

CARDIFF
UNIVERSITY

PRIFYSGOL
CAERDYDD

Stem Cell Pluripotency

Thesis submitted for the degree of Doctor of Philosophy

at

Cardiff University

by

Susan MacLean Hunter

May 2008

UMI Number: U585101

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U585101

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements:

I would like to thank Professor Sir Martin Evans, not only was he my supervisor and mentor, but also was my friend. PhD students and supervisors do not have to fall out.

I would also like to thank Fiona Mansergh who read this manuscript so thoroughly and has been my friend through thick and thin.

Thank-you also to all the past and present members of the Evans group: Nicky Walker, Julie Wilkins, Marleen Groenen, Rachel Ridgway, Michael Wride and Carl Daly.

Thank-you also to the guys in the office, I am a very messy person and to everyone working on the fourth and fifth floor.

My thanks also to the Joint Services team for their care of the animals and to the staff of Microarray Facility who produced the slides for this project and cared for the scanning equipment.

Finally love to my family, to my Mum who always thought I could do it and to my children, Kirsten and Ken who lived in squalor during the last writing up phase.

Abbreviations and Conventions

The following abbreviations are used throughout the text

A _(x)	Absorbance at x nm
aa-dUTP	aminoallyl-dUTP
aRNA	amplified RNA
ATP	Adenosine triphosphate
AVE	Anterior visceral endoderm
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary DNA
CTP	Cytosine triphosphate
dATP	Deoxy-Adenosine-Triphosphate
dCTP	Deoxy-Cytidine-Triphosphate
dGTP	Deoxy-Guanosine-Triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
d.p.c.	days post coitum
DTT	dithiothreitol
dTTP	Deoxy-Thymine-Triphosphate
dUTP	Deoxy-Uricil-Tri-Phosphate
EC	Embryonal carcinoma (cells)
EDG	equivalent day of gestation
EEct	Embryonic ectoderm
EPL	Early primitive ectoderm like (cells)
ES	Embryonic stem (cells)
EST	Expressed sequence tag
ExE	Extraembryonic ectoderm
FBS	Foetal bovine serum
GTP	Guanosine triphosphate
h.p.c.	hours post coitum
HRP	Horse radish peroxidase
ICM	Inner cell mass
IVT	<i>In vitro</i> transcription
kb	Kilo bases
lbs	pounds
LIF	Leukaemia inhibitory factor
M	Molar concentration
mol	mole
mRNA	Messenger RNA
N	Normal

NBS	Newborn bovine serum
NTP	Ribonucleotide
<i>pc</i>	Post Coitum
PCR	Polymerase chain reaction
PGC	Primordial germ cells
pmt	Photo multiplier tube
RNA	Ribonucleic acid
RNase	Enzyme which degrades RNA
RT-PCR	Reverse transcription polymerase chain reaction
Sq.in.	Square inch
TE	Trophectoderm
TS	Trophectoderm stem cells
UTP	Uracil triphosphate
VE	Visceral endoderm
U	Units of activity
v/v	volume by volume
w/v	weight by volume
ϵ	Extinction coefficient
Φ	Quantum yield

Throughout this thesis, microarraying definitions are as established by Duggan *et al.*, 1999. The DNA present on the spot is referred to as the “probe”, and the labelled cDNA used to hybridise to the microarray as the “target”.

The term “totipotent” is used to describe a cell from which the whole conceptus can be derived. “Pluripotent” applies to cells which can contribute to many different tissues in a whole animal when supported by other cells or can differentiate to cells from multiple lineages (Solter, 2006).

In many publications the abbreviation aRNA is used interchangeably for amplified and antisense since in most cases it is both. In this thesis amplified RNA can be in the sense orientation so this abbreviation has not been used to denote amplified RNA.

Acknowledgements:.....	3
Abbreviations and Conventions.....	2
Summary	9
1. Introduction.....	11
1.1 Origins and Properties of ES Cells.	12
1.1.1 Teratocarcinomas.....	12
1.1.2 Pluripotent Tissues in the Embryo.....	16
1.1.3 Derivation of ES Cells from the Pluripotent Compartment of the Embryo	20
1.1.4 Extrinsic Factors and the Maintenance of Pluripotency in ES cells.	22
1.1.5 Intrinsic Factors and the Maintenance of Pluripotency.	29
1.1.6 EG Cells.....	34
1.1.7 EPL Cells	34
1.1.8 Pluripotent Cells from Human.....	36
1. 2 Microarrays.....	37
1.3 Aims of the Project	39
2. Materials and Methods.....	41
2.1 Cell Culture.....	41
2.1.1 Media.	41
2.1.2 Culture of Embryonic Stem Cells.....	42
2.2. Animal Husbandry, Embryo and Tissue Isolation.....	43
2.2.1. Timed Matings.....	43
2.2.2 Embryo Isolation.....	44
2.2.3 Tissue Isolation.....	45
2.3 Molecular Methods	47

2.3.1 Purification of RNA.....	47
2.3.2. Ethanol Precipitation of Nucleic Acids.....	49
2.3.3 Analysis of Nucleic acids By Optical Density.....	50
2.3.4 Analysis of Labelled Target by Optical Density.....	50
2.3.5. Slide Gels.....	50
2.3.6. Analysis of RNA by gel Electrophoresis.....	51
2.4 Microarray.....	52
2.4.1. Slides.....	52
2.4.2. Blocking Of CMT-GAPS™ Coated Slides	52
2.4.3. Using Target Labelled with Cy3/Cy5	53
2.4.4 Post Hybridization.	53
2.4.5 Scanning.....	54
3. Labelling of Target	56
3. 1 Introduction.....	56
3.1.1 Labelling Molecules.....	57
3.1.2 Labelling of Target	58
3.2 Materials and Methods.....	66
3.2.1 Direct Labelling	68
3.2.2. Indirect Labelling.....	72
3.2.3 Comparison of Arrays Hybridised with Target prepared by the Atlas™ PowerScript™ Fluorescent Labelling Kit, the Template being Destroyed either by RNase H or NaOH.	81
3.3 Results and Discussion	84
3.3.1 Direct Labelling	84

3.3.2 Indirect Labelling: Testing parameters of fluorescently labelled target production by aminoallyl labelling.....	86
3.3.3 Kits.....	91
3.4 Conclusions.....	100
4. Amplification of Target	106
4.1 Introduction.....	106
4.1.1 Signal Amplification.....	106
4.1.2 Sample Amplification.....	107
4.2 Materials and Methods.....	116
4.2.1 Amplification using the method of Wang.....	119
4.2.2 Amplification using Beads.....	123
4.2.3 Amplification using the MessageAmp™ amplified RNA Kit.....	123
4.2.4 Amplification using the RiboAmp™ Kit (Arcturus).....	123
4.2.5 Analysis of Amplification Products Using PCR.....	123
4.3 Results and Discussion	126
4.3.1 Amplification using Oligo-(dT)-T7.....	126
4.3.2 Dynabeads.....	134
4.3.3 MessageAmp™ (Ambion) amplified RNA Kit.....	138
4.3.4 RiboAmp™ (Arcturus) Kit.....	140
4.4 Conclusions.....	145
5. Optimisation of Amplification.....	149
5.1. Introduction.....	149
5.1.1 Sample Preparation	149
5.1.2 Image Analysis.....	151
5.2.3 Normalisation and Data Analysis	151

5.2. Materials and Methods.....	155
5.2.1 Target Preparation.....	155
5.2.2 Slide Preparation and Hybridisation	156
5.2.3 Image Analysis.....	156
5.2.4 Normalisation and Analysis.....	158
5.3 Results and Discussion	160
5.3.1 Amplification and Labelling.....	160
5.3.2 Analysis.....	163
5.4 Conclusions.....	176
6. Embryonic stem cells and transcriptomic analysis	179
6.1 Introduction.....	179
6.1.1. Design of Experiment	183
6.1.2. Image Analysis.....	186
6.1.3. Analysis.....	187
6.1.4. Characteristics of the Embryonic Stages.	187
6.2. Materials and Methods.....	193
6.2.1. Cell Culture.....	193
6.2.2. Reference Pool	193
6.2.3 Collection of Tissues.....	195
6.3. Results and Discussion	195
6.3.1 ES Cells as Reference	195
6.3.2 Pooled Embryonic Tissues as Reference	197
6.6.3 Comparison of ImaGene™ and Spot.....	198
6.3.4 Comparison of ES Cells with Pluripotent Embryonic Tissue.....	210
6.4. Conclusions.....	228

7. Discussion232

 7.1 Labelling of Target232

 7.1.2 Amplification235

 7.1.2 Image Analysis.....237

 7.1.3 Background Subtraction and Normalisation.....239

 7.2 Embryonic Stem Cells240

 7.3 Future Work247

8. Bibliography249

Appendix 1 Developmental Stages.....270

Appendix 2. Formulae used in Analysis of Labelled Target275

Appendix 3: Commands for Analysis in Limma.....277

Appendix 4 Analysis of array data using NIA tool.....284

Summary

Embryonic stem cells (ES cells) are derived by explantation of the embryonic portion of the pre-implantation embryo into culture. These cells have unique properties which have made them invaluable in study of the function of genes *in vivo* and of cell differentiation *in vitro*. They can be grown in culture for extended periods of time in an undifferentiated state and induced to differentiate *in vitro*. While undifferentiated they can be genetically manipulated. Subsequent reintroduction of these cells into the blastocyst results in the cells being integrated and contributing to all the cells of the animal including the germ line thus leading to designed genetic change.

