistry

Supervisor: Alexander W. Bruce, Ph.D.

Department of Molecular Biology

Laboratory of Developmental Biology and Genomics

České Budějovice, December 2012

Ahiable JEA, 2012. Cloning candidate novel cell-fate genes (pre-implantation mouse development). BSc. Thesis (in English), 66 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

The aim of this study was to clone novel cell-fate genes: *Socs3*, *Gata1*, *Zfpm1*, *Zbtb32*, *Klf4*, *Foxa1* and *Dgcr8* into appropriate plasmid vectors for generation of RNA species to be microinjected into mouse embryo cell clones.

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography.

I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages.
Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defense in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

Date .....

Signature .....

Abstract

Generating experimentally useful recombinant DNA constructs that permit the clonal over-expression or knock-down of specific candidate cell-fate influencing genes in the pre-implantation mouse embryo is an essential pre-requisite to studying their role in early mouse development. Through experiments, our three aims have been successfully achieved: Firstly, we cloned novel cell fate influencing candidate gene´s protein coding sequences- *Gata1*, *Socs3* and *Zbtb32*- into the appropriate plasmid vector, pRN3P, to permit *in vitro* generation of mRNA species that can be microinjected. Originally, this was planned for four other additional candidate cell fate related genes: *Zfpm1*, *Klf4*, *Foxa1* and *Dgcr8* although not realised. Also, specific DNA sequence were added at the 5´end of the target gene protein coding region inserts that encoded a HA-tag at the N-terminus of recombinant proteins that would allow the discrimination of the resultant recombinant proteins from the endogenous ones after microinjection. These constructs can now be used in functional studies in the embryo. Furthermore, T_7 linked PCR products for some candidate genes were generated. These were then used to synthesize long dsRNA species by IVT. These long dsRNAs can also be microinjected into embryo cells to elicit gene expression knock-down in functional studies. Finally, we synthesise mRNA for the fluorescent proteins markers, GAP43-RFP and GAP43-GFP, EGFP and DsRed by *in vitro* transcription using pre-existing pRN3P-fluorescent gene clones. These are necessary for functional studies as they are co-injected into cells of the embryo with mRNA or long dsRNA for candidate cell fate influencing genes

Acknowledgements

My foremost and sincere appreciation goes to my supervisor, Alexander Bruce Ph.D., for his help, patience and time spent in the course of his supervision making this thesis a success. His leadership, confidence and encouragements have brought me so far.

I also wish to acknowledge the help received from RNDR Alena Krejci and the whole team of the laboratory of development biology and genomics during my laboratory work.

Moreover, l would like to thank my program coordinators, CSc. and a. Univ.-Prof. Mag. Dr. Nobert Müller (Johannes Kepler University Linz, Austria) and Prof. RNDr. Libor Grubhoffer (University of South Bohemia in České Budějovice, Czech) for making this Biochemistry cross border curriculum a unique one. They have been ever ready in assisting us and this makes our discipline a smooth one.

All this would not have been possible, if my family hadn´t been my backbone. Eric, my dearest husband, I so much treasure your endless love, support and encouragements. Honey, I love you forever and always. On the same note, I express a heart full of gratitude to my parents and sisters: Olivia, Hannes, Laura and Tamara; not forgetting Sedem and Patricia. Thank you all for being there for me. I love you all!

Finally, l would like to say thank You, almighty God for seeing me through it all.

Table of Contents

1 Introduction

1.1 Pre-implantation stages of mouse embryo development and the derived cell lineages

An amazing period of mouse embryo development is represented by the first 4.5 days. During this phase the development of a peri-implantation blastocyst consisting of three distinct cell lineages will ensue from a single fertilised oocyte (see Figure 1.1).

Trophectoderm (TE), the first lineage, comprises of embryonic tissues that will give rise to the placenta and also provide the developing embryo with positional cues and support in the uterus[1]. At the blastocyst stage it is occupied by outer cells that have initiated differentiation and engendered gene expression vital for trophectoderm function and implantation.

The second population, the primitive endoderm (PE), is a second extra-embryonic and differentiating tissue which lies in immediate contact with the blastocyst cavity[1] and lines the surface of the inner cell mass (ICM)[2]. The PE will give rise to the endodermal membranes of the yolk sac later in development

The pluripotent epiblast (EPI), which is the third lineage, represents the progenitor cells for every succeeding foetal tissue. It consists of cells which are located deep within the inner cell mass and are not in contact with the blastocyst cavity.[3-6]

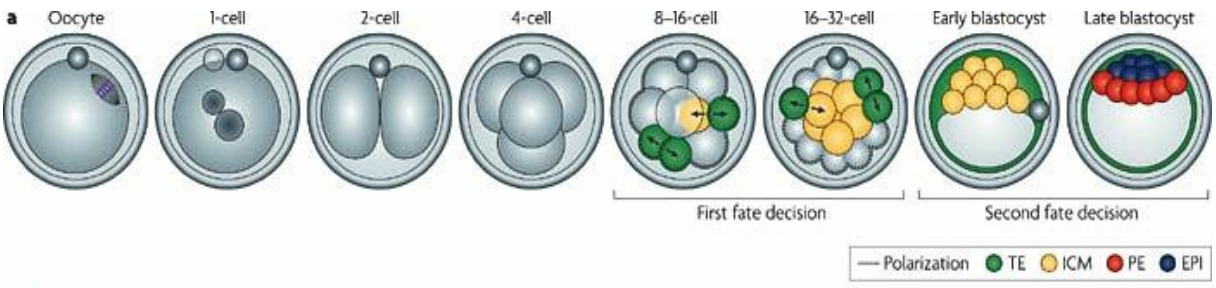


Figure 1.1: Pre-implantation stages of mouse development. Inner cell mass (ICM) divides asymmetrically starting at the 8-16-cell stage transition. In the first cell lineage, the outer cells become trophectoderm (TE). The second cell lineage comprises of the development of primitive endoderm (PE) at the surface of the ICM and the development of the epiblast (EPI) in the deeper layer. The late blastocyst shows the relative location of all the three cell lineages: TE, PE and EPI. Image taken from Zernicka-Goetz *et al.,*(2009)

1.2 Spatial and temporal regulation

The question however is how this single zygote undergoes these transitions of sequential cell divisions that eventually yield the peri-implantation blastocyst and the three described cell lineages? The development of the TE and ICM lineages (PE and EPI) are subject to spatial and temporal regulations during pre-implantation development and these can be conceptualised as two different stages: the first and the second cell fate decisions [6, 7].

1.2.1 The first cell-fate decisions

Asymmetric divisions of 8-cell and outer 16-cell stage blastomeres (a specific term for a pre-implantation embryo stage cell) that assign daughter cells to one relative inner and one outer position (a differentiative division) within the embryo[6] initiate the first cell decision in an early mouse development - expanded below. Alternatively, outer cells may divide symmetrically in that case both daughter cells remain outside (a conservative division)[6, 7].

The (a)symmetric divisions occur in two waves and these are at the 8- to 16-cell stage transition and the 16- to 32-cell stage respectively. Consequently the deposition of cells into different relative embryo locations leads them to become committed to particular developmental fates. Thus they become specified. This includes interaction between various transcription factors that specify TE and ICM cell fate[7-10].

The outside cells become actively committed to the TE through the action of the transcription factor Cdx2 (see Figure 1.2), that also acts to directly antagonise pluripotency related transcription (circuitry) factor associated with ICM[1, 7, 11, 12]. Whereas ICM cells remain in a pluripotent and undifferentiated state largely by the action of the transcription factors Oct4 and Nanog. Other factors, moreover, contribute robustness to this system evidencing the importance of temporal dynamics in transcription factor gene expression to lineage segregation. Tead4, a transcription factor expressing from the late 2-cell, activates zygotic*Cdx2*gene expression in only outer cells[1, 6, 13, 14]. Thus higher level of Cdx2 expression in these cells is more likely to promote their differentiation to TE.

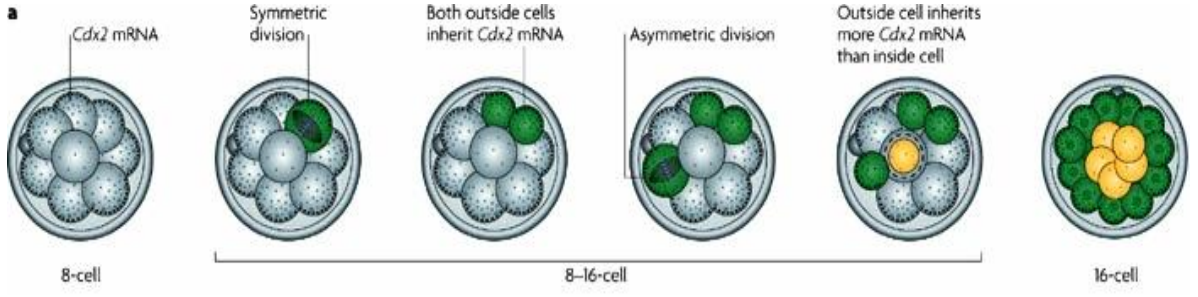


Figure 1.2: Transcriptional circuits in the first cell-fate decision. The symmetric divisions of 8-cell and outer 16-cell blastomeres assign daughter cells to inner and outer position. In outer cells Cdx2 expression is augmented leading to a differentiating TE cell-fate, whereas suppression of Cdx2 expression in inner cells couple with expression of pluripotency related transcription factor genes promotes formation of the pluripotent ICM. Image taken from Zernicka-Goetz *et al.,*(2009)

Although the expression of Tead4 is found in all cells, its transcriptional activation properties are restricted to only outside cells[1], hence explaining why only the outer cells express Cdx2. Examples of similar regulatory transcription factor genes important for controlling TE fate are Gata3 and Eomes[6, 14-16]. Indeed, Gata3 expression, like that of Cdx2, can also be regulated by Tead4[1, 16]. It has also been proposed that Cdx2 transcriptional expression is directly regulated by Gata3[1, 15]. Consequently, a combination of both temporal and spatial regulation of Cdx2 expression but a network of TE promoting transcription factors directly supports the appropriate differentiation of outer cells of the embryo towards TE[6].

Moreover, two hypotheses have been brought forward so as to understand how both inside and outside cells might begin to differentiate from each other: The first stresses the importance of a blastomere's relative position within the embryo and micro-environmental difference that this engenders in respect to either differentiation (TE) or retention of pluripotency (ICM) and is referred to as the "inside-outside" hypothesis[3, 7]. The second is derived from the discovery that the embryo's blastomeres become polarized along their apical-basal axis at the 8-cell stage[17]. The differentially inherited cellular compositions of the inside and outside cells that arise from the asymmetric cell division of such polarized blastomeres are proposed to provide them with instructional information regarding their fate (TE for outside cells and ICM for inner cells)[7]. Hence, according to the second hypothesis, it is not the cell position but instead cell division that originates the generation

of diverse traits of the two cell types. Furthermore, it is known that the molecules that regulate and are important for polarization of mouse blastomeres at the 8-cell stage are the molecules that regulate cell polarity in variety of systems (*e.g.*aPKC and Par3).

1.2.2 The second cell-fate decision

The second cell-fate decision involves the segregation of cells of the blastocyst ICM so that cells in direct contact with the blastocyst cavity develop to form the second extra embryonic tissue, the primitive endoderm (PE), whereas deeper ICM do not differentiate but rather expresses pluripotency genes and form the epiblast (EPI – the progenitors for all cells of the future body). The activation of the *Gata4* and *Gata6*transcription factor genes are required for successful PE differentiation[18 - 20] and result in the up-regulation of target genes required for PE integrity [4, 20 - 23]. It has been proposed that these transcription factors also antagonize the expression of pluripotency transcription factor genes associated with the derivation of the EPI lineage (*e.g. Nanog*[4, 20]).Indeed, in the immature early blastocyst ICM a mutually exclusive but mixed inter-cell expression pattern of PE (Gata6) and EPI (Nanog) marker gene protein expression has been disclosed. This so-called 'salt and pepper' expression pattern later sorts into appropriate cell layers associated with the PE and EPI in the mature blastocyst[4, 6].

Moreover non-invasive experimental evidence by Morris et al. tracking cells, by time-lapse microscopy, from the 8-cell stage to the late blastocyst stage suggests that the ultimate fate of ICM cells are considerably biased, bound by developmental timing of cell internalization (i.e. whether that be in the first wave of asymmetric division at the 8- to 16-cell stage transition in the case of EPI or in the second 16- to 32-cell wave for PE)[1, 6, 20, 24] (see Figure 1.3 a) and c)).

One deduction from this reported bias in inner cell fate is that there is a need for earlier segregation for EPI progenitors to uphold their pluripotent characteristics. Additionally, the importance of the timing of cell internalization in biasing cell fate has been confirmed by the latest discovery that cells internalized by the first wave, in the 16-cell stage embryo, up-regulate expression of the pluripotency-related transcription factor gene *Sox2*[1, 25]. As a

result, the pluripotency of a cell is improved depending on the time it spends inside the embryo thus forming the central tenant of the ´Time Outside-Time Inside` Hypothesis proposed in 2010 by Morris *et. al*.[1, 20, 24].

However, it should be noted that a second study by Yamanaka *et. al.* opposes the importance of the sequential internalization of cells upon their succeeding fate and proposes ´random´ mechanisms assign EPI and PE progenitors[3]. (see Figure 1.3 b) and c))

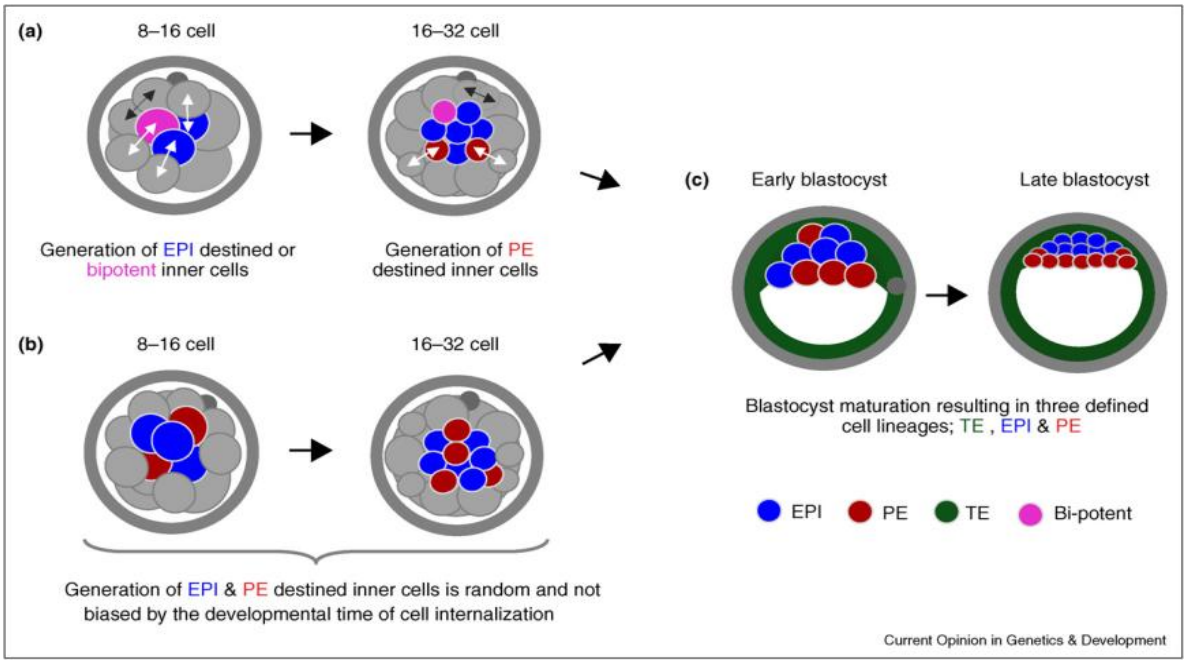


Figure 1.3: Models describing the origin of EPI and PE progenitor cells in the ICM of mouse embryos. **a)** The ´Time Outside-Time Inside` hypothesis according to Morris *et al*. (2010) proposes that cells sent inward during the first wave of asymmetric cell division, the 8- to 16-cell transition, are pluripotent and biased to form EPI (shown in blue). Whilst cells that are strongly biased to form PE (shown in red) are the ones internalized by asymmetric divisions during the second wave, the 16- to 32-cell stage transition. **b)** The `Random` model by Yamanaka *et. al.*(2010) suggests that timing of cell internalization is independent of EPI(shown in blue) and PE (shown in red) progenitor cell generation. **c)** The two models come together in an organised pattern in (first early then late) blastocyst maturation resulting in three defined cell lineages; TE, EPI and PE[1]. Image taken from Bruce, A. W *et al.,* (2010).

Another model, the ´position induction´ model, presents an early view that theoretically resembles the ´inside –outside` hypothesis for the first lineage segregation. In line with this model, cells on the surface of ICM lining the blastocyst cavity (that can be conceptualised as being outside) are activated to differentiate by hypothetical inductive signal that cannot be transmitted to the deeper ICM cells (that would represent the inner cells). It is therefore deduced from this model that early blastocyst ICM cells are both homogeneous and bipotent with respect to ultimate cell fate thereby contributing to both lineages, indicating a degree of flexibility of early ICM fate[20]. The position induction model is now somewhat dated given the finding that the early blastocyst ICM is heterogeneous for the expression of PE and EPI marker gene expression and a model based upon active ´cell sorting´ of these PE and EPI progenitors to their appropriate cell layers is now favoured.

Recently, there have been improvements regarding the understanding of the molecular determinants underlying PE specification. Sox17, a transcription factor, has come into view as novel PE marker[1, 24, 26].Also Sox17 has been proven to down-regulate pluripotency related gene expression and promotes extra-embryonic endoderm cell fate in ES cultures by competing with Nanog for the same chromatin binding sites[1, 26]. Experiments in which *Sox17*expression levels have been either artificially elevated or reduced testify to its ability to promote PE. The secreted signalling ligand Fgf4 is an additional PE promoting factor [1, 18].When embryos are cultured from the 8-cell stage in exogenously provided Fgf4 is the entire ICM of late blastocyst express the *Gata6*gene at the expense of *Nanog*[1, 5]. Apart from Nanog and Sox2, similar regulatory mechanisms for the transcription factor genes Oct4, Sall4 and Carm1 have been shown to be important for EPIderrivation[11, 27-34]. The genetic knockouts of the above mentioned factors (EPI: Nanog and Sox2; PE: Gata4, Gata6 and Sox17) are able to bring about phenotypically-overt pre-implantation development phenotypes[2, 6, 11, 14, 35, 36].However, they often reveal themselves much later in development[18, 30, 31, 37 - 41]. At times, this is conjoined with the inheritance of maternal factors[6, 30, 41]. These late phenotypes prove that significant pre-implantation cell-fate genes or roles can be obscured by the regulative properties of the embryo. Since in embryonic regulation, the embryonic tissue is capable of detecting changes in their size and location, and making the required adjustments to form the disk-shaped assembly of appropriate

structure[42]. Meaning the deficiency of an important role up to an identified developmental stage is not equal to not having any role at all.