The homology of these cells, however, to their tissue of origin is not unambiguous. The primary aim of this thesis was to apply global transcriptome analysis to investigate the homology of ES cells to the pluripotent compartment of the embryo.

Although ES cells can be grown in bulk, the tissue of origin, the embryonic portion of the peri-implantation embryo are small and inaccessible. It was therefore necessary to develop methods which would allow the transcriptome to be amplified without distorting the transcript profile. A linear amplification method proved to give the best result. The best method for fluorescently labelling the cDNA was shown to be enzymatic incorporation of aminoallyl dUTP followed by coupling to monoreactive Cy dyes. With these tools it was then possible to amplify the transcriptome of both colonies of ES cells and the embryonic portion of various peri-implantation embryos and apply the labelled cDNA to microarray slides. Statistical analysis of the results proved that the transcriptome of ES cells most resembles that of the embryonic ectoderm on day 5.5 of development.

Chapter 1

Introduction

1. Introduction

Embryonic stem cells (ES cells) are pluripotent cells isolated into culture from the pluripotent compartment of pre-implantation stage embryos (Evans and Kaufman, 1981; Martin, 1981). Under suitable conditions, they are capable of indefinite growth in culture in an undifferentiated state. They can be induced to differentiate *in vitro* and most importantly, when reintroduced into the embryonic environment of the blastocyst, will reintegrate into the embryo contributing to all tissues of the resulting adult mouse including the germ line. Because these cells can be grown in an undifferentiated condition, we can manipulate their genetic constitution (Thomas and Capecchi, 1987). On reintroduction to a blastocyst, lines of mice with precise genetic changes can be produced.

In the 25 years since murine embryonic stem cells were first isolated, a large technological base has evolved which exploits these unique features. Despite this, there is very little known about the origin of ES cells and their relationship to pluripotent stem cells *in vivo*. The need to fill this knowledge gap has become more pressing with the isolation of human cell lines that share some properties of the mouse cells.

The recent advent of microarray technology (Schena *et al.*, 1995) and RNA amplification (Van Gelder *et al.*, 1990) allows us to compare gene expression profiles of thousands of genes from small samples of cells and tissue fragments. Similarities and differences in gene expression between samples will allow us to provide a molecular phenotype: a “fingerprint” which should allow assignment of identity of samples and reveal relationships between them.

1.1 Origins and Properties of ES Cells.

1.1.1 Teratocarcinomas.

The successful derivation and culture of ES cells rests firmly on the work carried out with teratocarcinomas and their derivative cell lines. Teratocarcinomas are malignant tumours most commonly found in the gonads. These tumours are composed of both malignant undifferentiated cells and a wide range of non-malignant differentiated tissues. Male mice of the 129 strain and female mice of the LT strain have a higher incidence of such tumours. In the case of 129 male mice the tumours were believed to be of germ cell origin. Confirmation of the germ cells ability to produce such tumours was provided by experiments where genital ridges from embryos on day 11.5–12.5 of gestation were transplanted to the adult testis (Stevens, 1964). Homozygous embryos of mice carrying the Sl^J or Sl^D alleles of the steel gene are nearly devoid of primordial germ cells (Bennett, 1956). Stevens (1967) backcrossed these alleles on to the 129/Sv background, took genital ridges from homozygous mutant embryos on day 12 of gestation and grafted them into the testis capsule of normal 129/Sv mice. The mutant embryos did not develop teratocarcinomas while the heterozygous littermates did. By using a recombinant inbred line between C57Bl6J and the LT/Sv strain female mice which developed ovarian teratomas at a high rate and were heterozygous for the electrophoretic marker *Gpi-1* ($Gpi-1^a/Gpi-1^b$); Eppig et al (1977) showed that the tumours arose from parthenogenetically activated oocytes initiating development in the ovary.

It was further demonstrated that the transplantation to ectopic sites of pre-implantation embryos up to the 8-cell stage resulted in teratocarcinomas (Stevens, 1968). Grafts of post-implantation embryos up to day 8 of gestation before

gastrulation is completed, also gave rise to teratocarcinomas. When post-gastrulation stage embryos were transplanted no teratocarcinomas were found (Solter *et al.*, 1970; Stevens, 1968; Stevens, 1970). Transplantation of the isolated tissues from embryos on day 6 of gestation demonstrated that only embryonic ectoderm gave rise to teratocarcinomas. Extra embryonic ectoderm only resulted in invasive giant trophoblast tumours and endoderm did not develop (Diwan and Stevens, 1976).

Evans (1972) showed that the undifferentiated component of spontaneously occurring teratocarcinomas could be expanded *in vitro*. Single cells from these cultures could be expanded clonally (Martin and Evans, 1974). Furthermore, these lines of embryonal carcinoma (EC) cells could be differentiated *in vitro* as embryoid bodies (Adams *et al.*, 1991). Further, lines of clonal origin could both be expanded and differentiated producing cell types from all three primary germ layers: ectoderm, mesoderm and endoderm demonstrating the pluripotent nature of these cells (Martin and Evans, 1975b). Brinster (1974) produced the first evidence that EC cells taken from embryoid bodies grown *in vivo* were able to contribute to the development of a chimaera when introduced into the blastocyst of a host embryo. Illmensee and Mintz (1976) reported extensive chimaerism including germ line from similar experiments. However, because these cells are derived by passing through an *in vivo* phase as a tumour in an ectopic site, and are subjected to the pressures of this abnormal environment, they are typically aneuploid. The ascites tumour cell line used by Illmensee and Mintz – OTT 6050 – does not contain a Y chromosome and their reported results are unrepeatable and considered unreliable. In later experiments EC cells were derived from a karyotypically normal tumour which had undergone relatively few subdivisions thus maintaining the normal karyotype before injection into blastocysts. These cells can contribute to the germ line of a chimera (Stewart and

Mintz, 1981). EC cells from in vitro cultured cell lines were shown to be able to contribute to extensive chimaerism but these mice often develop early teratocarcinomas or late tumours of differentiated cell type derived from the tissue culture cells (Papaioannou *et al.*, 1975).

Further evidence that these cells were a paradigm for the undifferentiated cells of the early embryo was provided by studies on markers expressed by both cells of the embryo and EC cells. Alkaline phosphatase is expressed by the inner cell mass (ICM) and EC cells (Jacob, 1978). An anti-F9 cell antibody (Artzt *et al.*, 1973) was made by injecting F9, an EC line originating from a 129 strain mouse, into a 129 mouse. This antibody reacted with the cell surface of EC cells and with the ICM of the blastocyst. The advent of monoclonal antibodies led to the production of the antibody to Stage Specific Embryonic Antigen-1 (SSEA1). This monoclonal antibody was made by inoculation of mice with F9 cells before fusing the lymphocytes from the spleens to a myeloma cell line. The antibody reacts to an antigen present on the cell surface of the morula stage embryo, the ICM of the blastocyst and the undifferentiated component of teratocarcinomas and EC cell lines (Solter and Knowles, 1978). Gooi *et al* (1981) showed the carbohydrate nature of the antigen to which SSEA1 and other monoclonal antibodies raised in a similar way react. All these observations reinforced the belief that undifferentiated cell lines could be derived directly from an embryo without the intermediate stage of tumour formation (Evans, 1981). ES cells are derived by explanting into culture the embryonic portion of the pre-implantation embryo.

However their homology to the pluripotent compartment of the embryo is not clear. Dewey *et al* (1978), using 2-d electrophoresis, carried out comparisons of pluripotent EC cells with pre-implantation embryos and detected both similarities and

dissimilarities to all stages examined. Martin *et al* (1978) also utilised 2-d electrophoresis to compare the proteome of EC cells with isolated ICMs and embryonic ectoderm and not only showed an overall homology of EC cells to the embryonic ectoderm but also showed the presence of a single protein uniquely held in common between this peri-implantation stage and EC cells suggesting identity with this stage. In a more detailed analysis Lovell-Badge *et al* (1980) showed substantial differences between EC cells and the ICM using 2-d protein electrophoresis. In this study the EC line OTT5568 S PCMB was grown in monolayer without feeders and differentiated as embryoid bodies. The undifferentiated cells and embryoid bodies at various times post formation were run on 2d-gels. Cells from the nullipotent EC line, Nulli SCC1 (Martin and Evans, 1975a), the endodermal line PSA5-E and STO, the feeders used to routinely maintain the OTT5568 S PCMB cells were also analysed. Isolated ICM from early blastocysts at about 86hpc, were also analysed. Comparison of EC cells, ICMs and STO cells showed a degree of dissimilarity between EC cells and the ICM comparable to the dissimilarity between EC cells and STOs. When the gels from the ICMs were examined for the presence of spots which varied over the time course of embryoid body differentiation, they appeared to have more spots in common with embryoid bodies 6-12 hours after initiation.

It was shown that the Forsmann antigen was expressed by the ICM from both early and late blastocysts (Willison and Stern, 1978) using immunofluorescence. Further to this work Stinnakre (1981) showed that although Forsmann antigen was expressed by the early post-implantation epiblast, as the epiblast developed further to form the embryonic ectoderm, expression of the Forsmann antigen became restricted to the endoderm. These observations taken together have suggested a homology of EC cells with the epiblast of the late peri-implantation embryo rather than the early ICM or the

embryonic ectoderm. However, the work of Gardner (1985) has shown by injecting isolated ICM and epiblast that over the peri-implantation period the ability to contribute to chimaeras drops dramatically giving half the chimaerism rates with epiblast from day 5 of gestation embryos than from preimplantation stage ICMs and no chimaeras have been reported from injection of ectoderm cells from day 6.5 days post coitum (d.p.c.) into blastocysts. Furthermore, Wells *et al* (1991) demonstrated that over this time period, the rate of derivation of ES lines dropped from 13.2% for 3.5-d.p.c embryos to 2.5% for epiblast from embryos 5.5-d.p.c.

The belief that the equivalent of EC cells could be derived directly from the embryo without a period as a tumour was converted to reality with the landmark work of Evans and Kaufman (1981). EC cell lines were used to optimise culture conditions which allowed ES cells to be derived. An important observation arising from the initial maintenance of pluripotentiality *in vitro* (Evans, 1972) was that in most cases EC cells grew poorly except in mixed cultures with their differentiated derivatives. This led to the use of mitotically inactivated fibroblasts as feeders (Martin and Evans, 1974).

1.1.2 Pluripotent Tissues in the Embryo.

The totipotent fertilised egg undergoes a series of cleavage divisions. These are characterised by the cells becoming successively smaller with each round of cell division and little change in the size of the embryo until expansion of the blastocyst cavity. It has been revealed that over the preimplantation period the embryo switches from maternal mRNA stored in the oocyte to zygotic mRNA transcribed *de novo* from the genome in three waves (Hamatani *et al.*, 2004a); a minor wave on fertilisation followed by two waves, the major-ZGA (zygote gene activation) the at 2-4 cell stage

and mid-preimplantation gene activation (MGA) that peaks at 8 cell stage and is complete by transition of the morula to blastocyst.

After the 3rd cleavage division, the eight cell stage, a process of flattening and cell polarisation called compaction begins (Pratt *et al.*, 1982). The 4th cleavage divisions can be either conservative resulting in the production of two polar cells or differentiative resulting in one polar cell and one non polar (Johnson and Ziomek, 1981; Sutherland *et al.*, 1990). Thus a morula, made up of two cells types which on the basis of position and morphology can be distinguished, is formed. The cells with an outside position are polarised with a smooth surface and a villus surface preparatory to forming an epithelium, the trophoctoderm (TE). Those with an internal position have no villus cap. Although predictions can be made about the cellular fate of these cells in the undisturbed embryo, experimental manipulation has shown the considerable capacity of the embryo for regulation.

As the term totipotent is usually reserved for the ability to generate the whole conceptus, cells of the four and eight cell stage will be described as pluripotent since an individual cell cannot develop to form a complete conceptus (Solter, 2006). However, single cells at the eight cell morula stage have been shown to retain their pluripotent properties by aggregation of single cells with blastomeres expressing electrophoretically distinct markers and implantation into pseudopregnant foster mothers. Analysis of the post implantation conceptus shows cells contribute to all tissues including the extraembryonic membranes which are only derived from the trophoctoderm (Kelly, 1977). More recently, triplets have been generated by aggregation of single eight cell blastomeres with electrophoretically distinct tetraploid blastomeres (Tarkowski *et al.*, 2005). So far it has not been possible to produce a

conceptus with less than two aggregated blastomeres from 16-cell stage embryos, however aggregates made up of either 16 internal apolar blastomeres from 16-cell stage embryos, or from 16 outer polar blastomeres will develop to term (Ziomek *et al.*, 1982).

The first differentiation event in the pre-implantation embryo is the formation of the trophoctoderm (TE) from the outer polar cells of the morula. Blastulation begins towards the end of the 5th cleavage division (Smith and McLaren, 1977). After implantation the TE, which expresses FGFR2 (fibroblast growth factor receptor 2), overlying the ICM proliferates in response to FGF4 (fibroblast growth factor 4) which is produced by the epiblast (Rossant and Cross, 2001). This polar TE grows into a column of extraembryonic ectoderm (ExEct) which pushes down into the blastocoel cavity with the epiblast at its distal end. The mural TE is restricted to the outer surface of Reichert's membrane while the ICM differentiates further to produce a layer of primitive endoderm on its blastocoelic surface. This primitive endoderm migrates round the inner surface of the trophoctoderm to form the parietal endoderm (that manufactures the extracellular matrix which makes up the Reichert's membrane) while the endoderm which remains in contact with the ICM becomes the visceral endoderm (VE). Over this peri-implantation period the ICM is proliferating and is often at this stage referred to as the epiblast. Experiments with immunosurgically isolated ICM from early cavitating blastocysts show that at this early stage the ICM retains the ability to form trophoctoderm (Handyside, 1978) but this property is lost and by the late preimplantation stage isolated ICMs only differentiate endoderm.

Apoptotic signals from the VE and survival signals from the basement membrane trigger the epiblast to form a pro-amniotic cavity and to transform into a pseudo-

stratified epithelium; the embryonic ectoderm (EEct)(Coucouvanis and Martin, 1995). The distal VE at this time thickens and some genes become expressed in a regional manner (Pfister *et al.*, 2007). Labelling of individual cells of the epiblast with HRP (Lawson *et al.*, 1991) showed that pluripotency was maintained and the labelled cells contributed to all three germ layers of the mid-streak embryo. Movement of the distal visceral endoderm (DVE) cells to form an anterior signalling centre, the AVE, is believed to be the symmetry breaking event which determines the anterior end of the embryo (Beddington and Robertson, 1998; Thomas *et al.*, 1998).

The process of embryogenesis is controlled by complex networks of growth factors and cytokines. The delineation of the TE requires expression of *Cdx-2* (caudal type homeo box 2) (Strumpf *et al.*, 2005) during compaction and *eomes* (eomesodermin) expression is required to transform TE to trophoblast (Russ *et al.*, 2000). The ICM expresses *Pou5f1* (pou domain, class 5, transcription factor 1), *sox2* (SRY-box containing gene 2) and *nanog* (Nanog homeobox) (Avilion *et al.*, 2003; Chambers *et al.*, 2003; Mitsui *et al.*, 2003; Nichols *et al.*, 1998; Niwa *et al.*, 2000). Gradients of ligands for the BMP, WNT and NODAL pathways are established from the junction of the EEct and ExEct *i.e.* the proximal EEct and distal ExEct, ligands appear to be widely expressed throughout the EEct and VE while signals from the DVE endoderm cells normally antagonize signals received by the proximal epiblast and thus act to maintain anterior identity (Beddington and Robertson, 1999) Nodal antagonists such as *Lefty1* and *Cer1* are expressed in the DVE (Brennan *et al.*, 2001). The nodal gradient becomes restricted to the posterior of the embryo where the primitive streak will initiate. Embryos that lack nodal fail to gastrulate and are deficient in mesoderm (Conlon *et al.*, 1994; Zhou *et al.*, 1993).

1.1.3 Derivation of ES Cells from the Pluripotent Compartment of the Embryo

ES cells are directly isolated into culture without transformation or immortalisation and can be maintained for many generations without change to the karyotype. They are morphologically indistinguishable from EC cells and furthermore, express SSEA-1 and alkaline phosphatase. They also differentiate via embryoid bodies. Most importantly, on injection into a blastocyst, they give chimaeras at a high rate and only rarely give rise to teratocarcinomas. Furthermore, the cells are integrated into the gametes and are transmitted to subsequent generations.

The first ES lines were isolated by plating out whole blastocysts from 129SvE mice which had undergone a period of diapause (Evans and Kaufman, 1981). Embryonic diapause is an adaptation which allows blastocysts to survive for several weeks without implantation while the mother has a suckling litter. When the litter is weaned the blastocysts implant and development proceeds to term. The state of delay can be experimentally induced by removing the source of oestrogen and providing exogenous progesterone. After the blastocysts had attached to the culture surface and outgrown for a few days, the clump of embryonic cells was picked, dissociated with trypsin and plated on to mitotically inactivated feeder cells. Blastocysts which have undergone a period of delay of implantation appear to give an improved rate of forming ES lines although it is not a requirement. Lines can also be derived from ICM's which have been isolated by immunosurgery (Martin, 1981; Solter and Knowles, 1975).