1.3 Introduction to the experimental screen designed to identify novel cell fate related genes relating to pre-implantation development

In the Laboratory of Developmental Biology and Genomics (the lab of my supervisor) an experimental screen aimed at using the joint handles of temporal and spatial patterns of gene expression within the pre-implantation embryo was conducted to identify candidate gene mRNAs with possible functional role in influencing cell fate.

Arising from this screen were seven novel candidate cell-fate influencing genes that were to be concentrated upon in this study (the list of candidate genes is much longer but the seven candidates represent high priority candidates identified by my supervisor) – expanded upon below (*i.e.* to try and clone as many of these genes into our *in vitro* transcription plasmid vector, pRN3P, as possible or derive dsRNA constructs to effect RNAi in cell clones of the embryo). The seven genes are: *Socs3*, *Gata1*, *Zfpm1*, *Zbtb32*, *Klf4*, *Foxa1* and *Dgcr8*. Summarised below are some general information on each of the seven genes:

1.3.1 General information on the genes to be cloned or to be used to derive long dsRNAi

1) Socs3

SOCS proteins represent major regulators of the immune responses and exert their effects in classical negative-feedback regulatory loops. Socs3 also known as Suppressor of Cytokine Signalling 3 is temporarily expressed by multiple cell lineages within the immune system and functions predominantly as a negative regulator of cytokines that activate the JAK-STAT pathway[43]. In the pre-implantation mouse embryo it's mRNA expression peaks in the outer cells of the 16-cell stage conceptus.

- donor plasmid for cloning or dsRNAi generation; *not required can used genomic DNA as the gene is mono-exonic.*

- compatible restriction enzyme digests (*i.e.* restriction enzyme recognition sites not present in the target insert sequence) required to prepare recipient pRN3P plasmid and target gene insert (complete coding sequence, plus N-terminal HA epitope tag) for cloning; *BglII and XbaI (using NEB buffer 2 + BSA at 37^{o}C overnight).*
- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation; *60^{o}C for both insert generation and dsRNAi PCR template DNA generation.*
- Anticipated successful PCR product lengths; *cDNA insert + linkers and HA tag = 741bp & dsRNAi template DNA = 603bp.*

2) Gata1

Gata binding protein 1, also known as globin transcription factor 1, is a gene that encodes a protein belonging to the GATA family of transcription factors. In erythroid development, it regulates the switch of foetal haemoglobin to adult haemoglobin. Mutations in this gene have been accompanied with X-linked dyserythropoietic and thrombocytopenia[43]. In our screen of potential novel cell fate related genes the *Gata1* gene showed substantial enrichment in mRNA levels in the outer cells of the 16-cell stage embryo. Moreover, there exists precedent for other GATA family members playing important roles during pre-implantation mouse embryo cell-lineage derivation (*e.g.Gata3*, *Gata4* and *Gata6*).

- donor plasmid for cloning or dsRNAi generation; *pSPORT:Gata1 (stock concentration 1816 ng/ml).*
- Compatible restriction enzyme digests required to prepare recipient pRN3P plasmid and target gene insert for cloning; *BglII and XbaI (using NEB buffer 2 + BSA at 37^{o}C overnight).*
- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation; *60^{o}C for both insert generation and dsRNAi PCR template DNA generation.*
- anticipated successful PCR product lengths; *cDNA insert + linkers and HA tag = 1342bp & dsRNAi template DNA = 203bp*

3) Zfpm1

Zfpm1 (Zinc finger protein, multiple 1), also known as Fog1 (friend of GATA protein 1) is a transcription regulator that has a vital role in erythroid and megakaryocytic cell differentiation. It is an essential transcriptional cofactor that acts via the formation of a heterodimer with transcription factors of the GATA family GATA1, GATA2 and GATA3. Subjected to cell-type and promoter context, such Zfpm1-Gata heterodimers can either activate or repress transcriptional activity of their target genes[43]. In the screen of potentially novel cell-fate related genes, *Zfpm1* expression is enriched in the inner versus outer cells of the 16-cell sate embryo and potentially modulates the functional output of already characterised GATA factors during pre-implantation development (*e.g.Gata4* and *Gata6* in the PE versus EPI cell-fate choice in the early blastocyst ICM).

- donor plasmid for cloning or dsRNAi generation; *it has not been possible to obtain a clone from literature based sources, but we will try and clone from cDNA prepared from embryos and/ or ES cells*
- compatible restriction enzyme digests required to prepare recipient pRN3P plasmid and target gene insert for cloning; *EcoRI and BamHI (using NEB buffer 3 + BSA at 37^{o}C overnight)*
- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation; *60^{o}C for both insert generation and dsRNAi PCR template DNA generation*
- anticipated successful PCR product lengths; *cDNA insert + linkers and HA tag = 3088bp & dsRNAi template DNA = 624bp*

4) Zbtb32

Zbtb32 (zinc finger and BTB domain containing 32) is a gene that encodes aDNA-binding protein that binds to the to a 5'-TGTACAGTGT-3' core sequence. It can function as a transcriptional trans-activator and transcriptional repressor. Most certainly, it exerts its repressor effect by preventing Gata3 from binding to its DNA recognition elements and has been involved in regulating the differentiation and activation of helper T-cells[43]. In the pre-

implantation mouse embryo, its mRNA expression is enriched in outer blastomeres of the 16-cell stage embryo.

- donor plasmid for cloning or dsRNAi generation; *pCR4TOPO:Zbtb32 (62 ng/ml).*
- compatible restriction enzyme digests required to prepare recipient pRN3P plasmid and target gene insert for cloning; *BglII and XbaI (using NEB buffer 2 + BSA at 37^{o}C overnight).*
- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation;*60^{o}C for both insert generation and dsRNAi PCR template DNA generation*
- anticipated successful PCR product lengths; *cDNA insert + linkers and HA tag = 1498bp & dsRNAi template DNA = 612bp*

5) Klf4

Klf4 (Krueppel-like factor 4) is a gene that encodes Klf4transcription factor protein. Klf4 has been proved to be a good indicator of stem-like capacity in embryonic stem cells (ESCs) and in mesenchymal stem cells (MSCs)[43].

- donor plasmid for cloning or dsRNAi generation; pBRPyCAG:Klf4-IP (*205ng/ml*)
- compatible restriction enzyme digests required to prepare recipient pRN3P plasmid and target gene insert for cloning; *BglII and XbaI (using NEB buffer 2 + BSA at 37^{o}C overnight)*
- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation;*60^{o}C for both insert generation and dsRNAi PCR template DNA generation*
- anticipate successful PCR product lengths; *cDNA insert + linkers and HA tag = 1552bp & dsRNAi template DNA = 593bp*

6) Foxa1

Foxa1 (Fork-head box A1) also known as hepatocyte nuclear factor (HNF1A) is a gene that encodes a member of the fork-head class of DNA-binding proteins. FOXA1 is best known as a

transcriptional activator for liver-specific transcripts, namely albumin and transthyretin. In mice, it has been shown to function during differentiation of the pancreas and liver and to participate in the regulation of metabolism[43]. In our pre-implantation development screen, *Foxa1* peak mRNA expression is during the 8- and 16-cell stages with enriched outer cell expression at the 16-cell stage.

- donor plasmid for cloning or dsRNAi generation; pXY-Asc:Foxa1 (*1614 ng/ml*)
- compatible restriction enzyme digests required to prepare recipient pRN3P plasmid and target gene insert for cloning; *EcoRI and NotI (using NEB buffer 3 + BSA at 37^{o}C overnight)*
- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation;*60^{o}C for both insert generation and dsRNAi PCR template DNA generation*
- anticipated successful PCR product lengths; *cDNA insert + linkers and HA tag = 1509bp & dsRNAi template DNA directed to CDS = 637, dsRNAi template DNA directed to 5'UTR = 440bp*

7) ***Dgcr8***

Dgcr8 (Digeorge Syndrome Critical Region Gene 8) encodes a subunit of the microprocessor complex that transfers the biogenesis of microRNAs from the primary microRNA transcript. The encoded protein is a double-stranded RNA binding protein that functions as the non-catalytic subunit of the microprocessor complex. For the binding of the double-stranded RNA substrate and facilitation of the cleavage of the RNA by the ribonuclease III protein, Drosha, this protein is needed[43].

- donor plasmid for cloning or dsRNAi generation; *we do not have a clone, but we will try and clone from cDNA prepared from embryos and/ or ES cells*
- compatible restriction enzyme digests required to prepare recipient pRN3P plasmid and target gene insert for cloning; *BglII and XbaI (using NEB buffer 2 + BSA at 37^{o}C overnight)*

- PCR annealing temperatures for generation of target gene inserts for cloning or long dsRNA generation; *60^{o}C for both insert generation and dsRNAi PCR template DNA generation*
- anticipated successful PCR product lengths; *cDNA insert + linkers and HA tag = 2372bp & dsRNAi template DNA = 212bp*

During our preliminary experiments we found that we were not amplifying cloning inserts for all the genes listed above. Therefore we ultimately narrowed our cloning efforts to a subset of targets (*Socs3*, *Gata1* and *Zbtb32*).

2 General and specific project aims

The ultimate and primary experimental aim of this Bachelor's project was to provide/ generate experimentally useful regents/ recombinant DNA constructs that could be used to dysregulate the expression of candidate cell fate influencing genes (by either over-expression or RNAi mediated knock-down) in defined subset of cells (also known as cell clones) of the developing pre-implantation embryo. Although outside the remit and timescale of this Bachelor's project, the aim of such experimental interventions would be to observe if the induced candidate gene's dysregulation could influence eventual cell-fate of the modified cell clone in the mature blastocyst in relation to its unmodified neighbouring cells within the same embryo. Accordingly, to achieve the necessary clonal over-expression or knock-down of a candidate cell fate related gene, either mRNA or long double stranded RNA (dsRNA) specific for these genes needed to be generated prior to their microinjection (together with mRNA for fluorescent proteins to provide a convenient marker of the dysregulated cell clone) into defined cells of the pre-implantation stage embryo.

Therefore the first specific aims of this project were to clone novel cell fate influencing candidate genes' protein coding sequences into an appropriate plasmid vectors to permit *in vitro* generation of microinjectable mRNA species. The original plan was for this to be done for each of the seven candidate cell fate related genes described in the previous section (*Socs3*, *Gata1*, *Zfpm1*, *Zbtb32*, *Klf4*, *Foxa1* and *Dgcr8*). We aimed to clone our candidate

genes into a plasmid vector called pRN3P (see Figure 2.1 below). There is a need to clone the genes into this particular vector as; **a)** it permits *in vitro* transcription (IVT) to generate mRNA for micro-injection from the T_3bacteriophage RNA polymerase promoter sequence upstream of the multiple cloning site and **b)** by cloning into this vector the mRNA will be flanked by 5´and 3´ UTR from the frog *β-globin* gene and this will impart very good stability upon the IVT generated mRNAs when injected into the embryo - *i.e.* they will not be degraded easily and will persist during the pre-implantation stages of development.

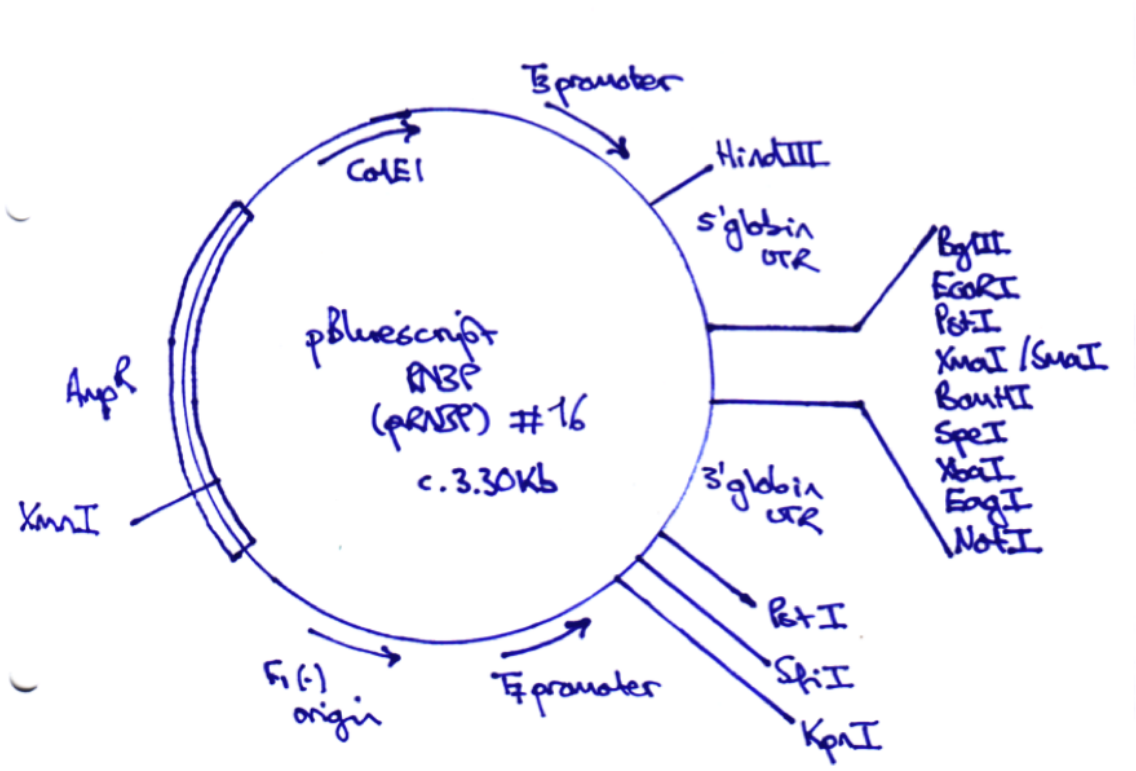


Figure 2.1: Plasmid map of pRN3P: Note the T_3 bacteriophage RNA polymerase promoter sequence, 5'/3' frog b-globin gene UTR's (flanking the multiple cloning site).

Furthermore, a specific DNA sequence was to be added at the 5´end of the target gene protein coding region inserts that would encode a 10 amino acid epitope tag (known as a HA-tag) at the N-terminus of resultant recombinant proteins in the embryo. The addition of this HA-tag was designed to allow the discrimination of endogenous and recombinant protein expressions after successful microinjection.

The second specific aim was to generate candidate gene specific double stranded PCR products that incorporated 5' promoter sequences for T_7 bacteriophage RNA polymerase, that could then be utilised to synthesise long dsRNA species by IVT. These long dsRNAs could then effect target gene knock-down in the embryo by targeting their mRNAs for degradation by the RNAi pathway.

The third specific aim was to synthesise mRNAs for the fluorescent proteins markers, GAP43-RFP and GAP43-GFP (both fluorescently mark the cell membranes, either red or green respectively), EGFP (homogenously green) and DsRed (homogenously red) by IVT using pre-existing pRN3P-fluorescent gene clones. The derived fluorescent mRNAs could then be used to identify the dysregulated cell clones in the embryo after co-microinjection with either candidate cell fate gene mRNA or long dsRNA.

3 Materials and methods

The immediately following sections will describe the procedures employed to generate the cloned target genes of interest and derive either HA-tagged mRNA for microinjection and over-expression in the embryo or long dsRNA for RNAi studies, by *in vitro* transcription (IVT). A description of more common place/ general laboratory techniques used will follow.

3.1 PCR primers for generating full-length inserts for cloning or double stranded DNA templates for long dsRNA generation

The chosen PCR primers were designed (and ordered from Sigma) to amplify the whole full-length coding sequence of target candidate genes that would also incorporate a HA tag plus Kozak translation initiation sequence and restriction enzyme sites for cloning at the 5' and 3' ends into the pRN3P vector. Similarly, the T_7 linkered primers for generating PCR products for long dsRNA production were designed in the way such that they could amplify the whole length of the complementary target gene DNA sequence.

The PCR primer pairs used to generate the candidate gene inserts (*i.e.* for Gata1,Klf4, Socs3 and Zbtb32) for cloning into pRN3Pare summarised in Table 3.1.Note that each of the forward primers contains a spacer sequence ('gactat' to permit restriction enzyme binding), the required compatible restricted enzymes site (*BglII* for all three genes, shown in purple), a kozak translation initiation sequence ('gccacc' – in blue), a start codon (ATG in green), the HA tag sequence (the last codon for which is highlighted in blue) and the gene specific sequence (starting from the second codon). In the case of the reverse primers that same spacer sequence is used, followed by the required compatible restriction enzyme site (*XbaI* in all three cases). The translational 'stop' codon is shown (in red) followed by the gene specific sequence.

Table 3.1.: Successfully employed forward (f) and reverse (r) primer sequences designed for cloning full length candidate gene-HA tagged (N-terminal) constructs.

Gata1 (f)	5'- gactatAGATCTgccacc**ATG**ggctacccatacgatgttcctgactatgctGATTTTCCTGGTCTAGGGGCCCT – 3'
Gata1 (r)	5' – gactatTCTAGATTAAAGTGGAGCATCATACTGATCCAG – 3'
Klf4 (f)	5'- gactatAGATCTgccacc**ATG**ggctacccatacgatgttcctgactatgctAGGCAGCCACCTGGCGA – 3'
Klf4 (r)	5' – GACTATtctaga**TTA**AAAGTGCCTCTTCATGTGTAAGGC – 3'
Socs3 (f)	5'- gactatAGATCTgccacc**ATG**ggctacccatacgatgttcctgactatgctGTCACCCACAGCAAGTTTCCCG – 3
Socs3 (r)	5' – gactatTCTAGATCAAGAACTGAGTGGGGCGATC – 3'
Zbtb32 (f)	5'- gactatAGATCTgccacc**ATG**ggctacccatacgatgttcctgactatgctCCCCAGACCCCCACAAGACTAATC – 3'
Zbtb32 (r)	5' – gactatTCTAGATCAGGTGGCAGCAGAGGAGG – 3'

The PCR primer pairs used to generate the double stranded PCR products required as template to generate the candidate gene specific long dsRNA are shown in Table 3.2. Note the T_7 bacteriophage RNA polymerase promoter sequence (shown in purple lower-case letters) to the 5' end of the primer, prior to the gene specific region (shown in capital letters).