The success rate in isolating ES lines appears to have a strong genetic component with the most efficient strain being 129 which yields lines at up to 30% (Robertson,

1987). The other commonly used strains which have given rise to ES lines, C57 Bl/6 and BALB/C, yield lines at much lower rates. Mice of the strain CBA have been considered non-permissive for ES line derivation, although this has been shown not to be the case (Brook and Gardner, 1997). Interestingly, formation of EC is much less affected by genetic composition (Damjanov *et al.*, 1983). This has led to various strategies to enhance the rate of success. Only two factors so far convincingly improve success rates. Delay of implantation clearly plays a role in the successful derivation of ES lines. Although this was believed to be due to a larger number of cells in the ICM the increase is marginal. Copp (1982) showed that the ICM of delayed blastocysts on both the equivalent day of gestation (EDG) 5 and 10 contained approximately 29 cells and the ICM of non delayed blastocysts 4 days and 17h *post coitum* contained, shortly before implantation, about 43 cells. Nichols *et al* 2001 showed that delayed ICMs on EDG-6.5 and 8.5 had approximately 32 and 42 cells respectively and are not significantly different from each other.

The inner cell mass of the late blastocyst stage has a layer of primitive endoderm on the blastocoelic surface. When ICMs are isolated by immunosurgery, this layer grows in culture to surround the whole ICM. The primitive endoderm produces a layer of basement membrane which is involved in the conversion of the epiblast to a pseudo columnar epithelium and also produces a signal leading to the apoptotic events which result in the formation of the pro-amniotic cavity (Coucouvanis and Martin, 1995). This event has been generally held as poor prognostically when isolating ES lines (Robertson, 1987). It has been suggested that denuding the ICM of endoderm removes signals which promote differentiation (Handyside *et al.*, 1989). Furthermore Brook and Gardner (1997) demonstrated that microsurgical removal of both the

trophectoderm and endoderm from delayed blastocysts facilitates the isolation of ES lines including non-permissive strains such as CBA/Ca.

The earliest stage at which ES cell lines have been established is from 8-cell compacted morula (Tesar, 2005). The latest reported stage at which ES lines have been derived is the epiblast of day 5.5 hatched, post-implantation egg cylinder stage embryo (Wells *et al.*, 1991).

1.1.4 Extrinsic Factors and the Maintenance of Pluripotency in ES cells.

Several cytokines have been shown to have relevance in the maintenance of ES cells *in vitro*. However their role in maintenance of the pluripotent compartment *in vivo* is less clear. LIF, BMPs, WNTs and FGFs have all been demonstrated to play a role in the balancing act between differentiation and pluripotency in ES cells.

The early work of Martin and Evans (1974) indicated that feeder cells produced a trophic factor that supported undifferentiated growth of the co-cultured ES cells. This was confirmed when it was shown that feeders could be replaced by conditioned media from the STO feeder cells or from Buffalo Rat Liver cells (BRL cells) (Smith and Hooper, 1987; Smith and Hooper, 1983). Subsequently the active factor was isolated and shown to be a single cytokine; leukemia inhibitory factor (LIF) (Smith *et al.*, 1988; Williams *et al.*, 1988). Further evidence that LIF is the relevant factor is provided by the experiments which showed that feeders lacking a functional *Lif* gene do not support ES proliferation in an undifferentiated state (Stewart *et al.*, 1992). Removal of LIF from the culture medium of ES cells in monolayer culture leads to differentiation to primitive endoderm like cells (Niwa *et al.*, 1998; Niwa *et al.*, 2000). It has been shown that ES cells can be maintained for at least a short period of time

without a source of LIF (Berger and Sturm, 1997; Dani *et al.*, 1998) suggesting the presence of an unidentified factor or factors in serum which can repress the differentiation of ES cells to endoderm.

LIF is a member of a family of cytokines related to interleukin-6 (IL-6) which have multiple functions and usually drive differentiation of adult stem cells. It was later shown that undifferentiated growth of ES cells could be supported by other members of the IL-6 family, which act via the gp130 receptor (Conover *et al.*, 1993; Pennica *et al.*, 1995; Yoshida *et al.*, 1996). LIF, oncostatin-M (OSM) and cardiotrophin-1 (CT-1) bind to the leukaemia inhibitory factor receptor α (LIFR α) forming a heterodimeric receptor complex consisting of LIFR α and gp130. Ciliary neurotrophic factor (CNTF) requires the CNTFR (also expressed by ES cells) to engage the LIFR-gp130 heterodimer (Stevens, 1970). IL-6, whose receptor is not expressed by ES cells, acts independently of LIFR α through a hexameric complex containing a homodimer of gp130 and two IL-6 and IL-6R modules (Murakami *et al.*, 1993; Nichols *et al.*, 1994; Yoshida *et al.*, 1994; Yoshida *et al.*, 1996). Thus, gp130 is the common receptor component involved in cytokine mediated ES cell renewal which was originally identified as the signal transducing subunit of the IL-6 receptor complex (Hibi *et al.*, 1990; Taga *et al.*, 1989).

Gp130, a member of the cytokine receptor superfamily (Bazan, 1990) has no tyrosine kinase activity and so recruits the Janus kinase (JAK), a non receptor tyrosine kinase (Darnell *et al.*, 1994). The JAK kinases are believed to be associated with the unpolymerised receptor modules. Ligand binding induces oligonucleotidimerization and by a largely unknown mechanism, brings the associated JAK molecules into close proximity culminating in phosphorylation and activation of the receptor complex

(Stahl and Yancopoulos, 1994). In ES cells JAK-1, JAK-2 and TYK-2 are activated in response to LIF stimulation (Ernst *et al.*, 1996). The phosphorylated tyrosines on the receptor form binding sites for proteins containing the Src-homology (SH-2) domain, in ES cells STAT-3 (signal transducer and activator of transcription 3) which are in their turn phosphorylated by the JAK kinases and released from the receptor complex as dimers. On dimerization, they are translocated to the nucleus where they bind to the DNA and direct specific gene expression.

The Gp130/JAK receptor complex in ES cells direct two major pathways; a stem cell renewal pathway through STAT-3 and a pathway which promotes cell differentiation via SHP-2 (also known as Ptpn11, protein tyrosine phosphatase, non-receptor type 11) (Burdon *et al.*, 1999). It has been shown that activated STAT-3 is sufficient to support stem cell renewal in the absence of LIF (Matsuda *et al.*, 1999) and cells stably expressing a dominant negative form of STAT-3 cannot be maintained in an undifferentiated form (Niwa *et al.*, 1998)

On the other hand it has been shown that SHP-2, which by associating with either GRB-2 (growth factor receptor bound protein 2) and SOS (son of sevenless) or GAB-1 (GRB-2 associated binding protein 1) and phosphoinositol kinase, activates the MAP Kinase pathway through RAS (rat sarcoma virus oncogene 1) and ERK (extra-cellular signal related kinase). The antagonistic effect on stem cell renewal can be partly explained by the phosphokinase activity of SHP-2. However blockade of ERK activating enzyme MEK-1 (mitogen activated protein kinase-1) with the specific blocking agent PD05809 reduces ES differentiation both in monolayer culture and in embryoid bodies.

Reciprocal patterns of LIF-R expression in the ICM and LIF in the trophectoderm add weight to the theory that Gp130 signalling is important for the expansion of the pluripotent compartment of the peri-implantation embryo (Nichols *et al.*, 1996). Paradoxically, knockout of the *Lif* gene has no effect on embryonic development (Stewart *et al.*, 1992). Knockout of the *Lif* receptor or of Gp130, the common component of the IL-6 cytokine family of receptors, results in death at mid-gestation (Ware *et al.*, 1995; Yoshida *et al.*, 1996). Disruption of the *Stat-3* gene results in early embryonic lethality but not until onset of gastrulation (Takeda *et al.*, 1997)

Experiments with *gp130*^{-/-} mice (see Sect 1.1.3) may throw some light on why LIF has an effect on ES cells while being redundant over the relevant period of development. Nichols *et al* (2001) have shown that gp130 signalling, mediated by LIF is essential to the survival of the ICM during diapause.

It has been shown that BMP2/4 (bone morphogenetic protein) , a member of the TGF- β (transforming growth factor- β) superfamily, can in concert with LIF support undifferentiated growth of mouse ES cells in the absence of other extrinsic factors, supplied by feeders or serum, (Ying *et al.*, 2003b) and appears to work by suppressing both the differentiation of neuronal lineages through SMAD (small mothers against decapentaplegic) and Id (inhibitor of differentiation) and by blocking the MAPK (mitogen activated phosphokinase) pathway which promotes differentiation under the influence of LIF (Qi *et al.*, 2004). The role of different signalling cascades has been hampered by the complexity of the media used to maintain the cells in their undifferentiated state. The ability to grow ES cells in the absence of complex factor mixtures such as serum should help in the elucidation of the interacting pathways.

Array experiments with ES cells have also highlighted the canonical WNT (named for the drosophila wingless gene and its mammalian homologue *Int-1*) and TGF β pathways as having putative roles in the maintenance of pluripotency (Sato *et al.*, 2003). The TGF β pathway again appears to operate on the MAPK pathway.

WNTs are a family of secreted protein which binds to the cell surface receptor Frizzled resulting in inhibition of the kinase Gsk-3 β (glycogen synthase kinase-3 β), a component of the Axin/APC complex which acts to mark β -catenin for degradation. Axin is a protein with several protein interaction domains and acts as a scaffold for the complex and APC (adenomatous polyposis coli) is a tumour suppressor protein commonly deleted in familial and sporadic colorectal cancer. Inhibition of the phosphorylation of β -catenin results in translocation to the nucleus where it interacts with the factors Tcf (T-cell factor) and Lef (lymphocyte enhancer factor) to regulate target genes. Mouse embryo fibroblasts (MEFs) which are often used as feeders for ES cells produce the WNT ligands. WNTs have been implicated in maintenance of the multipotent cells of the haemopoetic system. Moreover, inhibition of WNT by SFRP2 (secreted frizzled-related protein-2), an extracellular antagonist of WNT which competes for the frizzled receptor, appears to direct cells down a neuronal route (Aubert *et al.*, 2002)

It has been shown that both human and mouse ES cells can be maintained short term in the pluripotent state by WNT-3a provided by either media conditioned by cells which stably produce it or by the inhibitor of GSK-3 β (Sato *et al.*, 2004) derived from Tyrian purple known as BIO (Meijer *et al.*, 2003). Also both WNT3a and WNT4 are expressed by the blastocyst (Lloyd *et al.*, 2003). It has been suggested that the effect of WNT on neuronal differentiation is due to downstream regulation of BMPs

(Haegele *et al.*, 2003). However, *in vivo* analysis suggest that, like the JAK/STAT pathway, their ability to maintain ES cells in an undifferentiated state may be an artefact of culture rather than a reflection of this very important developmental stage characterised by the expansion of the epiblast. Disruption of β -catenin by disruption of the endogenous gene does not affect formation of the embryonic egg cylinder but appears to be important to axis formation (Huelsenken *et al.*, 2000).

It has been recently shown that WNT-3a in the absence of LIF and other serum factors can produce a transient expression of *Cdx2* which is sufficient, on the subsequent addition of the permissive factors, for the production of trophectoderm stem cells (TS) (He *et al.*, 2008). This is the first time the ability of mouse ES to differentiate to TE simply by manipulating the *in vitro* growth conditions has been demonstrated. Moreover, it shows how transient signals might be able to initiate differentiation cascades *in vivo*.

It should not be forgotten that ES cells grow on a substrate which even when non-cellular still provides signals which are potentially important in maintenance of the pluripotent state. *In vivo*, the epiblast is induced to form a cavity and an epithelium by a combination of signals from the primitive endoderm and the extra-cellular matrix it produces (Coucouvanis and Martin, 1995).

The signalling networks involved in the control of pluripotency are becoming ever more complex. The different pathways initiated on the cell surface each seem to have both pro- and anti-differentiative effects and each pathway may interact to modulate the others. The WNT pathway appears to acts in concert with the BMP pathway to repress neural differentiation (Haegele *et al.*, 2003). It also appears to interact with the JAK-STAT pathway by up-regulation of *Stat3* (Hao *et al.*, 2006). In contrast LIF and

the JAK-STAT pathway in concert with BMP are believed to interfere with *Lef1* expression which is required to activate *Cdx2* which induces TE (He *et al.*, 2008; Strumpf *et al.*, 2005). It is important to remember that the context of any signalling is context dependant. Although WNT induces CDX2, the continued presence of WNT acts as a negative feedback loop, and the terminal differentiation of the stimulated cells can only be achieved when the signalling molecules are changed to a permissive combination.

Although the mechanism is unclear, the differentiation of the endoderm *in vivo* seems to be controlled by GRB2 and the MAPK pathway as *Grb2* *-/-* embryos lose *Gata-6* expression and produce no primitive endoderm (Chazaud *et al.*, 2006). The BMP and STAT3 pathways have to be balanced in their effects to maintain the undifferentiated state. Expression of a constitutive BMP receptor or over-expression of SMAD1/4 (Ying *et al.*, 2003b) will cancel out the STAT3 pathway leading to non neuronal differentiation. ERK and p38 MAP kinase which are activated by LIF or FGF signalling act to promote differentiation to neurons (Ying *et al.*, 2003a; Ying and Smith, 2003). The PI3/AKT (phosphatidylinositol-3/ a serine threonine kinase) also inhibits the ERK MAPK pathway (Paling *et al.*, 2004). Disruption of any one of these pathways *in vivo* does not lead to peri-implantation loss.

Interestingly, *c-myc* is a direct target of STAT-3 (Cartwright *et al.*, 2005), and of the TGF β and WNT pathways (Kanehisa and Goto, 2000) and is one of the four genes which on co-introduction to somatic cells results in transformation to ES like cells which can contribute to chimaeras (Takahashi and Yamanaka, 2006).

1.1.5 Intrinsic Factors and the Maintenance of Pluripotency.

In contrast with ES cells, the pluripotent compartment of the embryo is neither static nor permanent. The pluripotent cells of the epiblast are transcriptionally distinct from those of the embryonic ectoderm.

The mouse embryo can develop from the one cell stage to the blastocyst stage in the absence of extrinsic factors and embryogenesis appears to be driven by asymmetrical gene expression inherent in the embryo although the mechanism of establishment of the asymmetry is unclear. On differentiation of the trophectoderm, signalling between the two tissues can guide further development. After implantation, the embryo is provided with nutrients and signalling molecules from the mother. However, many of the important developmental cues appear to still be provided by signalling centres established within the embryo.

The provision of extrinsic signals in the growth medium of ES cells leads to the regulation of nuclear transcription factors which regulate downstream targets which define cell type. Many of these factors associated with pluripotent cell lines and tissues have been discovered. Of these, OCT4, a POU domain transcription factor and the product of the *Pou5f1* gene, has been widely studied. *In vitro* it is expressed in ES, EC and EG cells (Okamoto *et al.*, 1990; Rosner *et al.*, 1990; Yeom *et al.*, 1996) in all pluripotent tissues of the embryo: oocytes, all cells of preimplantation embryos until the blastocyst stage when it is restricted to the ICM, the epiblast and embryonic ectoderm (Palmieri *et al.*, 1994; Scholer *et al.*, 1989) when it becomes again restricted to the primordial germ cells (Rosner *et al.*, 1990). Deletion of the gene in the mouse results in an early embryonic lethal. Although cells are allocated to the ICM resulting in a blastocyst with normal appearance, the embryo is only composed of

trophectoderm cells and develops no further (Nichols *et al.*, 1998). Regulation of this gene is critical to the differentiation of ES cells. Repression of *Pou5f1* leads to the cells adopting trophectodermal morphology while up-regulation leads to formation of primitive endoderm (Niwa *et al.*, 2000). However, OCT4 functions within a complex system of other factors which in the embryo lead to the dynamic process which is embryogenesis. One of the best studied of the genes which are involved in the co-regulation of other genes is *Sox2*.

SOX2 (an SRY related HMG box transcription factor *Sox2*) is a member of a group of proteins which act as partners for transcription factors which then regulate gene expression in the context of the factors present. In the chick lens SOX2 partners with the δ EF3 protein to direct δ -crystallin gene expression (Kamachi *et al.*, 1995). In F9 EC cells it partners with OCT4 in the activation of the *Fgf4* gene (Ambrosetti *et al.*, 2000) and *Hand1* which are involved in the proliferation or differentiation of trophectoderm in the blastocyst (Niwa, 2001; Rossant and Cross, 2001). It also partners OCT4 in its own regulation and the regulation of OCT4 (Okumura-Nakanishi *et al.*, 2005; Tomioka *et al.*, 2002).

Sox2 RNA is first expressed in some cells of the morula. The protein SOX2 is present in all blastomeres of the cleavage stage embryo and also becomes restricted to the ICM and epiblast at the blastocyst stage (Avilion *et al.*, 2003). Expression persists in the epiblast stage through to gastrulation when mid streak it becomes restricted to the anterior embryonic ectoderm which is destined to form neurectoderm. However, a new site of expression of *Sox2* is initiated in the ectoplacental cone where there is no *Pou5f1* expression. Disruption of the *Sox2* gene leads to peri-implantation death with loss of the epiblast at a stage slightly later than for *Pou5f1* knockout embryos.

Partnership with both OCT4 and KLF4 is not only important in the specification of ES cells being one of the three proteins which can reprogram mouse fibroblasts (Nakagawa *et al.*, 2008) but also for expression of *Lefty1* (Nakatake *et al.*, 2006). Deletion of *Lefty1* leads to loss of the left right axis during differentiation of the embryo (Meno *et al.*, 1998) and may contribute to axis specification in the early postimplantation epiblast (Takaoka *et al.*, 2006)

A gene for a homeodomain containing transcription factor was first identified as *Enk* (Wang *et al.*, 2003) and subsequently renamed as *Nanog* (Chambers *et al.*, 2003; Mitsui *et al.*, 2003) has been proposed as a candidate gene for self-renewal. *In vitro* it is expressed in ES cells, embryonic germ cells (EG cells) and both LIF dependant and independent EC cells (Chambers *et al.*, 2003). *In vivo*, *Nanog* is first detected in the inside cells of the compacted morula, the cells destined to form the ICM. Expression in the blastocyst is restricted to the ectoderm and is down regulated in the late blastocyst before implantation. *Nanog* is also expressed in the germ cells of the genital ridge on day 11.5 of gestation (Chambers *et al.*, 2003) a tissue from which EG cells can be derived (Matsui *et al.*, 1992; Resnick *et al.*, 1992) DPPA3, also known as Stella and considered to be a marker of pluripotency (Bowles *et al.*, 2003), is present in pre-migratory primordial germ cells on day 7.5 of gestation. However, *nanog* was absent from these cells (Hatano *et al.*, 2005).