Table 3.2.: Successfully employed forward (f1) and reverse (r1) primer sequences designed for generating double stranded PCR products required to generate candidate gene specific long dsRNA.

Gata1 (f1)	5' –taatacgactcactatagggCAGCATCAGCACTGGCCTAC – 3'
Gata1 (r1)	5' - taatacgactcactatagggCACACAGTTGAGGCAGGGTA – 3'
Klf4 (f1)	5' - taatacgactcactatagggGTTAGCAAAGGAAGCCCAGA – 3'
Klf4 (r1)	5' - taatacgactcactatagggGGTAAGGTTTCTCGCCTGTG – 3'
Socs3 (f1)	5' - taatacgactcactatagggTCACCCACAGCAAGTTTCC – 3'
Socs3 (r1)	5' - taatacgactcactatagggCACGTTGGAGGAGAGAGGTC – 3'
Zbtb32 (f1)	5' - taatacgactcactatagggCCAGACCCCCACAAGACTAA – 3'
Zbtb32 (r1)	5' - taatacgactcactatagggTACCTCGCATGACCTCCTCT – 3'

3.2 PCR reaction set up

3.2.1 Pfusion PCR insert generation for cloning

The Pfusion variant of DNA high-fidelity DNA polymerase (Finzymes) was used to generate the inserts for cloning the coding region of the candidate cell-fate influencing genes. The composition of each PCR reaction tube, for each of the four inserts to be generated for cloning is given in Table 3.3.Note that, the individual tubes were pulse spun in a micro-centrifuge to ensure the entire mixture was collected at the bottom of the tube prior to the subsequent PCR reaction. Also note that in all cases the template DNA used was a pre-existing and sequence verified linearized plasmid clone, albeit cloned into a different vector, for the specific gene. There was one exception in the case of the *Socs3* gene, that was amplified directly from genomic DNA (100 ng) because it is a single exon gene that permits such a strategy.

Table 3.3.: Pfusion PCR reaction mixture per sample for generation of full length candidate gene inserts for subsequent cloning into pRN3P.

	[µl]
Pfusion polymerase enzyme	0.125
10x Pfusion Buffer	2.5
dNTP mix (10 mM each)	2
Forward oligo primer(10 µM)	0.75
Reverse oligo primer(10 µM)	0.75
HPLC grade water	13.875
Template DNA (1 ng/µl)	5
Total	25

3.3 PCR machine cycle conditions

The tubes were placed into a PCR cycler and the following programme used to subject the individual reactions to 35 cycles of amplification. The programme parameters were as follows (the 'heated lid' function was used to prevent sample evaporation):

1. 98°C for 0´30" – initial denaturation of template (and primers) plus activation of the Pfusion polymerase
2. 98°C for 15´ - cycling denaturation step
3. 60°C for 0´38" – primer annealing step
4. 72°C for 2´ - product extension step
5. Go to step 2, 34 times – number of cycle repetitions
6. 72°C for 8´ - terminal extension step (to ensure all products are maximally extended)
7. 4°C forever – storage step

Following completion of the PCR cycle, 5% of the reaction was combined with 6x gel loading dye and loaded onto an agarose gel, alongside 1Kb+ DNA marker (Invitrogen), and electrophoresed to separate DNA fragments on the basis of size – i.e. length (see section 3.9.1). After completion of the electrophoresis the agarose gel was stained with ethidium

bromide, that allows visualisation of the DNA in the gel when exposed to U.V. light, to confirm successful amplification of the insert.

Following confirmation of successful PCR amplification of gene specific cloning inserts, the remainder of the PCR reactions were subject to phenol/ chloroform organic extraction and the separated DNA ethanol precipitated and the resultant DNA pellets resuspended in HPLC grade water (30 µl - see section 3.9.2). The purified PDCR DNA cloning inserts were then digested with the appropriate restriction enzymes (*Bgll l*and *Xbal*) in the presence of BSA overnight at 37^{o}C. Following the restriction digestion the DNA inserts were re-extracted, precipitated and resuspended (in 30 µl) and quantified by U.V. spectrophotometry using a nanodrop spectrophotometer.

In the case of purifying successfully amplified T_7 linked PCR products,(i.e. those to be used in long dsRNA generation), the DNA was also subject to phenol/ chloroform organic extraction, ethanol precipitation and resuspension in HPLC water. However, at this stage it was quantified by U.V. spectrophotometry using the nanodrop spectrophotometer, rather than being digested with restriction enzymes (as in the case of the preparation of the cloning inserts). In this form the DNA was prepared for use as a template in subsequent IVT (as described below in section 3.9.3).

3.4 Restriction enzyme digest of cloning insert PCR samples

The following procedures were carried out on ice.

To each of 30 µl sample of confirmed and amplified PCR cloning inserts (*i.e.* for Klf4-HA tagged, Gata1-HA tagged, Zbtb32-HA tagged and Socs3-HA tagged samples) the following were added: 5 µl of 10x Buffer 2 (NEB Biolabs), 5 µl of 10x BSA (NEB Biolabs), 2.5 µl of *Bglll*, 2.5 µl Xbal (NEB Biolabs) and5 µl of HPLC water (*i.e.* a final total volume of 50 µl). Samples were then vortexed, spun and kept overnight at 37°C to digest the PCR cloning inserts and ready them for cloning into the pRN3P plasmid vector. After completion of the restriction digestions the inserts were phenol/ chloroform extracted, precipitated with ethanol overnight and the resultant pellet (following centrifugation at full speed at 4^{o}C for 20 minutes) resuspended in 30 µl of HPLC water. The concentration of the restriction digested

prepared inserts was then determined on the nanodrop by U.V. spectrophotometry see Table 3.4).

Table 3.4.: Concentrations of restriction digested prepared cloning inserts:

HA Samples	Concentration [ng/µl]	Ratio A260/A280
Socs3 -HA	26.9	1.74
Gata1-HA	375.4	1.88
Zbtb32-HA	2777.7	1.87

3.5 Digestion of pRN3P vector to accept candidate gene inserts

In the lab was some pre-existing good quality empty pRN3P plasmid vector into which we planned to clone our restriction digest prepared candidate cell-fate gene inserts. Prior to preparing the pRN3P vector to accept these inserts (by reciprocal restriction enzyme digest) the DNA sequence of the multiple cloning site (MCS) was checked by sequencing through it using a T_3 bacteriophage RNA-polymerase promoter specific oligo (see Figure 2.1). After confirming the integrity of the restriction enzyme sites within the MCS of pRN3P, the plasmid was digested with the same restriction enzymes (*BglII* and *XbaI*) as used to prepare the inserts, as follows:

1. 5 µg of pRN3P plasmid was digested with *BglII* and *XbaI* in a 50 µl reaction using 15 units of each enzyme. The digestions were incubated in the presence of BSA for 2 hours at 37°C, pulse spun to bring water condensation down from the lid and then left at 37°C overnight.
2. 2 µl of the restricted plasmid DNA was then ran on an agarose gel, alongside non-cut plasmid, to confirm that the digestion had worked and yielded a linearized plasmid.
3. Following confirmation of the linearization, the remaining restriction digest mixture volume was increased to 135µl by addition of HPLC grade water and 15 µl of alkaline phosphatase buffer and 1 µl alkaline phosphatase enzyme (Roche) was added. The alkaline phosphatase step was included to prevent the plasmid from religating with itself (in the case of partial digestions, thus contributing artificially high background after transformation) although non-compatible restriction enzymes (in terms of the over-hangs they produce) were used (*i.e. BglII* and *XbaI*). The dephosphorylation reactions

were left for 20 minutes at 37°C, then the enzyme was heat-inactivated by a 20 minute incubation at 65°C.

4. Following the inactivation of the alkaline phosphatase, the reaction tube was pulse spun and HPLC grade water to final volume of 300 µl was added. The restriction enzyme digested and dephosphorylated pRN3P plasmid DNA was extracted by phenol/chloroform/isoamyl alcohol (IAA) organic extraction (300 µl) after which the aqueous phase was further extracted using an equal volume chloroform/IAA. To the aqueous phase a one tenth volume of3M sodium acetate pH5.2 and 2.5x volumes of 100% ethanol were added and the digested and dephosphorylated pRN3P plasmid DNA precipitated overnight at -20^{o}C.
5. Following centrifugation in a chilled bench top microfuge (+4^{o}C at 12,000 rpm for 20 minutes) the resultant DNA pellet had the supernatant removed and was washed in 70% ethanol (400 µl) using the same centrifugation conditions. The air dried pellet was then resuspend in 12 µl of HPLC grade water and digested and dephosphorylated pRN3P DNA concentration measured using the nanodrop U.V. spectrophotometer.

3.6 Setting up cloning ligation reactions and ligation transformation into *E-coli*

Optimally 200 ng of prepared vector and 300 ng of prepared inserts in 8 µl are required in an ideal ligation reaction (allowing for addition of 1 µl each of ligase enzyme and 10x buffer).This is not always possible owing to the low concentration of recovered restriction digested inserts. Therefore the appropriate amounts of both prepared vector and prepared insert (for each of the three candidate cell-fate related genes: *Gata1*, *Socs3* and *Zbtb32*) were combined in 1.5 ml eppendorf tubes and the volume increased to 40 µl with HPLC grade water. To this 4 µl of 3M sodium acetate (pH5.2) and 100 µl of 100% Ethanol were added and the samples were left in freezer overnight to precipitate both the vector and insert DNAs. The precipitates were centrifuged in a chilled bench top microfuge at full speed (12,000 rpm) for 30 minutes at 4°C after which the supernatants were discarded from the small DNA pellets. The pellets were then washed with 70% ethanol (100 µl) and left to dry at room temperature.

The DNA precipitates were thoroughly resuspended in 8 µl of HPLC grade water, after which 1 µl of 10x DNA ligase buffer and 1 µl T_4 derived DNA ligase (Roche) were added. The ligation reactions were mixed by pipetting and placed in a 16°C water bath overnight. 5 µl of the ligations were then transformed into chemically competent *E-coli* cells and then plated on ampicillin containing LB-agar plates (a 100 ng/µl ampicillin concentration was used and served to select for individual *E-coli* into which an insert ligated pRN3P plasmid, containing the *amp*^{r} resistance gene, had been introduced). After this they were incubated at 37°C overnight to allow ampicillin resistant clones to grow. Additionally, ligations with only linearized plasmid vector DNA without inserts were transformed to assess the background levels of religation or carryover of ampicillin resistant plasmids used as template in the original PCR reactions used to generate the inserts.

After overnight growth at 37^{O}C, approximately 10 colonies (*i.e.* transformants) from our experimental transformation plates were picked and streaked out on another ampicillin containing LB-agar plate. In all cases we had substantially more colonies on our experimental plates than on our corresponding negative control plates. The streaked colonies on the second plate were then incubated at 37^{O}C overnight and used as template for colony PCR reactions aimed at identifying transformants harbouring pRN3P plasmids containing the required candidate cell-fate related gene inserts. These PCRs were conducted using the same primer pairs originally used to derive the inserts. They were also performed together with water negative controls and donor plasmid positive controls to ensure that the PCR conditions were working and free from contaminating template DNA. Following a successful amplification, the bacterial colony that had the pRN3P plasmids with required inserts cloned into it could be easily identified, after they had been run on a gel. This permitted one to go back to the second streaked plate and identify the correct clone required to inoculate a liquid culture and thus prepare large quantities of the desired plasmid.

For each of the candidate cell-fate related genes that we were attempting to clone in this study, positive clones (identified by colony PCR) were identified and used toinoculate3 ml LB media liquid cultures that contain 100 ng/µl ampicillin. These were then incubated at 37°C overnight in a shaking incubator (250 rpm). Successfully grown bacterial cultures were then subject to column based mini-prep plasmid isolation performed according to the manufactuers protocol (“Qiagen Plasmid Purification Handbook” p.20-23). Following

successful purification approximately 500 ng of each mini-prep was linearized using the *XbaI* restriction enzyme (that should only cut once in the pRN3P MCS site used for cloning) and ran on an agarose gel alongside linearized normal (*i.e.* empty) pRN3P to ensure if it was correct size (*i.e.* larger – insert plus vector). One clone per candidate cell-fate related gene we were attempting to clone, that gave a linearized plasmid of the correct size, was then sent for DNA sequencing. After the confirmation by sequencing and aligning with reference cDNA sequence for the genes, glycerol stock was made for each successfully cloned gene. This was done by mixing 0.5 ml of LB culture with 0.5 ml of 50% of glycerol and put at -80°C. Then larger scale QiagenMIDI-plasmid-preps were performed to provide high concentrations (see Table 3.5) of high purity plasmid DNA to be used as template for in vitro transcription (IVT) to derive mRNAs (for HA-tagged Gata1, Klf4, Socs3 & Zbtb32) that could be micro-injected into the embryo. Note that pre-existing pRN3P based clones of experimentally useful fluorescent genes (*i.e.* DsRed, GAP43-RFP, EGFP2 and IRESGFP) were also purified by MIDI-prep and quantified (see table 3.5).

Table 3.5.: Plasmid quantification with Nanodrop

Plasmids	**Concentration [ng/µl]**	**Ratio A260/A280**
PRN3P:DsRed conc.	1456.1	1.88
PRN3P:DsRed dil.(1:10)	132.9	1.88
PRN3P:GAP3RFP con.	2103.9	1.87
PRN3P:GAP3RFP dil.(1:10)	186.5	1.84
PRN3P:EGFP2 con.	1710.9	1.87
PRN3P:EGFP2 dil.(1:10)	159.5	1.88
PRN3P:IRESGFP conc.	1772.7	1.89
PRN3P:IRESGFP dil.(1:10)	164.9	1.88
$Gata1_B$ conc.	3095.7	1.87
$Gata1_B$ dil. (1:10)	203.4	1.86
$Zbtb32_D$ conc.	3117.8	1.87
$Zbtb32_D$ dil.(1:10)	322.2	1.86
$Socs3_2$ conc.	4100.8	1.81
$Socs3_2$ dil.(1:10)	364.5	1.85

3.7 IVT protocols for genes cloned into pRN3P

- 2 μg of pRN3P plasmid was linearized overnight by digestion at 37^{o}C with the restriction enzyme *SfiI* in the presence of BSA. *SfiI* was used because the recognition site in the pRN3P plasmid is downstream of the 3' frog β-globin gene UTR and as a consequence the IVT derived mRNAs would terminate there, rather than progressing around the backbone of the plasmid (all of the cloned inserts in this study do not contain any internal *SfiI* sites).
- Phenol-chloroform extract the reaction as followed:
    1. Add HPL water to a total volume of 300 μl
    2. Add 300 μl phenol (pH (8)
    3. Vortex and spin at full speed for 5 min
    4. Transfer the upper layer (aqueous phase) into a new tube
    5. Add 300 μl of chloroform to the aqueous phase
    6. Vortex and transfer upper aqueous layer into a new tube
    7. Precipitate by adding one-tenth volume (i.e. 30 μl) of 3M NaAc (pH 5.2) and 2.5 times the total volume (i.e. 750 μl) of 100% EtOH and keep at 20°C overnight.
- After 30 minutes full speed centrifugation at 4°C in a bench top microfuge the small plasmid DNA pellet was resuspended in 7 μl of HPLC grade water (and 1 μl ran on an agarose gel alongside uncut plasmid to confirm successful linearization).
- Following confirmation of linearization the IVT was set up as follows using the "Message Machine T_3" kit from Ambion (per tube). *Note 5' to the cloned gene within the pRN3P vector is a promoter sequence for T_3 bacteriophage RNA polymerase:*

6 μl	linearized pNR3P plasmid
2 μl	10x T3 buffer
2 μl	T3 RNA polymerase
10 μl	2x NTP-cap mix
Total	20 μl

Note: Set reactions were set up on ice and the T_3 RNA polymerase added at room temperature. Also, that the 0.5 ml tubes were covered with parafilm to prevent sample loss via evaporation and incubated for exactly 4 hours at 37°C.

- The reactions were pulse spun and add 1 µl of Ambion kit DNase I added at the incubator. The reactions were then incubate for further 20 minutes at 37°C to destroy the IVT template plasmid DNA.
- The volume was increased to 300 µl with HPLC grade water and the samples phenol/chloroform then chloroform extracted. Samples were then precipitated with 100% ethanol (750 µl) and one-tenth volume (30 µl) of 3M NaAc pH 5.2 overnight at -20^{o}C.
- Precipitated RNA was spun down by full speed centrifugation (+4^{o}C) for 30 minutes in a bench top microfuge and the pellet further washed in 70% ethanol (2x 15 minutes).
- After air drying, the RNA pellet was resuspended in 20 µl of HPLC grade water.
- 0.5 µl was run on an agarose gel to confirm that IVT has worked and a band of the expected size was obtained.

3.8 Production of dsRNA from T_7-linked PCR products

The double stranded DNA T_7-linked PCR products that were generated and purified were sequence specific to the candidate cell-fate related genes whose expression we want to knock-down by RNAi in the embryo. T_7 bacteriophage-derived RNA polymerase plus UTP, ATP, GTP and CTP were therefore added to the double stranded DNA T_7-linked PCR products, that were then transcribed along both forward and reverse DNA strands to yield single-stranded but sequence complementary RNA molecules that could be made to hybridise and produce the required long dsRNA constructs. We used the Megascript T_7 IVT kit from Ambion.

3.8.1 Procedure of dsRNA production from T7-linked PCR products

I. Design gene-specific 22-24 nucleotide primers containing the T_7 RNA polymerase promoter sequence (reproduced below – promoter is underlined) to produce a double stranded DNA template fragment of 300-600 bp.

T7 +1

TAATACGACTCACTATAGGGAGA

II. Amplify IVT template by PCR from previously cloned full length open reading frame plasmid construct or pre-synthesised ES cell cDNA.

III. Produce dsRNA using Ambion MEGAscript®T_7 Kit

1. Thaw the frozen reagents, keep the enzyme and nucleotides on ice
2. Assemble transcript reaction at room temperature (otherwise the buffer could precipitate) as follows (actual volumes used for each of three constructs of interest is given in Table 3.6):

to 20 µl	Nuclease-free water
2 µl	ATP solution
2 µl	CTP solution
2 µl	GTP solution
2 µl	UTP solution
2 µl	10 x Reaction Buffer
0.1-1 µl	linear template DNA*
2 µl	Enzyme Mix

*Use 0.1-0.2 µg PCR-product template.