Over expression of *Nanog* maintains ES cells in an undifferentiated state in the absence of LIF but still requires the expression of *Pou5f1*. ES cells deficient in *Nanog* differentiate to form primitive endoderm but not mesoderm. In line with the belief that the EC line P19 represents a later developmental stage more akin to embryonic

ectoderm, *Nanog* is expressed at a much lower level. Also expression in EPL (early primitive ectoderm like) cells (Section 1.1.5) is reduced (Wang *et al.*, 2003)

Nanog deficient embryos at day 5.5 appear to be only composed of disorganised extra-embryonic endoderm with no epiblast or extra-embryonic ectoderm. *Nanog* deficient blastocysts at day 3.5 appear to be normal but cells in ICMs isolated from these embryos fail to proliferate and differentiate to parietal endoderm. Although *Nanog*^{-/-} ES cells can be derived from mutant *Nanog*^{-/-} blastocysts they differentiate slowly into parietal endoderm (Mitsui *et al.*, 2003).

The anti-differentiative effects of *Nanog* apply to all three lineages depending on the context. NANOG can block the formation of primitive endoderm in response to the removal of LIF in serum containing medium (Chambers *et al.*, 2003) or on the formation of embryoid bodies (Hamazaki *et al.*, 2004) in the presence of LIF and serum. Neuronal differentiation is blocked in the absence of LIF and BMP in serum free culture (Ying *et al.*, 2003b). Mesodermal differentiation has been shown to be inhibited by up-regulation of *Nanog* brought about by the binding of the transcription factor T (the product of the *Brachyury* gene) and STAT3 (regulated by LIF) to an enhancer element of the *Nanog* gene and NANOG binds to SMAD1 (Suzuki *et al.*, 2006b). Cells which have been induced to form mesodermal progenitors by stimulation with BMP regenerate ES cells in response to *Brachyury*, in combination with LIF (Suzuki *et al.*, 2006a).

Although neither *Oct4* nor *Nanog* can be regarded as a master gene of pluripotency, as neither gene can prevent differentiation on its own, NANOG, OCT4 and SOX2 have been shown to act in concert both to self regulate and to regulate other genes. Although the STAT3 pathway does not appear to interact with Oct4; differentiation

cannot be prevented by *Oct4* expression if LIF has been withdrawn or on blockade of the Stat3 pathway. It has been shown that co-transfection of *Oct4*, *Sox2*, *Klf4* and *c-myc* into fibroblasts is able to reprogram fibroblasts to *Nanog* expressing cells which resemble ES cells including the ability to contribute to chimaeras (Takahashi and Yamanaka, 2006; Wernig *et al.*, 2007) and to the germ line (Okita *et al.*, 2007) although the mice generated frequently develop tumours due to reactivation of the transfected *c-myc* gene. The re-specification of fibroblasts to ES cells occurs at low frequency which may be in part due to the requirement of the genes to be expressed at the correct level. Subsequently it has been shown that the transfection of *c-myc* is dispensable (Nakagawa *et al.*, 2008). The oncogene *c-myc* enhances DNA replication leading to a more relaxed chromatin structure. The authors noted that the induced ES cells took longer to emerge without *c-myc* and at a lower frequency. I suggest that the ability of *Sox2* to open the DNA in a local fashion (Pevny and Lovell-Badge, 1997) may play a role in initiating the epigenetic state required to establish the transcriptional network.

The ease with which ES cells can be genetically manipulated has led to the belief that the DNA is open and that transcriptional control is leaky (Evans *et al.*, 1997). Furthermore, when ES cells differentiate the nuclei shrink and the distribution of heterochromatin changes (Niwa, 2007). In addition, the epigenetic status of ES cells is unique in that the DNA bears co-localised histone marks characteristic of both inactive and active genes (Bernstein *et al.*, 2006) that are associated with genes expressed at low levels. In general, chromatin marks associated with active genes are abundant in ES cells (Azuara *et al.*, 2006; Jeong-Heon Lee, 2004) and rapid changes occur in the histones during self renewal (Meshorer *et al.*, 2006). Since pluripotency can be restored to ES cells which are deficient for certain of the genes which are

involved in this open chromatin state Niwa (2007) has proposed that the epigenetic processes facilitate the transcription factor network rather than maintain pluripotency.

1.1.6 EG Cells

Primordial germ cells can first be detected by alkaline phosphatase staining in the extra-embryonic mesoderm at the base of the allantoic bud on day 7.5 from where they migrate to the hindgut area then to the genital ridge where they stay in the developed gonad. As discussed in above (Section 1.1.1), teratocarcinomas can be established from post- migratory germ cells.

Although pre-migratory germ cells have not so far been shown capable of forming teratocarcinomas, it has been shown that limited culture of germ cells could be achieved by culture in the presence of LIF and the Steel growth factor (Matsui *et al.*, 1991). Further, addition of bFGF to cultures of primordial germ cells resulted in their conversion to EG cells, so named to denote their origin, which resembled ES cells and which could also be maintained indefinitely (Labosky *et al.*, 1994; Matsui *et al.*, 1992; Stewart *et al.*, 1994). However, lines derived from later stage embryos progressively lose their ability to contribute to normal embryogenesis as they lose their imprinted patterns (Tada *et al.*, 1998). EG cells, like their germ cells founders, have the ability to erase imprints (Tada *et al.*, 1997). ES cells do not appear to have this capacity.

1.1.7 EPL Cells

Progression from ICM to primitive ectoderm involves differentiative stages characterised by differences in the expression of the genes *Rex1*, *Fgf5*, *Gbx2*, *Psc1*,

CRTR-1 and *PRCE* resulting in distinguishable pluripotent cell populations (Pelton *et al.*, 1998).

	Day3.5	Day4.5	Day4.75	Day5	Day5.25	Day5.5
<i>Rex1</i>	Red	Red	Pink	Blank	Blank	Blank
<i>CRTR1</i>	Red	Red	Pink	Blank	Blank	Blank
<i>Psc1</i>	Red	Red	Red	Pink	Blank	Blank
<i>Fgf5</i>	Blank	Blank	Blank	?	Red	Red
<i>PRCE1</i>	Blank	Pink	Red	Pink	Blank	Blank
<i>Oct3/4</i>	Red	Red	Red	Red	Red	Red

	ES	EPL
<i>Rex1</i>	Red	Blank
<i>CRTR1</i>	Pink	Blank
<i>Psc1</i>	Red	Blank
<i>Fgf5</i>	Blank	Red
<i>PRCE1</i>	Pink	Red
<i>Oct3/4</i>	Red	Red

Figure 1.1: Diagram of expression of genes in the pluripotent compartment of embryos associated with differentiation over the period from the early blastocyst to formation of the primitive ectoderm. Below expression of the same genes in EPL and ES cells. Red indicates high expression, pink low expression and blanks no expression. Data extracted from Pelton *et al.*, (1998).

Early primitive ectoderm cells (EPL cells) have been isolated as a distinct cell type by the culture of ES cells in conditioned media from a human hepatocarcinoma cell line HepG2 (Lake *et al.*, 2000; Rathjen *et al.*, 1999). Analysis of expressed markers, morphology, cytokine responsiveness and differentiation potential both *in vivo* and *in vitro*, have led the investigators to suggest that EPL cells are more related to the embryonic primitive ectoderm and are distinct from ES cells. These cells do not contribute to chimaeras but can be reverted to the ES state by removal of the conditioned media and replacing LIF whereupon they regain their ability to contribute to chimaeras.

1.1.8 Pluripotent Cells from Human.

It has proved remarkably difficult to produce ES cells from other species. Where cells which share properties with mouse ES cells have been isolated it has proved either impossible to confirm by the making of chimaeras for technical reasons, as in the hamster or for ethical reasons as in the human.

ES like cells have been isolated from primate blastocysts and this includes human. These ES cells hold the promise of therapy for people with degenerative illnesses or accidents resulting in the loss of organs. Before this dream can be realized, we must understand not only how they are similar to the mouse but also how they differ. Unlike the mouse ES cells, human ES cells (hES) can differentiate to form trophectoderm. This suggests that hES cells may represent an earlier stage of development before the delamination of the primitive endoderm. Also, only the pro-differentiative ERK/MAPK pathway and not the Jak/STAT pathway appear to be responsive to LIF (Sato *et al.*, 2004) BMP promotes differentiation to trophectoderm rather inhibiting it. Blocking of BMP with the antagonist Noggin leads to differentiation into neurons (Pera *et al.*, 2004) which appears to be a common effect in mouse and in *Xenopus* (Munoz-Sanjuan and Brivanlou, 2002) suggesting a common function of BMP in blocking neuronal differentiation. The other arm of the TGF β family appears to be more important in controlling pluripotency in hES cells. The factors TGF β , activin and nodal activating SMAD2/3 are able to inhibit differentiation (James *et al.*, 2005; Vallier *et al.*, 2004). FGF2 acts as a competence factor, by a poorly understood mechanism, with nodal and activin to increase this inhibitory effect (Vallier *et al.*, 2005).