Table 3.6.: The actual quantities used in the transcript reaction

Actual measurements used for transcript reaction			
	Socs3	**Zbtb32**	**Klf4**
H_2O↑ 20 ml [ml]	7.5	7.5	7.5
ATP [µl]	2.0	2.0	2.0
CTP [µl]	2.0	2.0	2.0
GTP [µl]	2.0	2.0	2.0
UTP [µl]	2.0	2.0	2.0
10 x Reaction Buffer [µl]	2.0	2.0	2.0

3. Incubate at 37°C for 4 hours.
4. Add 1 µl TURBO DNase, mix well and inoculate 15 min at 37°C (this removes the IVT template DNA)

IV. Run 0.5 µl or less on a gel to check for the appropriately-sized band

V. Phenol-chloroform extract the reaction

1. Add 115 µl nuclease-free water and 15 µl ammonium acetate stop solution (from the Ambion kit)
2. Extract with 150 µl phenol mixture for RNA (pH 4.5)
3. Extract the aqueous phase with 150 µl chloroform
4. Add 150 µl of 100% isopropanol to the aqueous phase and precipitate for 30 min at -20°C
5. Pellet the RNA by centrifugation at maximum speed for 30 min at +4°Cina bench top microfuge
6. Discard the supernatant, air dry the pellet and resuspend in 10 µl nuclease-free water

VI. Anneal the complementary RNA strands by incubating at 75°C for 5 min and allowing the reactions to cool down to room temperature

VII. Treat with AmbionRNase Cocktail™ to remove any single-stranded RNA species

1. Assemble reaction

165 µl	Nuclease-free water	
10 µl	2 x RNAse digest buffer*	
10 µl	Annealed RNA	
0.35 µl	RNAse Cocktail (Ambion)	
	* 5 M NaCl	1.2 µl
	1 M Tris-Cl pH 7.4	0.2 µl
	0.5 M EDTA pH 7.5	0.2 µl
	H_2O	8.4 µl

2. Incubate for 30 min at 37°C
3. Stop reaction by adding 10 µl of 20% SDS and 2 µl proteinase K 25 mg/ml
4. Incubate for 20 mi at 37°C

VIII. Phenol-chloroform extract the reaction

1. Add 40 µl ammonium acetate stop solution (Ambion kit)
2. Extract with 400 µl phenol mixture for RNA (pH 4.5)
3. Extract the aqueous phase with 400 µl chloroform

4. Add 400 µl 100% isopropanol to the aqueous phase and precipitate for 30 min at 20°C
5. Pellet the RNA by centrifugation at maximum speed for 30 min at +4°C in a bench top microfuge
6. Discard the supernatant, air dry the pellet and resuspend in 15 µl nuclease-free water

IX. Run 0.5 µl or less on a gel to check for the appropriately-sized band and quantify the dsRNA yield by U.V. spectrophotometry using a nandrop. Confirmed long dsRNAs were then aliquoted and stored at -80^{o}C.

3.9 General laboratory methods

3.9.1 Procedure for preparing an agarose gel

1g of agarose (AgaroseServa) was added to 100 ml of 1x TAE (Tris-Acetate-EDTA^{+}) buffer and dissolved by heating in a microwave. The molten gel was cooled slightly and 2 µl of ethidium bromide (EtBr) added and mixed (to allow visualisation of the DNA under U.V. illumination). The cooled molten gel was then carefully poured into a plate, avoiding the generation of bubbles, containing a toothed comb (to generate the wells into which PCR DNA samples can be added). The gel was then allowed to cool down to room temperature and therefore set.

3.9.2 Procedure for phenol-chloroform extraction of successfully amplified PCR products

1. Add HPLC water to the remaining PCR reaction mixture to a total volume of 200 µl
2. Add 200 µl phenol (pH 8)
3. Vortex and spin at full speed for 5 min
4. Transfer the upper aqueous layer into a new tube
5. Add 200 µl of chloroform
6. Vortex, centrifuge and transfer upper aqueous layer into a new tube
7. Add 20 µl 3M $Na^{+}Ac^{-}$ pH 5.2 (sodium acetate) and 2.5 times the total volume of (*i.e.*

500 µl) of 100% ethanol (EtOH)

8. Vortex and incubate in the freezer (-20^{o}C) overnight to precipitate the DNA
9. Centrifuge at 4°C with maximum speed for 20 min
10. Discard supernatant and resuspend the DNA pellet in 25 µl of HPLC water

3.9.3 Transformation of plasmids/ ligation mixtures into chemically competent *E-coli*

For all transformations employed 1 µl of pRN3P:GAP43-GFP (10 ng/ µl) plasmid was also transformed in 'side-by-side' reactions to act as a positive control.

Step 1 to 3 are carried on ice

1. Thaw chemically competent *E.coli* from the -80^{o}C freezer.
2. Measure 97 µl of *E.coli* into the appropriate number of tubes for the number of transformations to be undertaken.
3. To each tube add either Add 1 µl of positive control pRN3P:GAP43-GFP or 5 µl of each ligation.
4. Incubate/ heatshock the tubes at 42°C in preheated water bath for 45 seconds.
5. Transfer tubes back onto ice for a further 2 minutes
6. Add 500 µl of SOC broth that was preheated to 37°C on the heating block
7. Shake the tubes in a shaker (250 rpm) at 37°C for 2 hour to allow time for the *E-coli* to recover and for successful transformants to express the ampicillin resistance gene that will be used in the subsequent selection step.
8. Centrifuge the tubes at maximum speed in a bench top microfuge for 1 min to pellet the cells
9. Discard the majority of the supernatant leaving a few µl above the cell pellet.
10. The pellet was then resuspended and transferred onto an agar plate containing ampicillin (100µg/ ml) and spread across the surface of the plate under sterile conditions. The plates were then incubated at 37°C overnight to allow the successful transformants to grow into recognisable colonies.

3.9.4 Procedure for preparation of 1 litre LB Agar

1. Add the following to 800 ml H_2O
    a. 10 g Bactotryptone
    b. 5 g yeast extract
    c. 10 g NaCl
2. Adjust the pH to 7.5 with conc. NaOH
3. Add 15 g agar
4. Dissolve the agar in the solution using microwave
5. Adjust the volume to 1 l with H_2O
6. Sterilize by autoclaving

4 Results and discussion

4.1 Cloning into pRN3P

4.1.1 Generation of PCR of inserts for cloning

Given below (Figure 4.1a) is an agarose gel resolving PCR products generated for the cloning of the candidate genes of interest into pRN3P. We originally planned to clone 7 genes, incorporating N-terminal HA tags, (*Socs3*, *Gata1*, *Zfpm1*, *Zbtb32*, *Klf4*, *Foxa1* and *Dgcr8*) but we were only able to obtain DNA products for 4 (*Socs3*, *Gata1*, *Klf4* and *Zbtb32).* The gel in Figure 4.1 shows the successful amplification of the protein coding sequences plus HA tags for *Gata1*and *Zbtb32* each totalling the expected base-pair length of 1342 and 1498 respectively. The product in the *Klf4* lane is too small compared to the theorised length of 1552 bp and is likely unspecific. Therefore the cloning of Klf4 was no longer pursued. In the case of *Socs3* cloning, the initial attempt at amplifying the insert was unsuccessful (see Figure 4.1a) but after repetition we obtained a product of the correct/ anticipated length of 741 base-pairs (Figure4.1b). Therefore we obtained specific inserts for the cloning of the HA-tagged protein coding regions of candidate cell-fate influencing genes. These three genes were for *Gata1*, *Socs3* and *Zbtb32*.

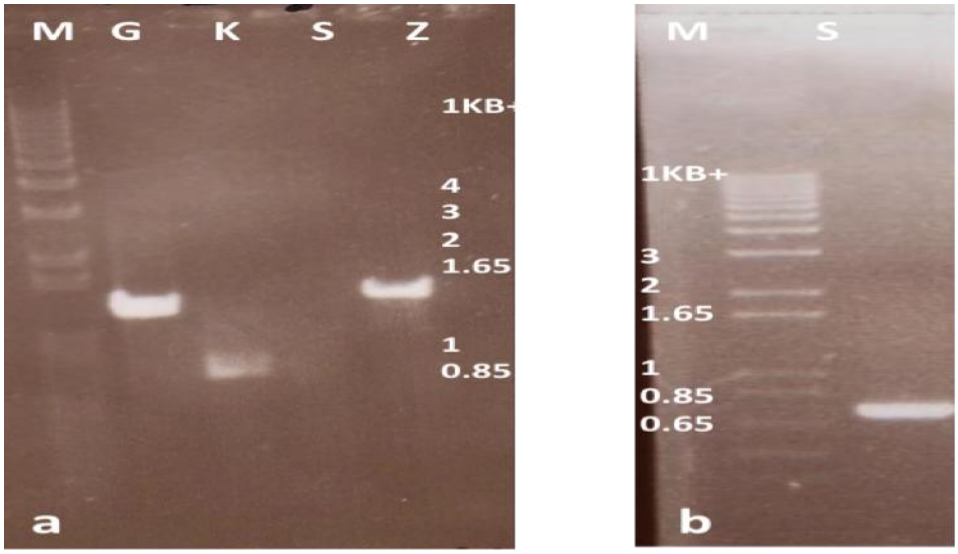


Figure 4.1: Gels showing the inserts for cloning: M: 1Kb+ dsDNA marker ladder; G: Gata1 HA; K: Klf4 HA, S: Socs3 HA and Z: Zbtb32 HA; the numbers represent the size of the marker ladder in kb. **a)** First successful amplification of inserts. **b)** Repeated amplification of Socs3 insert that had failed in **a.**

4.1.2 Identification of positive bacterial transformants

Following the PCR–based generation of cloning inserts for the *Gata1*, *Socs3* and *Zbtb32* genes (Figure 4.1), cloning into the IVT plasmid vector pRN3P was conducted as described in the materials and methods section. Following transformation into *E-coli* and selection of successful transformants by growth on ampicillin containing agar plates a number of colonies for each gene were obtained. Using a colony PCR based strategy, clones that harboured pRN3P plasmid with the desired gene insert cloned into the MCS were identified. Figure 4.2 shows an illustrative example of a colony PCR-base gel for the cloning of the *Zbtb32* gene. In this example 8 colonies (designated ZA-to ZH) were screen with 6 giving a PCR product of the correct size for the *Zbtb32* gene (1498 bp, denoted by the arrow - note that the same primer pair were used as were previously employed to generate the insert, see Figure 4.1).

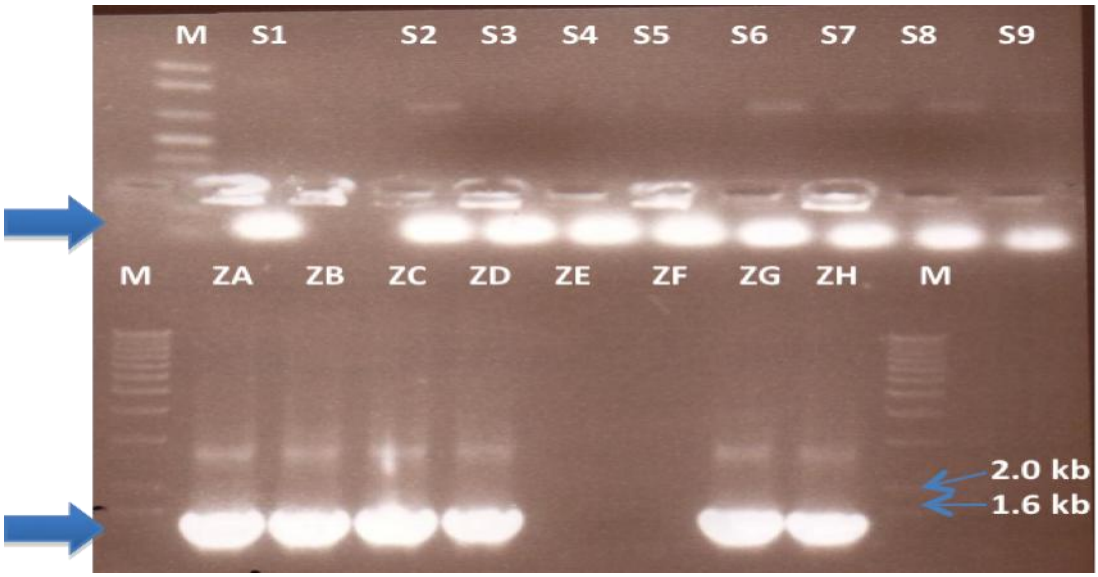


Figure 4.2: Illustrative example of a colony PCR gel, screening for *Zbtb32* positive pRN3P (and *Socs3* positive pRN3P containing bacterial clones. The upper part of the Socs3 positive pRN3P, however, was omitted in original picture) M: 1Kb+DNA marker ladder (from bottom to top: 1.65, 2, 3,..., 12 Kb). ZA-ZH representing eight clones that were screened: The arrows denote the PCR products of the expected size for positive clones harbouring pRN3P:Zbtb32-HA-N.

Following successful identification of clones harbouring pRN3P plasmid with either *Gata1*, *Socs3* or *Zbtb32* derived inserts, mini-preps were inoculated and plasmid DNA purified as described in the materials and methods. These were then subjected to restriction enzyme digest to linearize the plasmids. The linearized plasmids were then ran together with a DNA marker on an agarose gel to assess their length, in relation to that theorised by the cloning strategy (i.e. vector, 3.3Kb, plus insert). All the isolated clones for the three genes had lengths that were consistent with that theorised by successful cloning of the HA-tagged inserts into pRN3P (data not shown).

After the successful confirmation of bacterial clones harbouring pRN3P with gene specific inserts by both colony PCR and restriction enzyme based sizing of linearized plasmids referred to above, samples of the mini-prep DNA were sent for sequencing using the pRN3P upstream and downstream T_3 and T_7 promoter sequences (see Figure 2.1) as priming sites. We sequenced from both sides of the insert in order to ensure we had complete coverage of the insert. This is because in a good conventional sequencing run it can reasonably be expected to get no more than 800-900 base-pairs of sequence. As both the *Gata1* and *Zbtb32* derived inserts are greater than 1.3 kb we need to sequence from both ends. In

appendix 1 can be found the sequence alignments of our clones together with reference sequence derived from GenBank. For the reasons outlined above, an alignment for each gene insert in both forward and reverse sequence reads is given. It is important to note that in the case of the *Zbtb32* cloning, the originally selected clone (ZK) has a two base-pair deletion in the HA-tag region that was not there in another clone sequenced (ZD appendix 1). This mistake most likely derives from a mis-synthesised PCR primer but serves to illustrate the need to sequence verify any cloned DNA, as the above two mentioned methods of validation (i.e. colony PCR and sizing by restriction enzyme base linearization) would have missed this error. Following the sequencing and alignment of clones for all three genes (*Gata1*, clone G8; *Zbtb32*, clone ZD and *Socs3*, clone S2) to reference sequence, it was confirmed that we had successfully cloned three HA-tagged candidate cell-fate influencing genes into pRN3P. We have designated these plasmids; pRN3P:Gata1-HA-N, pRN3P:Zbtb32-HA-N and pRN3P:Socs3-HA-N and have placed them as glycerol stocks into the lab's -80°C plasmid archive. The plasmid maps of these three genes can be found in appendix 2.

4.2 IVT generation of mRNA for microinjection

Following the successful cloning of pRN3P:Gata1-HA-N, pRN3P:Zbtb32-HA-N and pRN3P:Socs3-HA-N, medium scale and purer MIDI preps of them were made to act as templates in IVT reactions to generate HA-tagged *Gata1*, *Zbtb32* and *Socs3* mRNAs for microinjection into the cells of the pre-implantation mouse embryo.

As described in the materials and methods, in order to perform the IVTs, the plasmids first need to be linearized by a restriction enzyme that recognises a unique restriction site downstream of the cloned insert. In the pRN3P plasmid there is a site for the enzyme *SfiI* that can fulfil this purpose. The DNA sequence of the *SfiI* recognition site is unusually long (13 bp) and not often found in cloned inserts. Indeed it is not found in any of *Gata1, Zbtb32* or *Socs3's* cloned protein coding regions and as such it was used to linearize the plasmid and render it an appropriate temple for IVT.

Figure 4.3 shows the results of the linearization, whereby uncut plasmids were run alongside a sample of linearized plasmid for each candidate cell-fate influencing gene clone. In each

case the condition that contained the *SfiI* gave a band of the expected size for the linearized plasmid (4.6 kb for *Gata1*, 4.4 kb for *Socs3* and 4.8 kb for *Zbtb32* indicated by arrows in Figure 4.3) that was larger than the supercoiled species in the conditions without restrict enzyme added. Curiously, we also observed smaller bands in the conditions plus *SfiI*. We could not explain the origin of these bands given the correct size of the linearized plasmid bands (i.e. upper bands). We therefore decided to proceed to use these prepared templates in IVT reactions and observe if the derived mRNAs were of the correct size.

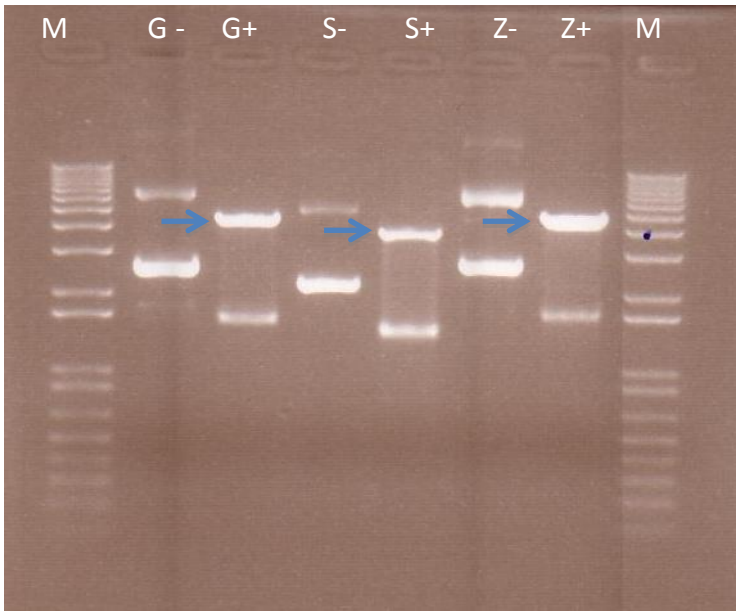


Figure 4.3: Linearization of plasmid clones with *SfiI* restriction enzyme showing M: 1Kb+ DNA marker ladder(from bottom to top: 0.1, 0.2, 0.3, 0.5, 0.65, 0.89, 1, 1.65, 2..., 12 Kb), G-: pRN3P:Gata1-HA-1 negative (minus *SfiI*), G+: pRN3P:Gata1-HA-1 positive (plus *SfiI*), S-: pRN3P:Socs3-HA-1 negative (minus *SfiI*),, S+: pRN3P:Socs3-HA-1 positive (plus *SfiI*),, Z-: pRN3P:Zbtb32-HA-1 negative (minus *SfiI*), and Z+: pRN3P:Zbtb32-HA-1 negative (plus*SfiI*). Arrows denote linearized plasmid of the correct size in conditions plus *SfiI*, that were then used a template in T3 bacteriophage based IVT reactions to generate HA-tagged candidate cell-fate influencing gene mRNAs.