1. 2 Microarrays

Microarrays are ordered arrays of probes (also known as elements) printed in columns and rows on solid, planar substrates thus giving a unique address to each probe. This allows the best use of current technologies such as robots, scanners and software. The probes can be nucleic acid, protein or even tissue. Probes can be almost anything which can be deposited on the substrate e.g. protein, nucleotides or even tissue sections.

A major application of microarray technology is transcriptome analysis. Other methods for analysis of gene expression include Northern blotting (Alwine *et al.*, 1977), S1 nuclease protection (Berk and Sharp, 1977), differential display (Liang and Pardee, 1992), sequencing of cDNA libraries (Adams *et al.*, 1991; Okubo *et al.*, 1992) and serial analysis of gene expression (SAGE) (Velculescu *et al.*, 1995). Microarrays have the advantage of allowing this analysis on a global scale

For transcriptional analysis the probes, which are tethered to the matrix, are either oligonucleotides or cDNA. The matrix is usually a standard glass microscope slide with a chemically modified surface.

The best known commercially produced oligonucleotide array platform is the GeneChip® produced by Affymetrix Inc, CA, U.S.A. which consists of a quartz chip on which up to 500,000 probe locations can exist on an area of 1.28cm². The probes are synthesised in situ by light-directed chemistry (Affymetrix). The GeneChip® arrays are termed one colour platforms because only one sample is applied to each chip and absolute expression levels are measured for each gene being queried, several

positive and negative oligonucleotides are present. Accuracy of measurement relies on the small array-array differences achieved during production of the GeneChip®.

The other commonly used technology is the two colour array where cDNA PCR products are tethered to the slides. Two samples are independently fluorescently labelled and then co-hybridised to the slide effectively eliminating the array-array differences. Relative gene expression levels are determined by measuring the ratio of the intensities of fluorescence for both Cy3™ and Cy5™ dye at each probe/spot.

Although the one colour platforms have the advantage of producing absolute levels for each gene, the selection of positive and negative oligonucleotide for each gene is critical. A particular issue with amplified material is the 3' bias introduced in the reverse transcription of the mRNA. Affymetrix, in recognition of this problem, have started to produce chips designed for this 3' biased material.

The NIA 15K gene set was compiled to address the problem of the lack of coverage of genes important during early mammalian development. A PCR based cDNA library construction method was used to compile a total of 52,374 ESTs from pre- and peri-implantation embryos (Ko *et al.*, 1998) embryonic and extra embryonic tissues on 7.5-d.p.c., female gonad and mesonephros on 12.5-d.p.c. and newborn ovary (Tanaka *et al.*, 2000). From the 52,374 ESTs, 15,264 putative unique genes were extracted.

The majority of these probes were derived from the 3' end of the cDNAs, an important consideration as linear amplified RNA is used as the target and have an average length of 1.5kb. Only 30% of the sequences represent known genes, the rest are either novel (~50%) or are similar but not identical to known genes (~20%).

1.3 Aims of the Project

In this thesis I have used microarrays to phenotype ES cells and pluripotent tissues of the embryo to address the question of homology of ES cells to the pluripotent cellular compartment of early embryos. To this end I have assessed techniques for extracting and handling RNA from small samples, amplification of the transcriptome and labelling of the samples. A number of software packages for the image analysis of the microarray slides and data mining were also scrutinised.

Chapter 2

Materials and Methods

2. Materials and Methods

2.1 Cell Culture

2.1.1 Media.

The following media and reagents were purchased:

Distilled water, tissue culture tested	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK
DPBS-A, Dulbecco's Phosphate Buffered Saline (1x) liquid, without Ca ²⁺ and Mg ²⁺ ,	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK
Foetal Bovine Serum (FBS), selected batches	PAA Laboratories GmbH, Linz, A-4020 Austria
Gelatin, 0.1% in water	StemCell Technologies, Vancouver, BC, Canada
Glutamine 200mM (100x)	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK
Murine LIF (ESGRO™)	Chemicon International, Temecula, California 92590, USA
MEM Alpha medium (1x), liquid with glutamine, without ribonucleosides and deoxyribonucleosides	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK
Newborn Bovine Serum (NBS), selected batches	PAA Laboratories GmbH, Linz, A-4020 Austria
Penicillin/Streptomycin Solution, liquid 5,000units/ml Penicillin and 5,000mg/ml Streptomycin in normal saline	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK
Trypsin-EDTA (1x) in HBSS, prepared as 0.5g/l Trypsin (1:250), 0.2g EDTA in Hanks' Balanced Salt Solution without Ca ²⁺ and Mg ²⁺	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK

Table 2.1: Cell Culture Media

Other Reagents

A 10⁻¹M (100x) stock solution of β-mercaptoethanol was prepared by adding 72μl of β-mercaptoethanol (Merck KGaA, 64293 Darmstadt, Germany) to 100ml DPBS-A.

Selected batches of Foetal Bovine Serum (FBS) and Newborn Bovine Serum (NBS) were heat inactivated before use by incubating at 56°C for 20 minutes. All reagents were made up and stored in sterile disposable plastic vessels (NUNC™, Falcon™).

2.1.2 Culture of Embryonic Stem Cells

Embryonic stem cells (ES cells) of the line IMT-11 (Bierbaum *et al.*, 1994) were used in all experiments. The precise genetic constitution of these cells is unclear; the original stock of mice was acquired from Dr E. Wagner who acquired his stock from Prof R. Kemmler. These mice were described as 129SvSl cp, therefore agouti, white bellied and dark eyed. The breeding program did not select for the Steel locus so this was most likely lost. Thus these ES cells are most similar to the D3 cell line (Doetschman *et al.*, 1985) although further genetic contamination of the original mouse stocks can not be ruled out.

These ES lines were maintained in their undifferentiated state in the medium MEM-Alpha supplemented with 10⁻⁴M β-Mercaptoethanol, 10⁻³ U/ml murine LIF (ESGRO™), 2mM glutamine, 10% FBS and 10% NBS (MEMα-ES). The cells were maintained at 37°C in a humidified atmosphere with 5% CO₂ on gelatin coated tissue culture grade plastic ware (NUNC™). The gelatinization was achieved by flooding the plates with 0.1% gelatin (Stem Cell Technologies) then aspirating the solution from the plates before use.

Subculture was carried out by washing the cells twice with Dulbecco's Phosphate Buffered Saline without Ca²⁺ and Mg²⁺ (DPBS-A, GIBCO™) at room temperature then rinsing the cells with Trypsin-EDTA at 4°C. When the cells had rounded up and detached from the plastic substrate, MEM-α was added and the cell suspension

triturerated to give a single cell suspension. Cells were counted and plated at 3×10^4 cells/cm², i.e. 2×10^6 cells were plated on a 100mm dish. After 48 hours of culture the media was changed and after further 24h cultivation a 100mm dish would yield approximately 1.5×10^7 cells.

2.2. Animal Husbandry, Embryo and Tissue Isolation

The following reagents were purchased.

Albumin, Bovine Fraction V Powder, 96-99% albumin, Embryo tested, BSA	Sigma-Aldrich Company Ltd, Poole, Dorset, UK
Depo-Provera™	Upjohn Ltd, Crawley, UK
Opti-MEM™-1 Reduced Serum Medium (1x), liquid	GIBCO™ Invitrogen Ltd, Paisley, Renfrewshire, UK
Tamoxifen	Sigma-Aldrich Company Ltd, Poole, Dorset, UK
Tyrode's Solution, Acidic	Sigma-Aldrich Company Ltd, Poole, Dorset, UK

Table 2.2: Reagents for collection of embryos.

2.2.1. Timed Matings.

Mice of the strain 129SvEv were maintained on a 12h light/12h dark light cycle. Food and water were available *ad libitum*. Females were caged with males and the females checked each morning for the presence of a vaginal plug. As oestrus females usually ovulate at the mid-point of the dark period, embryo stages are represented as days post coitum (d.p.c.) or as hours post coitum (h.p.c.) assuming that fertilization occurred at midnight preceding the discovery of the vaginal plug. For the timings of the different developmental stages see Appendix 1. Embryo isolation was carried out in sterile disposable dishes (NUNC™, Sterlin™). Embryos were manipulated by means of a

finely drawn out glass Pasteur pipette attached to a mouthpiece. For full details of all embryo manipulation techniques see (Nagy, 2003).

2.2.2 Embryo Isolation.

All embryos were collected and handled in OptiMEM-1™ (Gibco) supplemented with 4mg/ml BSA (OptiMEM+BSA). OptiMEM-1™ is a convenient ready made medium which is HEPES buffered and suitable for embryo handling outside of the CO₂ controlled atmosphere of the incubator (Ouhibi *et al.*, 1995).

2.2.2a Oocytes.

Oocytes were isolated between six and eight h.p.c from females which had been mated with vasectomised males. The oviducts were dissected from the mice into a dish of OptiMEM+BSA and the ampulla was opened with No 5 watchmaker's forceps allowing the escape of the oocytes with their associated cumulus cells into the medium. The cumulus cells were separated from the oocytes by incubating in hyaluronidase (Sigma), 0.3mg/ml in OptiMEM at room temperature until the cumulus cells drop off. The oocytes were then rinsed through several drops of OptiMEM+BSA to free them from cumulus cells, hyaluronidase and other debris. The oocytes were then transferred to an RNase free microcentrifuge tube in the smallest volume of medium before lysis in the buffer supplied in the Zymo Mini RNA Isolation kit™(Zymo).