Figure 4.4 shows an agarose electrophoresis gel of IVT reactions derived using the linearized plasmid DNA templates shown in figure 4.3. As can be seen successful transcription of the templates yielded mRNAs of the expected length for each of the HA-tagged genes. These mRNAs were quantified by U.V. spectrophotometry using a nanodrop, aliquoted and stored at -80^{o}C awaiting microinjection into the embryo.

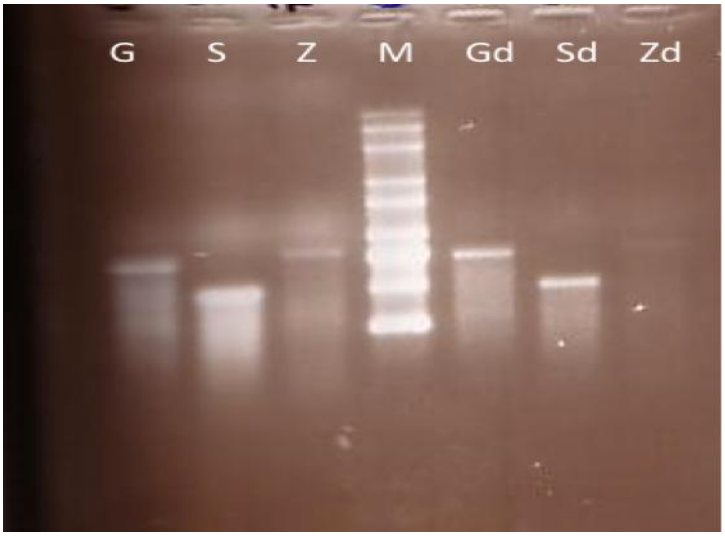


Figure 4.4: IVT reactions obtained from the linearized plasmid DNA templates illustrated in figure 4.3. showing G : Gata1-HA, S: Socs3–HA, Z: Zbtb32-HA, M: ss**RNA** marker (bottom to top 0.5, 1.0, 1.5, 2.0) , Gd: Gata1-HA 1:10 dilution, Sd: Socs3–HA 1:10 dilution and Zd: Zbtb32-HA 1:10 dilution.

4.3 Generation of T_7 linked PCR products for long dsRNA generation

Below, in figure 4.5, are the gels that show the T_7 linked PCR products generated to derive long dsRNA constructs (also shown in Figure 4.5) for RNAi mediated knock-down of candidate cell-fate influencing gene expression in the pre-implantation mouse embryo. In this case we were able to amplify PCR products for the *Klf4*, *Socs3* and *Zbtb32* genes and to derive long dsRNAs (as described in the materials and methods section) that have been archived at -80^{o}C ready for microinjection into the pre-implantation mouse embryo.

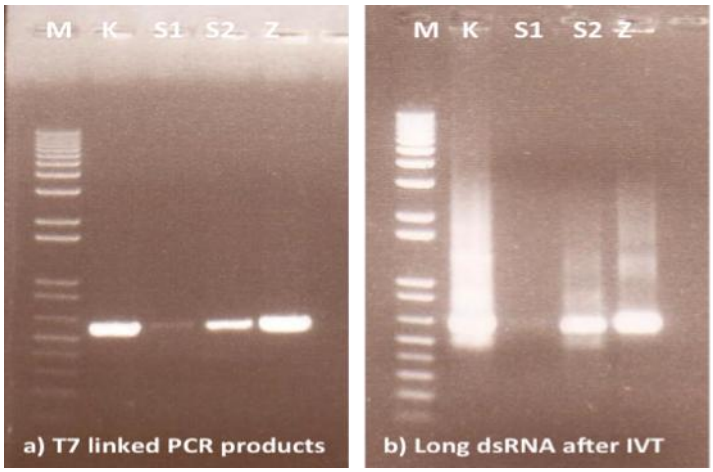


Figure 4.5: a) Electrophoresis of T7 linked PCR products and b) Electrophoresis of long dsRNA after IVT with M: 1 Kb+ DNA ladder marker from bottom to top: 0.1, 0.2, 0.3, 0.5, 0.65, 0.89,1, 1.65, 2, ...,12 Kb; K: Klf4 T7; S1: $Socs3_1$ T7; S2: $Socs3_2$ T7 and Z: Zbtbb32 T7.

4.4 Generation of mRNAs for fluorescent protein markers

During my research project, T_3bateriophage RNA polymerase driven IVTs of pre-existing pRN3P clones with fluorescent protein gene inserts was also undertaken. This was exactly analogous to the process used to derive the HA-tagged Gata1, Socs3 and Zbtb32 mRNAs. The fluorescent gene mRNA's prepared were for GAP43-GFP and GAP43-RFP (used to mark the plasma membrane green or red respectively) EGFP (green) and DsRed (red). These mRNAs were successfully amplified (Figure 4.6) and are in routine use in Dr. Alexander Bruce's laboratory.

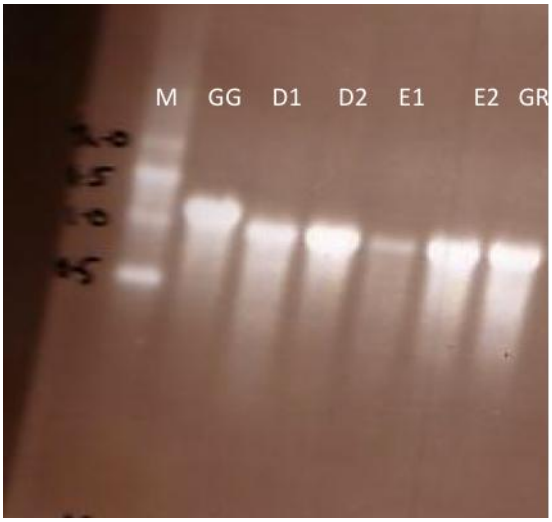


Figure 4.6: Fluorescent gene mRNA's indicating M: **ssRNA** marker (top to bottom: 2.0, 1.5 1.0, 0.5 Kb), GG: GAP43-GFP, D1-D2: DsRed, E1-E2: EGFP and GR: GAP43-RFP

5 Future experiments

During my research project it was impossible for us to do any embryo work due to the breakdown of an incubator and micro-injector. However, my supervisor, Alexander Bruce and his Ph.D. student will use both kinds of my prepared RNA´s in their experiments. Indeed preliminary experiments using the pRN3P-Zbtb32-HA-N derived HA-tagged Zbtb32 mRNA in embryo microinjections has confirmed the successful expression of the construct (Figure 5.1)

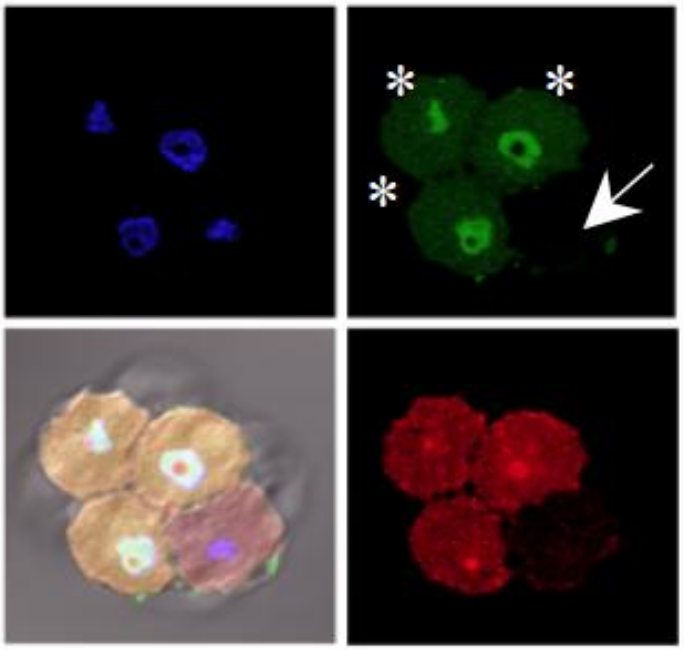


Figure 5.1: Successful HA-tagged Zbtb32 mRNA microinjection and expression in the pre-implantation mouse embryo. Single z-plane confocal stack of an 8-cell stage embryo in which one cell at the 2-cell stage was injected is shown. Left to right starting from top row: DAPI DNA counter-stain (blue), anti-HA tagged immune fluorescence, to detect the HA-tagged Zbtb32 (Alexa^{488}; green), fluorescent injection mRNA control (*DsRed*; red) & phase contrast merged with fluorescent images. Asterisks and arrow mark cells derived from the injected and non-injected cell clones respectively.

In relation to the genes we could not amplify by PCR, we would propose some possibilities for improvement: These would centre on altering the PCR cycling conditions on the PCR machine and particularly the annealing temperature.

6 Conclusions

Putting the main target of this thesis, *i.e.* the cloning of candidate novel cell-fate influencing genes regarding pre-implantation in mouse development, in focus, l have succeeded to clone some of the 7 genes (*Gata1*, *Socs3* and *Zbtb32*) into pRN3P. Consequently, these cloned inserts, representing the protein coding regions of the genes, now have N-terminal HA epitope tags and this theoretically will permit the distinguishing of recombinant protein expression from endogenous protein expression in the embryo cell after microinjection. This can be achieved by using an anti-HA antibody in immune-fluorescent staining. The plasmid maps of these successfully cloned genes can be viewed in the appendix 2. These constructs can now be used in functional studies in the embryo

In the same line, l was able to generate T_7 linked PCR products for some but not all candidate genes and successfully derive long dsRNA by IVT. These long dsRNAs can now also be microinjected into embryo cells to elicit gene expression knock-down in functional studies.

Additionally, l mediated IVTs from pre-existing plasmid clones to generate mRNAs for fluorescent proteins. These are necessary for functional studies as they are co-injected into cells of the embryo with mRNA or long dsRNA for candidate cell fate influencing genes. The successful expression of fluorescent protein being an indicator that the experimental mRNA or long dsRNA was successfully delivered into the cells. This is also important because in this way we can tell which cells in the embryo have been dysregulated, for they will fluoresce whereas the non-injected cells (and their progeny, following cell divisions) will not – see figure 5.1.

7 Appendix

Appendix 1: Sequence alignments for pRN3P:Zbtb32-HA-N, Gata1-HA-N and Socs3-HA-N against reference GenBank derived reference sequences

Below are reproduced alignments of the cloned candidate cell-fate gene plasmids (pRN3P:Zbtb32-HA-N, Gata1-HA-N and Socs3-HA-N) against reference GenBank sequence for the protein coding regions. In each case the sequence derived from a forward read using the T_3 promoter sequence upstream of the insert integration site is given first, followed by the reverse sequence derived from the downstream T_7 promoter. As one can reasonably only expect 800-900 base-pairs of good sequence per read, the point at which both forward and reverse reads overlap in reliable sequence is highlighted in blue text in the T_7 derived sequence (for pRN3P:Zbtb32-HA-N and Gata1-HA-N – not required for the small pRN3P:Socs3-HA-N insert), thus ensuring the entire insert was sequenced.

In the case of the pRN3P:Zbtb32-HA-N plasmid reads, two clones are represented owing to the identification of a two base-pair deletion in the HA-tag region of the originally isolated clone (designated ZK), most probably due to an error in primer manufacture. Therefore a second clone was picked (ZD) and sequenced and found not to contain the error.

Important sequence features in alignments shown below are highlight thus:

BglII site used in cloning

Kozak translational initiation sequence

HA epitope tag

Start of endogenous mouse *Zbtb32* gene DNA sequence

START codon

STOP codon

Point mutations

XbaI site used in cloning

Over-lapping and reliable sequence between forward and reverse sequence reads

pRN3P:Zbtb32-HA-1 clones D and K sequence alignment (using T_3 promoter in vector) *– note two base-pair deletion in the HA epitope tag region of the originally selected clone ZK compared to subsequently selected ZD clone.*

```
CLUSTAL 2.1 multiple sequence alignment

ZD3             CTGGRCTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCAGA 60
ZK3             CWGCRCTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCAGA 60
                * * ************************************************************

ZD3             TCTGCCACCATGGGCTACCCATACGATGTTCCTGACTATGCTCCCCAGACCCCCACAAGA 120
ZK3             TCTGCCACCATGG--TACCCATACGATGTTCCTGACTATGCTCCCCAGACCCCCACAAGA 118
                *************  ******************************************

ZD3             CTAATCAGCCCATATGGCTCTGATAGGTTGGTACAGTTAGCGGCTAGACTCCGGCCAGCA 180
ZK3             CTAATCAGCCCATATGGCTCTGATAGGTTGGTACAGTTAGCGGCTAGACTCCGGCCAGCA 178
                ************************************************************

ZD3             CTGTGTGATACCCTGATCACAGTAGGGGGCCTGGAGTTCCCGGCTCACAGCCTAGTGCTG 240
ZK3             CTGTGTGATACCCTGATCACAGTAGGGGGCCTGGAGTTCCCGGCTCACAGCCTAGTGCTG 238
                ************************************************************

ZD3             GCAGGTGCAAGCCCAAGGCTTGGCTGCAGGGGCCGGTGGGCTCTGGTTGAAGACATAAGC 300
ZK3             GCAGGTGCAAGCCCAAGGCTTGGCTGCAGGGGCCGGTGGGCTCTGGTTGAAGACATAAGC 298
                ************************************************************

ZD3             CCTTCCACCTTTGCTCAGATTCTGACCTTTGTCTATGGAGAGAGTATAGAGCTACAGCCT 360
ZK3             CCTTCCACCTTTGCTCAGATTCTGACCTTTGTCTATGGAGAGAGTATAGAGCTACAGCCT 358
                ************************************************************

ZD3             GGGGAGCTGGGGGACCTTGAAGAGGCAGCCAAAGCCTTGGGGGTACAGGCCTTGGAAGAG 420
ZK3             GGGGAGCTGGGGGACCTTGAAGAGGCAGCCAAAGCCTTGGGGGTACAGGCCTTGGAAGAG 418
                ************************************************************

ZD3             GCCTGTCAGAGAGCTCAAAAGGGCAAGGATGAAGATGAGCTGGATCCAGGACTGAAGAGG 480
ZK3             GCCTGTCAGAGAGCTCAAAAGGGCAAGGATGAAGATGAGCTGGATCCAGGACTGAAGAGG 478
```

```
                ************************************************************

ZD3             CACCAGCAATCAGAAGACTTCATGAGGGGCTCTGAGAGAGGACTTGGGAGTCCTGGAGAG 540
ZK3             CACCAGCAATCAGAAGACTTCATGAGGGGCTCTGAGAGAGGACTTGGGAGTCCTGGAGAG 538
                ************************************************************

ZD3             AAACAGAAGCCAGAGAAGGATTTTAGAAGTAATGGGAGAGAACAGGAGATGTCACACAAG 600
ZK3             AAACAGAAGCCAGAGAAGGATTTTAGAAGTAATGGGAGAGAACAGGAGATGTCACACAAG 598
                ************************************************************

ZD3             CATAAAGCACCCGGAGAGAGGCCTGAGATGGCAGGAGCAACTAGGATGATGAGCTCAGAG 660
ZK3             CATAAAGCACCCGGAGAGAGGCCTGAGATGGCAGGAGCAACTAGGATGATGAGCTCAGAG 658
                ************************************************************

ZD3             GAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAGGAGAGTCTTCGTGGGTGCCCT 720
ZK3             GAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAGGAGAGTCTTCGTGGGTGCCCT 718
                ************************************************************

ZD3             GATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTCATTCCCAGGCCCTGGTGGGCT 780
ZK3             GATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTCATTCCCAGGCCCTGGTGGGCT 778
                ************************************************************

ZD3             GAAGTCCCTAGGTTGGGGGGAGGGCCAGTCGGCCTTGTGGAGCATCTTGTTGTGGCCCTC 840
ZK3             GAAGTCCCTAGGTTGGGGG-AGGGCCAGTCGGCCTTGTGGAGCATCTTGTTGTGGCCCTC 837
                ****************** ****************************************

ZD3             CAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACTGCGGCCTGGCACGTCCGGGC 900
ZK3             CAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACTGCGGCCTGGCAGGTCCGG-C 896
                ************************************************** ****** *

ZD3             CTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAAGCTCTCTGGAGTCAGA-TC 959
ZK3             CTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAAGCTCTCTGGAGTCAGAATC 956
                ************************************************************ **

ZD3             AGCTGGCTTC-TCTAGCCCCACTCCAGGATCTTTTCCYCAGGCACTGAAACCMTCAGCCC 1018
ZK3             AGCTGGCTTCCTCTAGCCC-ACTCCAGGATCTTT-CCCCAGGCACTGAAACCCTCAGCCT 1014
                ********** ******** ************** ** ************** ******
```

```
ZD3             TG-CWGATAGAAACGTCTGGCA-GGGTTCACAG-CMCACTG-CAC-TGKGTGAGTCAGGA 1073
ZK3             TGGCAGATAGAAACGTCTGGCAAGGGTTCMCAGGCMCACTGGCACCTGTGKGAGTCAGAG 1074
                ** * ***************** ****** *** ******* *** ** * *******

ZD3             CSGCMACTKGACTGGCATCTCACCAACATCT-TCAACCTTATCCTG-TCGATT-CGGCCA 1130
ZK3             CGCACACTG-ACTG-CATCTCACCAACACCTCTTAACCYT-TCTGGCTCGCYTTYGGCCG 1131
                *    ***  **** ************* ** * **** * **  * ***  *  ****
```

pRN3P:Zbtb32-HA-1 clones D and K sequence alignment with published mouse *Zbtb32* gene protein coding sequence (using T_3 promoter in vector)

```
CLUSTAL 2.1 multiple sequence alignment

ZD3                 CTGGRCTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCAGA 60
ZK3                 CWGCRCTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCAGA 60
Zbtb32_ref_seq      ------------------------------------------------------------

ZD3                 TCTGCCACCATGGGCTACCCATACGATGTTCCTGACTATGCTCCCCAGACCCCCACAAGA 120
ZK3                 TCTGCCACCATGG--TACCCATACGATGTTCCTGACTATGCTCCCCAGACCCCCACAAGA 118
Zbtb32_ref_seq      ---------ATG-----------------------------CCCCAGACCCCCACAAGA 21
                            ***                             ******************

ZD3                 CTAATCAGCCCATATGGCTCTGATAGGTTGGTACAGTTAGCGGCTAGACTCCGGCCAGCA 180
ZK3                 CTAATCAGCCCATATGGCTCTGATAGGTTGGTACAGTTAGCGGCTAGACTCCGGCCAGCA 178
Zbtb32_ref_seq      CTAATCAGCCCATATGGCTCTGATAGGTTGGTACAGTTAGCGGCTAGACTCCGGCCAGCA 81
                    ************************************************************

ZD3                 CTGTGTGATACCCTGATCACAGTAGGGGGCCTGGAGTTCCCGGCTCACAGCCTAGTGCTG 240
ZK3                 CTGTGTGATACCCTGATCACAGTAGGGGGCCTGGAGTTCCCGGCTCACAGCCTAGTGCTG 238
Zbtb32_ref_seq      CTGTGTGATACCCTGATCACAGTAGGGGGCCTGGAGTTCCCGGCTCACAGCCTAGTGCTG 141
                    ************************************************************

ZD3                 GCAGGTGCAAGCCCAAGGCTTGGCTGCAGGGGCCGGTGGGCTCTGGTTGAAGACATAAGC 300
ZK3                 GCAGGTGCAAGCCCAAGGCTTGGCTGCAGGGGCCGGTGGGCTCTGGTTGAAGACATAAGC 298
```

```
Zbtb32_ref_seq      GCAGGTGCAAGCCCAAGGCTTGGCTGCAGGGGCCGGTGGGCTCTGGTTGAAGACATAAGC 201
                    ************************************************************

ZD3                 CCTTCCACCTTTGCTCAGATTCTGACCTTTGTCTATGGAGAGAGTATAGAGCTACAGCCT 360
ZK3                 CCTTCCACCTTTGCTCAGATTCTGACCTTTGTCTATGGAGAGAGTATAGAGCTACAGCCT 358
Zbtb32_ref_seq      CCTTCCACCTTTGCTCAGATTCTGACCTTTGTCTATGGAGAGAGTATAGAGCTACAGCCT 261
                    ************************************************************

ZD3                 GGGGAGCTGGGGGACCTTGAAGAGGCAGCCAAAGCCTTGGGGGTACAGGCCTTGGAAGAG 420
ZK3                 GGGGAGCTGGGGGACCTTGAAGAGGCAGCCAAAGCCTTGGGGGTACAGGCCTTGGAAGAG 418
Zbtb32_ref_seq      GGGGAGCTGGGGGACCTTGAAGAGGCAGCCAAAGCCTTGGGGGTACAGGCCTTGGAAGAG 321
                    ************************************************************

ZD3                 GCCTGTCAGAGAGCTCAAAAGGGCAAGGATGAAGATGAGCTGGATCCAGGACTGAAGAGG 480
ZK3                 GCCTGTCAGAGAGCTCAAAAGGGCAAGGATGAAGATGAGCTGGATCCAGGACTGAAGAGG 478
Zbtb32_ref_seq      GCCTGTCAGAGAGCTCAAAAGGGCAAGGATGAAGATGAGCTGGATCCAGGACTGAAGAGG 381
                    ************************************************************

ZD3                 CACCAGCAATCAGAAGACTTCATGAGGGGCTCTGAGAGAGGACTTGGGAGTCCTGGAGAG 540
ZK3                 CACCAGCAATCAGAAGACTTCATGAGGGGCTCTGAGAGAGGACTTGGGAGTCCTGGAGAG 538
Zbtb32_ref_seq      CACCAGCAATCAGAAGACTTCATGAGGGGCTCTGAGAGAGGACTTGGGAGTCCTGGAGAG 441
                    ************************************************************

ZD3                 AAACAGAAGCCAGAGAAGGATTTTAGAAGTAATGGGAGAGAACAGGAGATGTCACACAAG 600
ZK3                 AAACAGAAGCCAGAGAAGGATTTTAGAAGTAATGGGAGAGAACAGGAGATGTCACACAAG 598
Zbtb32_ref_seq      AAACAGAAGCCAGAGAAGGATTTTAGAAGTAATGGGAGAGAACAGGAGATGTCACACAAG 501
                    ************************************************************

ZD3                 CATAAAGCACCCGGAGAGAGGCCTGAGATGGCAGGAGCAACTAGGATGATGAGCTCAGAG 660
ZK3                 CATAAAGCACCCGGAGAGAGGCCTGAGATGGCAGGAGCAACTAGGATGATGAGCTCAGAG 658
Zbtb32_ref_seq      CATAAAGCACCCGGAGAGAGGCCTGAGATGGCAGGAGCAACTAGGATGATGAGCTCAGAG 561
                    ************************************************************

ZD3                 GAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAGGAGAGTCTTCGTGGGTGCCCT 720
ZK3                 GAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAGGAGAGTCTTCGTGGGTGCCCT 718
Zbtb32_ref_seq      GAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAGGAGAGTCTTCGTGGGTGCCCT 621
                    ************************************************************
```

```
ZD3                 GATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTCATTCCCAGGCCCTGGTGGGCT 780
ZK3                 GATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTCATTCCCAGGCCCTGGTGGGCT 778
Zbtb32_ref_seq      GATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTCATTCCCAGGCCCTGGTGGGCT 681
                    ************************************************************

ZD3                 GAAGTCCCTAGGTTGGGGGGAGGGCCAGTCGGCCTTGTGGAGCATCTTGTTGTGGCCCTC 840
ZK3                 GAAGTCCCTAGGTTGGGGG-AGGGCCAGTCGGCCTTGTGGAGCATCTTGTTGTGGCCCTC 837
Zbtb32_ref_seq      GAAGTCCCTAGGTTGGGGG-AGGGCCAGTCGGCCTTGTGGAGCATCTTGTTGTGGCCCTC 740
                    ****************** ****************************************

ZD3                 CAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACTGCGGCCTGGCACGTCCGGGC 900
ZK3                 CAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACTGCGGCCTGGCAGGTCCGG-C 896
Zbtb32_ref_seq      CAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACTGCGGCCTGGCAGGTCCGG-C 799
                    ************************************************** ****** *

ZD3                 CTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAAGCTCTCTGGAGTCAGA-TC 959
ZK3                 CTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAAGCTCTCTGGAGTCAGAATC 956
Zbtb32_ref_seq      CTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAAGCTCTCTGGAGTCAGAATC 859
                    ********************************************************** **
```

pRN3P:Zbtb32-HA-1 clone D sequence alignment with published mouse *Zbtb32* gene protein coding sequence (using T_7 promoter in vector)

```
CLUSTAL 2.1 multiple sequence alignment

Zbtb32_ref          GAACAGGAGATGTCACACAAGCATAAAGCACCCGGAGAGAGGCCTGAGATGGCAGGAGCA 540
ZD7R_revcom         ------------------------------------------------------------AGCA 4
                                                                                ****

Zbtb32_ref          ACTAGGATGATGAGCTCAGAGGAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAG 600
ZD7R_revcom         ACTAGGATGATGAGCTCAGAGGAGGTCATGCGAGGTATCGAGAGCCACAAGGGCTCTGAG 64
                    ************************************************************

Zbtb32_ref          GAGAGTCTTCGTGGGTGCCCTGATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTC 660
ZD7R_revcom         GAGAGTCTTCGTGGGTGCCCTGATCCCCTTTCCCCACCAGGTTCCCTCCTAACTAGCCTC 124
```

```
                ************************************************************

Zbtb32_ref      ATTCCCAGGCCCTGGTGGGCTGAAGTCCCTAGGTTGGGGGAGGGCCAGTCGGCCTTGTGG 720
ZD7R_revcom     ATTCCCAGGCCCTGGTGGGCTGAAGTCCCTAGGTTGGGGGAGGGCCAGTCGGCCTTGTGG 184
                ************************************************************

Zbtb32_ref      AGCATCTTGTTGTGGCCCTCCAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACT 780
ZD7R_revcom     AGCATCTTGTTGTGGCCCTCCAGATACGGTGCTCCCTTCTCCCATAGTACCCCCATCACT 244
                ************************************************************

Zbtb32_ref      GCGGCCTGGCAGGTCCGGCCTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAA 840
ZD7R_revcom     GCGGCCTGGCAGGTCCGGCCTCAAGATCAGAGGATCCCACTGACCTTGAACCACTCCAAA 304
                ************************************************************

Zbtb32_ref      GCTCTCTGGAGTCAGAATCAGCTGGCTTCCTCTAGCCCCACTCCAGGATCTTTTCCCCAG 900
ZD7R_revcom     GCTCTCTGGAGTCAGAATCAGCTGGCTTCCTCTAGCCCCACTCCAGGATCTTTTCCCCAG 364
                ************************************************************

Zbtb32_ref      GGCACTGAAACCCTCAGCCCTTGGCAGATAGAAACGTCTGGGCAAGGGTTCACAGGCACA 960
ZD7R_revcom     GGCACTGAAACCCTCAGCCCTTGGCAGATAGAAACGTCTGGGCAAGGGTTCACAGGCACA 424
                ************************************************************

Zbtb32_ref      CTGGCAACCTGTGTGAGTCAGGAGCGCACACTGAACTGCCCATCTCACCAACACCCTCCT 1020
ZD7R_revcom     CTGGCAACCTGTGTGAGTCAGGAGCGCACACTGAACTGCCCATCTCACCAACACCCTCCT 484
                ************************************************************

Zbtb32_ref      TTACCCTCTCCTGCTCGCTCTCGGCCCTATTCTTGCTCTGTGTGTGGAAAGAGGTTTTCA 1080
ZD7R_revcom     TTACCCTCTCCTGCTCGCTCTCGGCCCTATTCTTGCTCTGTGTGTGGAAAGAGGTTTTCA 544
                ************************************************************

Zbtb32_ref      CTCAAGCATCAGATGGAGACACATTACCGAGTCCACACAGGAGAGAAGCCCTTCTCCTGT 1140
ZD7R_revcom     CTCAAGCATCAGATGGAGACACATTACCGAGTCCACACAGGAGAGAAGCCCTTCTCCTGT 604
                ************************************************************

Zbtb32_ref      AGCCTCTGTCCTCAGCGCTCCCGGGATTTCTCTGCCATGACCAAGCACCTGAGGACACAT 1200
ZD7R_revcom     AGCCTCTGTCCTCAGCGCTCCCGGGATTTCTCTGCCATGACCAAGCACCTGAGGACACAT 664
                ************************************************************
```

```
Zbtb32_ref       GGGGCTGCTCCCTATCGATGCCCTCTGTGCAGGGCTGGCTGCCCTAGCCTGGCCTCCATG 1260
ZD7R_revcom      GGGGCTGCTCCCTATCGATGCCCTCTGTGCAGGGCTGGCTGCCCTAGCCTGGCCTCCATG 724
                 ************************************************************

Zbtb32_ref       CAGGCGCACATGCGTGGCCACTCGCCCAGCCGGCTGCCGCCTGGATGGACCATACGCTCC 1320
ZD7R_revcom      CAGGCGCACATGCGTGGCCACTCGCCCAGCCGGCTGCCGCCTGGATGGACCATACGCTCC 784
                 ************************************************************

Zbtb32_ref       ACCTTCCTCTACTCTTCCTCGAGGCCGACCCGGGCCTCCAGCTCTCCCGGTAGTCCTACC 1380
ZD7R_revcom      ACCTTCCTCTACTCTTCCTCGAGGCCGACCCGGGCCTCCAGCTCTCCCGGTAGTCCTACC 844
                 ************************************************************

Zbtb32_ref       TCCTCTGCTGCCACCTGA- 1398
ZD7R_revcom      TCCTCTGCTGCCACCTGAT 863
                 ******************
```

pRN3P:Gata1-HA-N clone B sequence alignment with published mouse *Gata1* gene protein coding sequence (using T3 promoter in vector)

This T_3 derived sequence identified four point mutations (mutation positions have been marked grey) relative to the GenBank reference sequence: two of them are silent mutations and the other two cause amino acid substitutions (D29S and N129S). We know it is not a mistake because these point mutations are present in an additional clone that has been sequenced and the same point mutations are present in the template *Gata1* plasmid used to generate the cloning insert by PCR.

```
CLUSTAL 2.1 multiple sequence alignment

G83              CKGCRCTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCAGA 60
Gata1_ref        ------------------------------------------------------------

G83              TCTGCCACCATGGGCTACCCATACGATGTTCCTGACTATGCTGATTTTCCTGGTCTAGGG 120
Gata1_ref        ---------ATG-------------------------------GATTTTCCTGGTCTAGGG 21
```

```
                    ***                                 ******************

G83             GCCCTGGGGACCTCAGAACCCTTGCCCCAGTTTGTGGATTCTGCCCTGGTGTCCTCACCA 180
Gata1_ref       GCCCTGGGGACCTCAGAACCCTTGCCCCAGTTTGTGGATTCTGCCCTGGTGTCCTCACCA 81
                ************************************************************

G83             TCAGGTTCCACAGGTTTCTTTTCCTCTGGGCCAGAGGGTTTGGATGCAGCATCTTCTTCC 240
Gata1_ref       TCAGATTCCACAGGTTTCTTTTCCTCTGGGCCAGAGGGTTTGGATGCAGCATCTTCTTCC 141
                **** ********************************************************

G83             ACTTCCCCAAATGCAGCCACGGCCGCAGCATCAGCACTGGCCTACTACAGAGAAGCTGAG 300
Gata1_ref       ACTTCCCCAAATGCAGCCACGGCCGCAGCATCAGCACTGGCCTACTACAGAGAAGCTGAG 201
                ************************************************************

G83             GCCTACAGACACTCCCCAGTCTTTCAGGTGTATCCACTGCTCAACAGTATGGAGGGAATT 360
Gata1_ref       GCCTACAGACACTCCCCAGTCTTTCAGGTGTATCCACTGCTCAACAGTATGGAGGGAATT 261
                ************************************************************

G83             CCTGGGGGCTCACCTTATGCTAGCTGGGCCTATGGCAAGACGGCACTCTACCCTGCCTCA 420
Gata1_ref       CCTGGGGGCTCACCTTATGCTAGCTGGGCCTATGGCAAGACGGCACTCTACCCTGCCTCA 321
                ************************************************************

G83             ACTGTGTGCCCCAGCCATGAGGATGCCCCTTCCCAGGCCCTGGAGGACCAGGAAGGGAAG 480
Gata1_ref       ACTGTGTGCCCCAGCCATGAGGATGCCCCTTCCCAGGCCCTGGAAGACCAGGAAGGGAAG 381
                ******************************************* ****************

G83             AGCAGCAACACGTTCTTGGACACCTTGAAGACGGAGCGGCTGAGTCCAGACCTCCTGACG 540
Gata1_ref       AGCAACAACACGTTCTTGGACACCTTGAAGACGGAGCGGCTGAGTCCAGACCTCCTGACG 441
                **** ********************************************************

G83             CTGGGGACTGCACTGCCTGCATCGCTCCCTGTCACCGGCAGTGCTTACGGGGGAGCTGAC 600
Gata1_ref       CTGGGGACTGCACTGCCTGCATCGCTCCCTGTCACCGGCAGTGCTTACGGGGGAGCTGAC 501
                ************************************************************

G83             TTTCCCAGTCCTTTCTTCTCTCCCACTGGGAGCCCTCTCAGCTCAGCAGCCTACTCTTCC 660
Gata1_ref       TTTCCCAGTCCTTTCTTCTCTCCCACTGGGAGCCCTCTCAGCTCAGCAGCCTATTCTTCC 561
                ************************************************** ******
```

```
G83             CCCAAGTTTCATGGAAGCCTGCCATTGGCCCCTTGTGAGGCCAGAGAGTGTGTGAACTGT 720
Gata1_ref       CCCAAGTTTCATGGAAGCCTGCCATTGGCCCCTTGTGAGGCCAGAGAGTGTGTGAACTGT 621
                ************************************************************

G83             GGAGCAACGGCTACTCCACTGTGGCGGAGGGACAGGACAGGTCACTACCTGTGCAATGCC 780
Gata1_ref       GGAGCAACGGCTACTCCACTGTGGCGGAGGGACAGGACAGGTCACTACCTGTGCAATGCC 681
                ************************************************************

G83             TGTGGCTTGTATCACAAGATGAATGGTCAGAACCGGCCTCTCATCCGGCCCAAGAAGCGA 840
Gata1_ref       TGTGGCTTGTATCACAAGATGAATGGTCAGAACCGGCCTCTCATCCGGCCCAAGAAGCGA 741
                ************************************************************

G83             ATGATTGTCAGCAAACGGGCAGGCACCCAATGCACTAACTGTCAAACGACCACTACAACA 900
Gata1_ref       ATGATTGTCAGCAAACGGGCAGGCACCCAATGCACTAACTGTCAAACGACCACTACAACA 801
                ************************************************************

G83             CTCTGGCGGAGGAATGCCAGCGGAGATCCGGTATGCAATGCCTGCGGCCTCTATTC--AG 958
Gata1_ref       CTCTGGCGGAGGAATGCCAGCGGAGATCCGGTATGCAATGCCTGCGGCCTCTATTTCAAG 861
                ************************************************   **

G83             CTCCATCAG-TGA-CCGCC-ACTGACCATGAG--AAGATGGA-TCCAGACGAGA---CGC 1009
Gata1_ref       CTCCATCAGGTGAACCGCCCACTGACCATGAGGAAAGATGGAATCCAGACGAGGAACCGC 921
                ********* *** ***** ************  ******** **********    ***

G83             A-GGCATCTG-CAAAGGAAAAGAG---CGGGGGTCGA-TCTG-CTGGA-CGGGAGCG-CT 1060
Gata1_ref       AAGGCATCTGGCAAAGGGAAAAAGAAGCGGGGGTCGAATCTGGCTGGAGCGGGAGCGGCT 981
                * ******** ****** *** **   ********** **** ***** ******** **

G83             GAGA---CAGCTGGTG--CTCATGT--GTAGCTG-TA-CA-CAGTA-TGCAA--TGKGGG 1107
Gata1_ref       GAAGGACCAGCTGGTGGCTTCATGGTGGTAGCTGGTAGCAGCAGTAGTGGGAATTGTGGG 1041
                **     *********   *****   ******* ** ** ***** **  *  ** ***

G83             -ARGTGCCCT--AGCTG---ACTGGTCACTGCAG-TACTGC--AACCCTA---------- 1148
Gata1_ref       GAGGTGGCCTCAGGCTTGGCACTGGGCACTGCAGGTACTGCCCACCTCTATCAGGGCCTA 1101
                 * *** ***   ***    ***** ******** ******  * * ***
```

pRN3P:Gata1-HA-N clone B sequence alignment with published mouse *Gata1* gene protein coding sequence (using T_7 promoter in vector)

```
CLUSTAL 2.1 multiple sequence alignment

Gata1_ref       AGTGCTTACGGGGGAGCTGACTTTCCCAGTCCTTTCTTCTCTCCCACTGGGAGCCCTCTC 540
G7R             ------------------------------------------------------------TC 2
                                                                             **

Gata1_ref       AGCTCAGCAGCCTATTCTTCCCCCAAGTTTCATGGAAGCCTGCCATTGGCCCCTTGTGAG 600
G7R             AGCTCAGCAGCCTACTCTTCCCCCAAGTTTCATGGAAGCCTGCCATTGGCCCCTTGTGAG 61
                ************** ********************************************

Gata1_ref       GCCAGAGAGTGTGTGAACTGTGGAGCAACGGCTACTCCACTGTGGCGGAGGGACAGGACA 660
G7R             GCCAGAGAGTGTGTGAACTGTGGAGCAACGGCTACTCCACTGTGGCGGAGGGACAGGACA 121
                ************************************************************

Gata1_ref       GGTCACTACCTGTGCAATGCCTGTGGCTTGTATCACAAGATGAATGGTCAGAACCGGCCT 720
G7R             GGTCACTACCTGTGCAATGCCTGTGGCTTGTATCACAAGATGAATGGTCAGAACCGGCCT 181
                ************************************************************

Gata1_ref       CTCATCCGGCCCAAGAAGCGAATGATTGTCAGCAAACGGGCAGGCACCCAATGCACTAAC 780
G7R             CTCATCCGGCCCAAGAAGCGAATGATTGTCAGCAAACGGGCAGGCACCCAATGCACTAAC 241
                ************************************************************

Gata1_ref       TGTCAAACGACCACTACAACACTCTGGCGGAGGAATGCCAGCGGAGATCCGGTATGCAAT 840
G7R             TGTCAAACGACCACTACAACACTCTGGCGGAGGAATGCCAGCGGAGATCCGGTATGCAAT 300
                ************************************************************

Gata1_ref       GCCTGCGGCCTCTATTTCAAGCTCCATCAGGTGAACCGCCCACTGACCATGAGGAAAGAT 900
G7R             GCCTGCGGCCTCTATTTCAAGCTCCATCAGGTGAACCGCCCACTGACCATGAGGAAAGAT 360
                ************************************************************

Gata1_ref       GGAATCCAGACGAGGAACCGCAAGGCATCTGGCAAAGGGAAAAAGAAGCGGGGGTCGAAT 960
G7R             GGAATCCAGACGAGGAACCGCAAGGCATCTGGCAAAGGGAAAAAGAAGCGGGGGTCGAAT 420
                ************************************************************
```

```
Gata1_ref       CTGGCTGGAGCGGGAGCGGCTGAAGGACCAGCTGGTGGCTTCATGGTGGTAGCTGGTAGC 1020
G7R             CTGGCTGGAGCGGGAGCGGCTGAAGGACCAGCTGGTGGCTTCATGGTGGTAGCTGGTAGC 480
                ************************************************************

Gata1_ref       AGCAGTAGTGGGAATTGTGGGGAGGTGGCCTCAGGCTTGGCACTGGGCACTGCAGGTACT 1080
G7R             AGCAGTAGTGGGAATTGTGGGGAGGTGGCCTCAGGCTTGGCACTGGGCACTGCAGGTACT 539
                ************************************************************

Gata1_ref       GCCCACCTCTATCAGGGCCTAGGACCTGTGGTACTGTCAGGGCCTGTCAGCCATCTTATG 1140
G7R             GCCCACCTCTATCAGGGCCTAGGACCTGTGGTACTGTCAGGGCCTGTCAGCCATCTTATG 599
                ************************************************************

Gata1_ref       CCTTTTCCTGGACCTCTGCTGGGATCGCCTACAACCTCCTTCCCCACCGGTCCTGCACCT 1200
G7R             CCTTTTCCTGGACCTCTGCTGGGATCGCCTACAACCTCCTTCCCCACCGGTCCTGCACCT 659
                ************************************************************

Gata1_ref       ACCACCAGCAGCACCAGCGTGATCGCCCCACTCAGTTCTTGA 1242
G7R             ACCACCAGCAGCACCAGCGTGATCGCCCCACTCAGTTCTTGA 693
                ******************************************
```

pRN3P:Socs3 -HA-1 clone 2 sequence alignment with published mouse *Socs3* gene protein coding sequence (using T_3 promoter in vector)

```
CLUSTAL 2.1 multiple sequence alignment

S23             CRKGGRCTWAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCA 60
mSocs3_ref      ------------------------------------------------------------

S23             GATCTGCCACCATGGGCTACCCATACGATGTTCCTGACTATGCTGTCACCCACAGCAAGT 120
mSocs3_ref      -----------ATG-----------------------------GTCACCCACAGCAAGT 19
                          ***                              ***************

S23             TTCCCGCCGCCGGGATGAGCCGCCCCCTGGACACCAGCCTGCGCCTCAAGACCTTCAGCT 180
mSocs3_ref      TTCCCGCCGCCGGGATGAGCCGCCCCCTGGACACCAGCCTGCGCCTCAAGACCTTCAGCT 79
```

```
                ************************************************************

S23             CCAAAAGCGAGTACCAGCTGGTGGTGAACGCCGTGCGCAAGCTGCAGGAGAGCGGATTCT 240
mSocs3_ref      CCAAAAGCGAGTACCAGCTGGTGGTGAACGCCGTGCGCAAGCTGCAGGAGAGCGGATTCT 139
                ************************************************************

S23             ACTGGAGCGCCGTGACCGGCGGCGAGGCGAACCTGCTGCTCAGCGCCGAGCCCGCGGGCA 300
mSocs3_ref      ACTGGAGCGCCGTGACCGGCGGCGAGGCGAACCTGCTGCTCAGCGCCGAGCCCGCGGGCA 199
                ************************************************************

S23             CCTTTCTTATCCGCGACAGCTCGGACCAGCGCCACTTCTTCACGTTGAGCGTCAAGACCC 360
mSocs3_ref      CCTTTCTTATCCGCGACAGCTCGGACCAGCGCCACTTCTTCACGTTGAGCGTCAAGACCC 259
                ************************************************************

S23             AGTCGGGGACCAAGAACCTACGCATCCAGTGTGAGGGGGGCAGCTTTTCGCTGCAGAGTG 420
mSocs3_ref      AGTCGGGGACCAAGAACCTACGCATCCAGTGTGAGGGGGGCAGCTTTTCGCTGCAGAGTG 319
                ************************************************************

S23             ACCCCCGAAGCACGCAGCCAGTTCCCCGCTTCGACTGTGTACTCAAGCTGGTGCACCACT 480
mSocs3_ref      ACCCCCGAAGCACGCAGCCAGTTCCCCGCTTCGACTGTGTACTCAAGCTGGTGCACCACT 379
                ************************************************************

S23             ACATGCCGCCTCCAGGGACCCCCTCCTTTTCTTTGCCACCCACGGAACCCTCGTCCGAAG 540
mSocs3_ref      ACATGCCGCCTCCAGGGACCCCCTCCTTTTCTTTGCCACCCACGGAACCCTCGTCCGAAG 439
                ************************************************************

S23             TTCCGGAGCAGCCACCTGCCCAGGCACTCCCCGGGAGTACCCCCAAGAGAGCTTACTACA 600
mSocs3_ref      TTCCGGAGCAGCCACCTGCCCAGGCACTCCCCGGGAGTACCCCCAAGAGAGCTTACTACA 499
                ************************************************************

S23             TCTATTCTGGGGGGCGAGAAGATTCCGCTGGTACTGAGCCGACCTCTCTCCTCCAACGTGG 660
mSocs3_ref      TCTATTCTGGGGGGCGAGAAGATTCCGCTGGTACTGAGCCGACCTCTCTCCTCCAACGTGG 559
                ************************************************************

S23             CCACCCTCCAGCATCTTTGTCGGAAGACTGTCAACGGCCACCTGGACTCCTATGAGAAAG 720
mSocs3_ref      CCACCCTCCAGCATCTTTGTCGGAAGACTGTCAACGGCCACCTGGACTCCTATGAGAAAG 619
                ************************************************************
```

```
S23             TGACCCAGCTGCCTGGACCCATTCGGGAGTTCCTGGATCAGTATGATGCTCCACTTTAAT 780
mSocs3_ref      TGACCCAGCTGCCTGGACCCATTCGGGAGTTCCTGGATCAGTATGATGCTCCACTTTAA- 678
                ************************************************************

S23             CTAGAGCGGCCGCGGATCTGGTTACCACTAAACCAGCCTCAAGAACACCCGAATGGAGTC 840
mSocs3_ref      ------------------------------------------------------------
```

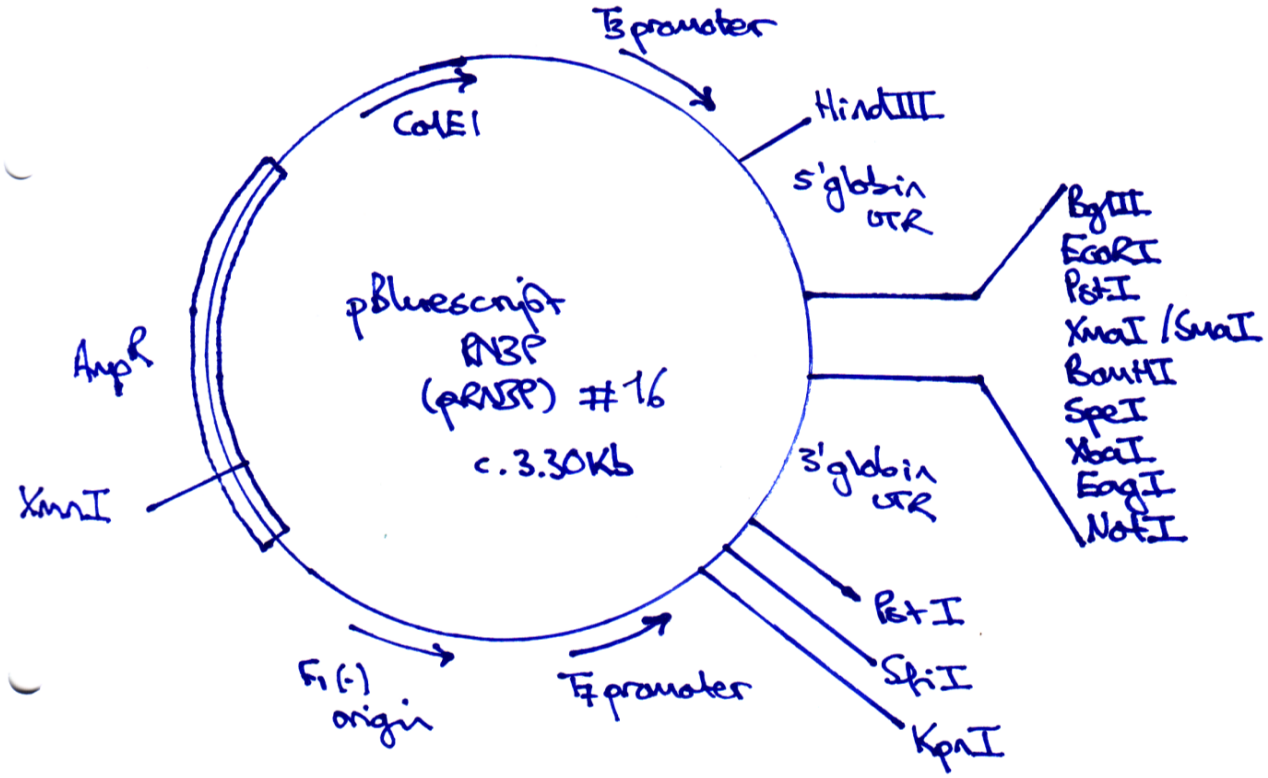
pRN3P:Socs3 -HA-1 clone 2 sequence alignment with published mouse *Socs3* gene protein coding sequence (using T_7 promoter in vector)

```
CLUSTAL 2.1 multiple sequence alignment

Socs3_Ref       -----------------------------------ATG----------------------
S7T             CAGAATAAACGCTCAACTTTGGCAGATCTGCCACCATGGGCTACCCATACGATGTTCCTG 420

Socs3_Ref       --------GTCACCCACAGCAAGTTTCCCGCCGCCGGGATGAGCCGCCCCCTGGACACCA 55
S7T             ACTATGCTGTCACCCACAGCAAGTTTCCCGCCGCCGGGATGAGCCGCCCCCTGGACACCA 480
                        **************************************************

Socs3_Ref       GCCTGCGCCTCAAGACCTTCAGCTCCAAAAGCGAGTACCAGCTGGTGGTGAACGCCGTGC 115
S7T             GCCTGCGCCTCAAGACCTTCAGCTCCAAAAGCGAGTACCAGCTGGTGGTGAACGCCGTGC 540
                ************************************************************

Socs3_Ref       GCAAGCTGCAGGAGAGCGGATTCTACTGGAGCGCCGTGACCGGCGGCGAGGCGAACCTGC 175
S7T             GCAAGCTGCAGGAGAGCGGATTCTACTGGAGCGCCGTGACCGGCGGCGAGGCGAACCTGC 600
                ************************************************************

Socs3_Ref       TGCTCAGCGCCGAGCCCGCGGGCACCTTTCTTATCCGCGACAGCTCGGACCAGCGCCACT 235
S7T             TGCTCAGCGCCGAGCCCGCGGGCACCTTTCTTATCCGCGACAGCTCGGACCAGCGCCACT 660
                ************************************************************

Socs3_Ref       TCTTCACGTTGAGCGTCAAGACCCAGTCGGGGACCAAGAACCTACGCATCCAGTGTGAGG 295
S7T             TCTTCACGTTGAGCGTCAAGACCCAGTCGGGGACCAAGAACCTACGCATCCAGTGTGAGG 720
                ************************************************************
```

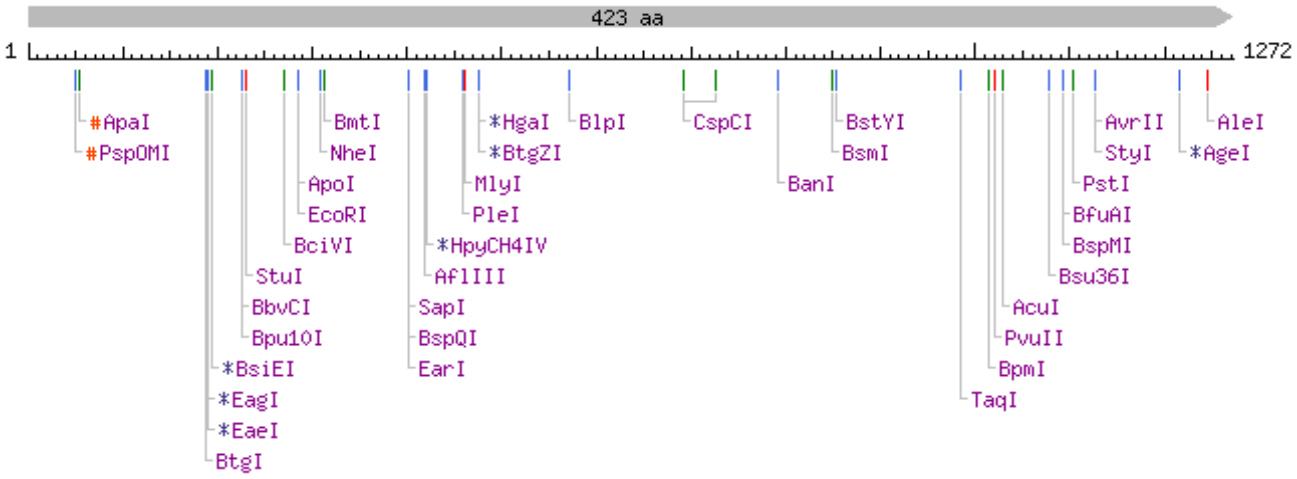
```
Socs3_Ref       GGGGCAGCTTTTCGCTGCAGAGTGACCCCCGAAGCACGCAGCCAGTTCCCCGCTTCGACT 355
S7T             GGGGCAGCTTTTCGCTGCAGAGTGACCCCCGAAGCACGCAGCCAGTTCCCCGCTTCGACT 780
                ************************************************************

Socs3_Ref       GTGTACTCAAGCTGGTGCACCACTACATGCCGCCTCCAGGGACCCCCTCCTTTTCTTTGC 415
S7T             GTGTACTCAAGCTGGTGCACCACTACATGCCGCCTCCAGGGACCCCCTCCTTTTCTTTGC 840
                ************************************************************

Socs3_Ref       CACCCACGGAACCCTCGTCCGAAGTTCCGGAGCAGCCACCTGCCCAGGCACTCCCCGGGA 475
S7T             CACCCACGGAACCCTCGTCCGAAGTTCCGGAGCAGCCACCTGCCCAGGCACTCCCCGGGA 900
                ************************************************************

Socs3_Ref       GTACCCCCAAGAGAGCTTACTACATCTATTCTGGGGGCGAGAAGATTCCGCTGGTACTGA 535
S7T             GTACCCCCAAGAGAGCTTACTACATCTATTCTGGGGGCGAGAAGATTCCGCTGGTACTGA 960
                ************************************************************

Socs3_Ref       GCCGACCTCTCTCCTCCAACGTGGCCACCCTCCAGCATCTTTGTCGGAAGACTGTCAACG 595
S7T             GCCGACCTCTCTCCTCCAACGTGGCCACCCTCCAGCATCTTTGTCGGAAGACTGTCAACG 1020
                ************************************************************

Socs3_Ref       GCCACCTGGACTCCTATGAGAAAGTGACCCAGCTGCCTGGACCCATTCGGGAGTTCCTGG 655
S7T             GCCACCTGGACTCCTATGAGAAAGTGACCCAGCTGCCTGGACCCATTCGGGAGTTCCTGG 1080
                ************************************************************

Socs3_Ref       ATCAGTATGATGCTCCACTTTAA---------------------- 678
S7T             ATCAGTATGATGCTCCACTTTAATCTAGAGCGGCCGCGATCGACA 1125
                ***********************
```

Appendix 2: Plasmid maps of the cloned genes pRN3P:Zbtb32-HA-N, Gata1-HA-N and Socs3-HA-N with *in silico* restriction digest of the insert show (NEB cutter). The 'CB No.' denotes the plasmids catalogue in the our lab's plasmid archive

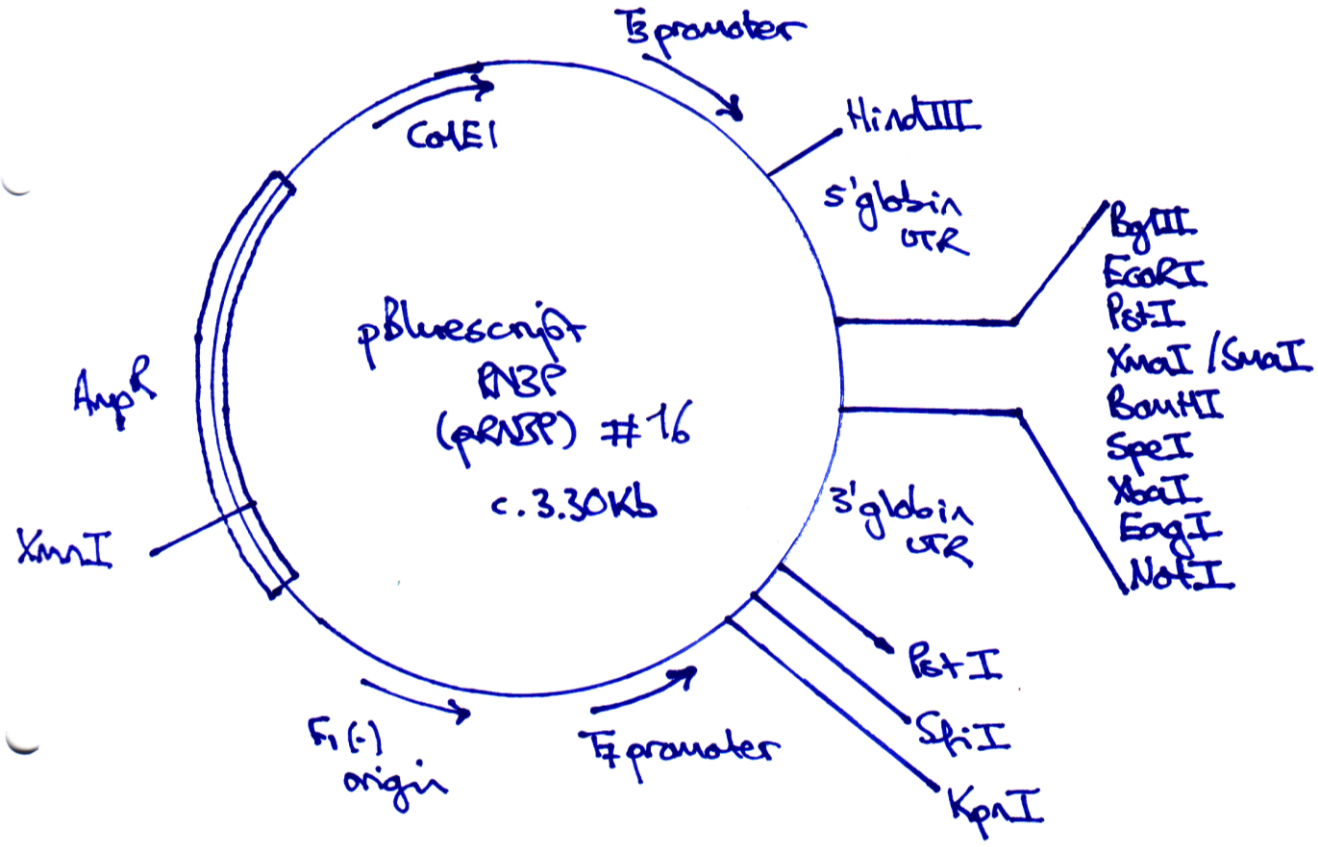
pRN3P:mGata1-HA-N (CB N^{o}.60)



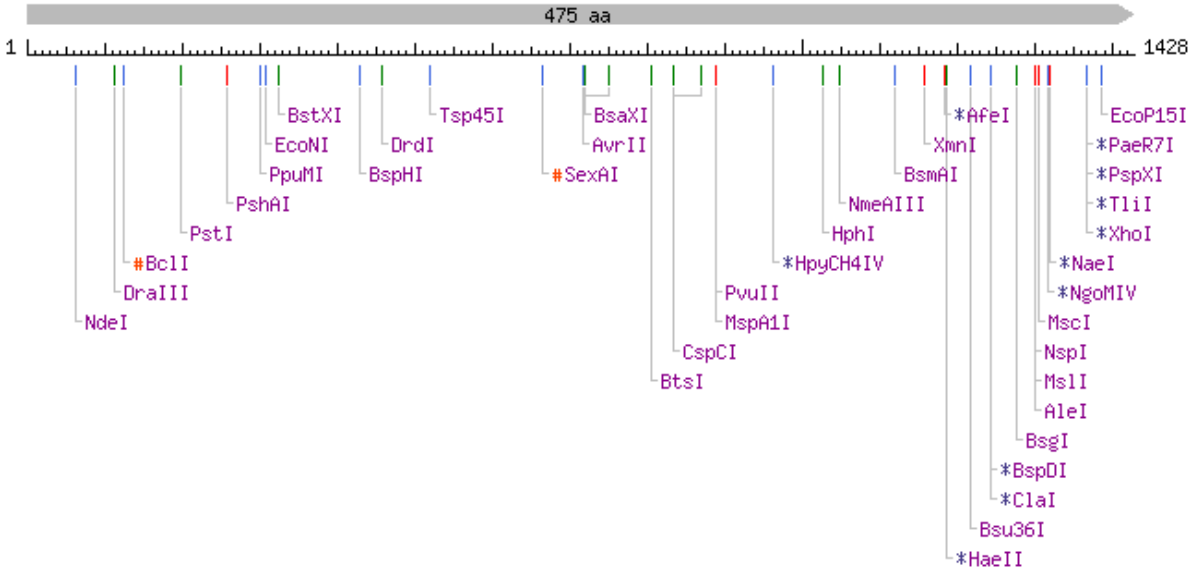
Cloned into *BglII* & *XbaI* sites



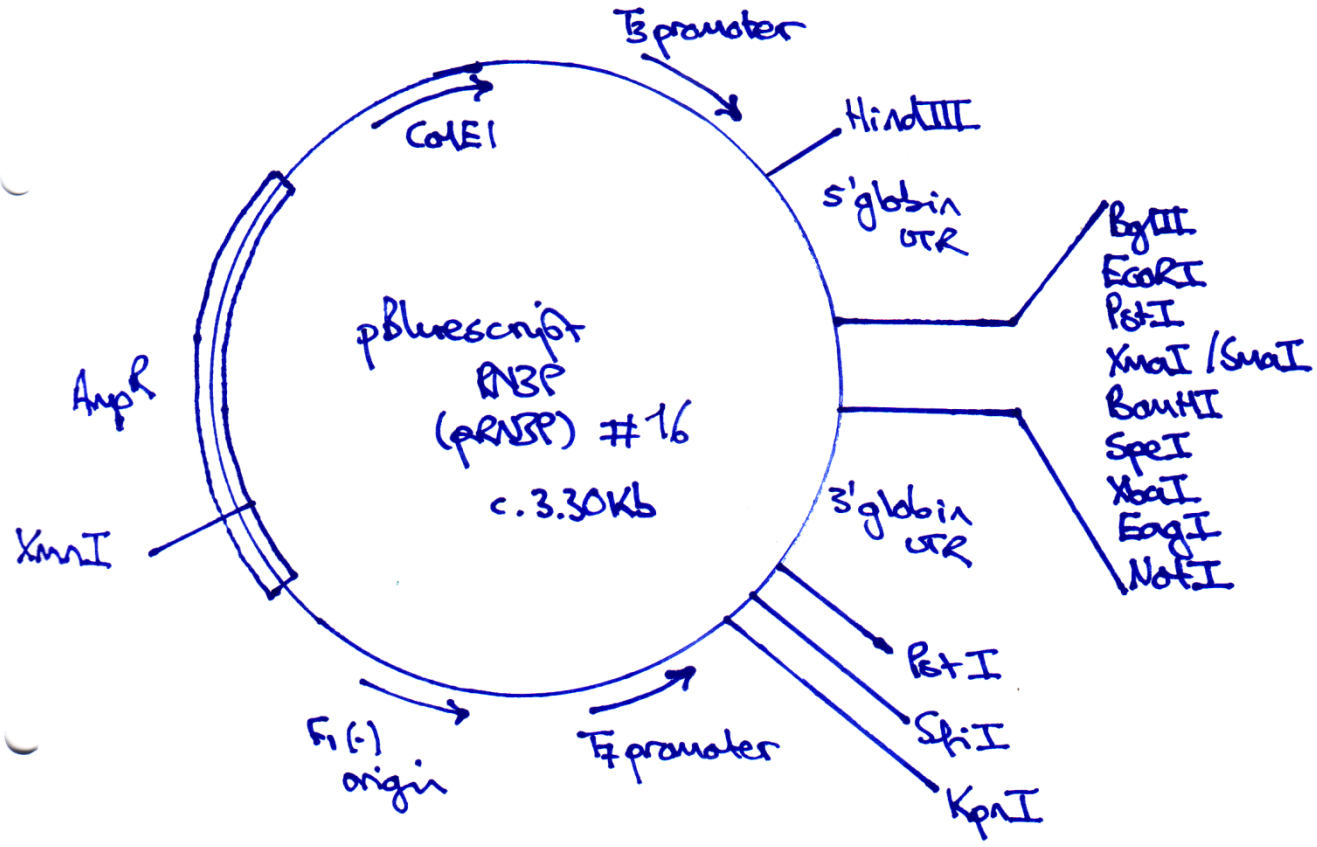
pRN3P:mZbtb32-HA-N (CB N°59.)



Cloned into *BglII* & *XbaI* sites



pRN3P:mSocs3-HA-N (CB N°.58)



Cloned into *BglII* & *XbaI* sites

8 Bibliography

1. Bruce AW*,* Zernicka-Goetz *M:* Developmental control of the early mammalian embryo: Competition among heterogeneous cells that biases cell fate. *Curr Opin Genet Dev* 2010, **20***:* 485–491.

2. Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, Rossant J: Cdx2 is required for correct cell fatespecification and differentiation of trophectoderm in the mouse blastocyst. *Development* 2005, **132**: 2093-2102.

3. Tarkowski AK, Wroblewska JJ: Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage.*Embryol. Exp. Morphol* 1967, **18:** 155-180.

4. Chazaud C, Yamanaka Y, Pawson T, Rossant J: Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Dev. Cell 2006*, **10**: 615-624.

5. Yamanaka Y, Lanner F, Rossant J: FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* 2010, **137**: 715-724.

6. Bruce AW: Identification and characterization of novel cell-fate influencing genes during pre-implantation mouse development. *Czech Sci Found* Part C.

7. Jedrusik A, Parfitt DE, Guo G, Skamagki M, Grabareck JB, Johnson PR, Zernicka-Goetz M: Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes Dev* 2008, **22**: 2692-2706.

8. Johnson MH, McConnell JML: Lineage allocation and cell polarity during mouse embryogenesis. *Semim Cell Dev Biol* 2004, **15**: 583–597.

9. Rossant J: Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. *Semin Cell Dev Biol* 2004, **15**: 573–581.

10. Zernicka-Goetz M: First cell fate decisions and spatial patterning in the early mouse embryo. *Semin Cell Dev Biol* 2004, **15**: 563–572.

11. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S: The HomeoproteinNanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* 2003, **113**: 631-642.

12. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J: Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. *Cell* 2005, **123**: 917-929.

13. Nishioka N, Inoue K, Adachi K, Kiyonari H, Ota M, Ralston A, Yabuta N, Hirahara S, Stephenson RO, Ogonuki N, Makita R, Kurihara H, Morin-Kensicki E, Noyima H, Rossant J, Nakao K, Niwa H, Sasaki H: The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Dev Cell* 2009, **16**: 398-410.

14. Russ AP, Wattler S, Colledge WH, Aparicio SAJR, Carlton MBL, Pearce JJ, Barton SC, AzimSurani M, Ryan K, Nehls MC, Wilson V, Evans MJ: Eomesodermin is required formouse trophoblast development and mesoderm formation. *Nature* 2000, **404**: 95-99.

15. Home P, Ray S, Dutta D, Bronshteyn I, Larson M, Paul S: GATA3 is selectively expressed in the trophectoderm of peri- implantation embryo and directly regulates Cdx2 gene expression. *J Biol Chem* 2009, **284**: 28729-28737.

16. Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, Guo G, Robson P, Draper JS, Rossant J: Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* 2010, **137**: 395-403.

17. Ziomek CA, Johnson MH: Properties of polar and apolar cells from the 16-cell mouse morula. *Rouxs Arch Dev Biol* 1981, **190**: 287–296.

18. Morrisey EE, Tang Z, Sigrist K, Lu M, Jiang F, Ip HS, Parmacek MS: GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* 1998, **12**: 3579-3590.

19. Koutsourakis M, Langeveld A, Patient R, Beddington R, Grosveld F: The transcription factor GATA6 is essential for early extraembryonic development. *Development* 1999, **126**: 723-732.

20. Zernicka-Goetz M, Morris SA, Bruce AW: Making affirm decision multifaceted regulation of cell fat in the early mouse embryo. *Nat Rev Genet* 2009, **10**: 467-477.

21. Yang DH, Smith ER, Roland IH, Sheng Z, He J, Martin WD,Hamilton TC, Lambeth JD, Xu XX: Disabled-2 is essential for endodermal cell positioning and structure formation during mouse embryogenesis. *Dev Biol* 2002, **251**: 27–44.

22. Miner JH, Li C, Mudd JL, Go G, Sutherland AE: Compositional and structural requirements for laminin and basement membranes during mouse embryo implantation and gastrulation. *Development* 2004, **131**: 2247–2256.

23. Gerbe F, Cox B, Rossant J, Chazaud C: Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Dev Biol* 2008, **313**: 594–602.

24. Morris SA, Teo RTY, Li R, Robson P, Glover DM, Zernicka-Goetz M: Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc Natl Acad Sci USA* 2010, **107**: 6364-6369.

25. Guo G, Huss M, Tong GQ, Wang C, Sun LL, Clarke ND, Robson P: Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst. *Dev Cell* 2010, **18**: 675-685.

26. Plusa B, Frankenberg S, Chalmers A, Hadjantonakis AK, Moore CA, Papalopulu N, Papaioannou VE, Glover DM, Zernicka-Goetz M: Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *J Cell Sci* 2005, **118**: 505-515.

27. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A: Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* 2003, **113**: 643-655.

28. Scholer HR, Ruppert S, Suzuki N, Chowdhury K, Gruss P: New type of POU domain in germ line-specific protein Oct-4. *Nature* 1990, **344**: 435-439.

29. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H, Smith A: Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* 1998, **95**: 379-391.

30. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Budge R: Multi potent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003, **17**: 126-140.

31. EllingU, Klasen C, Eisenberger T, Anlag K, Treier M: Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci USA* 2006, **103**: 16319–16324.

32. Zhang J, Zhang J, TamWL, Tong GQ, Wu Q, Chan HY, Soh BS, Lou Y, Yang J, Ma Y, Chai L, Ng HH, Lufkin T, Robson P, Lim B: Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* 2006, **8**: 1114-1123.

33. Torres-Padilla ME, Kouzarides T, Zernicka-Goetz M: Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* 2007, **445**: 214-218.

34. Wu Q, Bruce AW, Jedrusik A, Ellis PD, Andrews RM, Langford CF, Glover DM, Zernicka-Goetz M: CARM1 Is Required in Embryonic Stem Cells to Maintain Pluripotency and Resist Differentiation. *Stem Cells* 2009**, 27**: 2637-2645.

35. Nishioka N, Yamamoto S, Kiyonari H, Sato H, Sawada A, Ota M, Nakao K, Sasaki H: Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev* 2008, **125:** 270-283.

36. Yagi R, Kohn MJ, Karavanova I, Kaneko KJ, Vullhorst D, Depamphillis ML, Buonanno A: Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* 2007, **134**: 3827-3836.

37. Ma GT, Roth ME, Groskopf JC, Tsai FY, Orkin SH, Grosveld F, Engel JD, Linzer DI: GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. *Development* 1997, **124**: 907-914.

38. Kuo CT, Morrisey EE, Anandappa R, Sigrist K, Lu MM, Parmacek MS, Soudais C, Leiden JM: GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* 1997, **11**: 1048-1060.

39. Molkentin JD, Lin Q, Duncan SA, Olson: Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* 1997, **11**: 1061-1072.

40. Kanai-Azuma M, Kanail Y, Gad JM, Tajima Y, Taya C, Kurohmarul M, Sanai Y, Yonekawa H, Yazaki K, Tam PPL, Hayashi Y: Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* 2002, **129**: 2367-2379.

41. Yadav N, Lee J, Kim J, Shen J, Hu MC, Aldaz CM, Bedford MT: Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. *Proc Natl Acad Sci USA* 2003, **100**: 6464–6468.

42. *http://medical-dictionary.thefreedictionary.com/embryonic+regulation* (accessed on 1st December 2012).

43. Online Mendelian Inheritance in Man, OMIM (TM) - *http://www.ncbi.nlm.nih.gov/omim/*. (accessed on 3rd October 2012).