2.2.2b Fertilised Eggs.

Fertilised eggs were collected from timed matings (8-10 h.p.c.) with fertile males in exactly the same manner as for oocytes (2.2.2b)

2.2.2c Two-cell to Compacted Morulae.

The oviducts from timed matings with fertile males were dissected into dishes. Embryos were isolated from the oviducts by inserting a 33-gauge needle into the infundibulum of the oviduct and flushing with OptiMEM+BSA by means of a syringe.

2.2.2d Blastocysts

Pregnant females were sacrificed on days 3.5- 5 post coitum. Blastocysts were flushed from the uteri in OptiMEM™-1 supplemented with 4mg/ml BSA (OptiMEM+ BSA).

2.2.2e Post implantation Embryos.

Pregnant females were sacrificed on days 5.5-6.5 post-coitum and the decidua were dissected free of the uteri. Embryos were carefully removed from the decidua with tungsten needles.

2.2.2d Delayed Implantation.

Embryonic diapause was induced by intraperitoneal injection of 10µg of tamoxifen (Sigma) and subcutaneous injection of 1mg of DepoProvera™ (Hunter and Evans, 1999) on day 2.5-3 of gestation.

2.2.3 Tissue Isolation.

2.2.3a Production of Rabbit anti Mouse Spleen Cell Antibodies.

Mouse spleens were teased in DPBS-A then spun down at 200g for 5min. The red blood cells were lysed by washing in isotonic TRIS Ammonium Chloride, pH 7.2 (17mM TRIS base, 139mM Ammonium Chloride, pH7.2). At weekly intervals 2×10^8 cells were injected into a rabbit. After 4 weeks a test bleed was taken and the efficacy of the

antibody was tested. The anti-serum was heat inactivated at 56°C for 20min to destroy endogenous complement activity.

This antibody was kindly produced by Dr Eryl Liddell, School of Biosciences, Cardiff University.

2.2.3b Isolation of the Inner Cell Mass by Immunosurgery.

The inner cell masses of blastocysts were dissected by immunosurgery (Solter and Knowles, 1975). Briefly, where necessary, the zona pellucida was removed from the embryos with acidic Tyrode's solution (Nicolson *et al.*, 1975). The embryos were then incubated in rabbit anti mouse spleen antibody (see sect 2.2.3a.) diluted 1:10 in OptiMEM+BSA at room temperature for 20 minutes. The embryos were then washed through six drops of OptiMEM+BSA to remove all unbound antibody. The blastocysts were then incubated in Guinea pig complement diluted 1:5 in OptiMEM+BSA at 37°C for 20minutes. After washing through 3 drops of OptiMEM+BSA, the lysed trophectoderm was removed by pipetting through a finely drawn glass capillary

2.2.3c. Dissection of the Dy6 Epiblast.

Embryos were dissected from the uterus and decidua as described in Sect 2.2.2b. The Reichert's Membrane and ectoplacental cone were removed with tungsten needles and the embryonic portion was separated from the rest of the embryo. The visceral endoderm layer was removed by incubating the embryonic tissue in 0.25% pancreatin, 0.5% trypsin in DPBS-A at 4°C for 10min. The endoderm layer was then peeled off by pipetting through a glass capillary drawn to a diameter slightly smaller than the embryonic cylinder

2.3 Molecular Methods

When handling RNA, care was taken to avoid contamination with RNase. Gloves were worn at all times and changed frequently. Sterile disposable plastics were used wherever possible. Pipette tips with filters (Maximum Recovery™, Axygen Scientific Inc.) and microcentrifuge tubes (Sigma-Aldrich) were purchased nuclease free. All glassware was dry autoclaved in an oven at 240°C for a minimum of four hours. Solutions and water were treated with DEPC (0.1% diethylpyrocarbonate) then autoclaved at 15lbs/sq.in. for 15mins. Solutions which could not be treated this way such as TRIS, were either purchased as nuclease free concentrated stocks or made up from dedicated chemicals in RNase free glassware with DEPC treated water. Equipment such as electrophoresis tanks were sprayed with RnaseZap™ (Ambion) then rinsed with DEPC treated water.

Traces of DEPC can be detrimental to many enzyme reactions. In this case nuclease-free water (Ambion) was used.

All chemicals were purchased as Molecular Biology Grade.

2.3.1 Purification of RNA.

2.3.1a Isolation of total and cytoplasmic RNA.

The following kits for RNA isolation were purchased (Table2.2)

QIAGEN RNeasy™ Midi Kit	Qiagen Ltd., Crawley, Sussex, UK.
QIAGEN RNeasy™ Mini Kit	Qiagen Ltd., Crawley, Sussex, UK.
Zymo Mini RNA Isolation Kit™	Zymo Research, Orange, California, USA

Table 2.3: Kits purchased for isolation of RNA.

All three of these kits work on the same principle. The cells or the cytoplasm was disrupted in a high salt guanidinium buffer which provides optimal binding conditions for the RNA to the silica gel membrane. Contaminants were washed away with wash buffer and the RNA eluted with water.

Embryonic stem cells were washed twice with DPBS-A then were treated with trypsin-EDTA as for subculture. The cells were counted in a haemocytometer and collected as a pellet by centrifugation at 200g at room temperature. As ES cells are believed to be rich in RNA, to avoid overloading the columns, the midi column was used for all samples with more than 5×10^6 cells and the mini column was used for samples with less than 5×10^6 cells. The cells were then processed using the protocol provided with the kit for the isolation of total RNA from the cytoplasm of the cells to minimise DNA contamination of the RNA and to enrich for mature, fully spliced mRNA.

The Zymo Mini RNA Isolation kit™ works exactly as above but is designed for the extraction of RNA from very small samples down to 10^1 cells and elutes in 8µl of water. In this case to avoid losing precious material, we isolated total RNA rather than cytoplasmic. Isolated tissues from embryos or colonies of ES cells picked by means of a glass capillary and were collected in a minimal volume of DPBS-A in sterile microcentrifuge tubes. The cellular material was lysed in 200µl of Zymo RNA Extraction Buffer. The manufacturers' protocols were followed in all cases.

2.3.1b Purification of mRNA using Dynabeads™

Messenger RNA (mRNA) was extracted by means of a magnetic poly-T bead. These magnetic beads have oligo(dT)₂₅ covalently bound to their surface. The polyA tail

present on the mRNA hybridises to the oligo(dT) on the bead. By placing the tube in a magnetic stand, the supernatant with the impurities can be removed.

Resuspended beads (200µl) (Dynal, Oslo, Norway) were placed in a sterile microcentrifuge tube which was then placed in a magnetic stand (Promega Ltd., Southampton, UK). The beads were left for approximately 30 seconds to allow the suspension to clear, and then the supernatant was removed. The beads were pre-washed in 200µl 2x binding buffer (20mM Tris-HCl pH 7.5, 1M LiCl, 2mM EDTA), and again placed in a magnetic stand. Meanwhile, 150µg of total RNA (200µl) was incubated at 65°C for 2 minutes. The 2x binding buffer was removed from the beads and these were resuspended in a further 200µl of 2x binding buffer. The total RNA was added to the bead suspension, and mixed thoroughly for 5 minutes at room temperature. The vial was placed in the magnetic stand for 2 minutes and the supernatant removed. The bound mRNA was washed twice by the addition of 400µl of washing solution (10mM Tris-HCl pH 8.0, 0.15M LiCl, 1mM EDTA). After thorough mixing, the tube was placed back in the magnetic stand, and the supernatant removed. Care was taken to remove as much of the wash solution as possible. Elution solution (10mM Tris-HCl pH 8.0) was added to the tube (20µl) and was incubated at 80°C for 2 minutes. The tube was placed in the magnetic stand, and the supernatant containing the mRNA was removed to a sterile microcentrifuge tube. The mRNA was stored at -70°C until use.

2.3.2. Ethanol Precipitation of Nucleic Acids.

DNA was precipitated by the addition of 0.1 volume of 3M Sodium Acetate (Sigma) and 2.0 volumes of absolute ethanol and incubation at -20°C for 1-2 hour or overnight. The DNA was then pelleted by centrifugation for 20minutes at 14,000g at 4°C. The pellet was washed once in 70% ethanol then air dried.

RNA and cDNA was treated in the same manner except that 2.5 volumes of absolute ethanol were used.

2.3.3 Analysis of Nucleic acids By Optical Density

Nucleic acids absorb at a wavelength of 260nm. Given that an absorbance of 1_{OD} is equivalent to 50 $\mu\text{g ml}^{-1}$ of double-stranded DNA or 37 $\mu\text{g ml}^{-1}$ single-stranded DNA and 40 $\mu\text{g ml}^{-1}$ RNA, it was possible to quantify DNA and RNA concentrations. Furthermore, determining absorbencies at 230nm and 280nm provides a means of examining sample purity. Pure preparations of DNA and RNA will have an A_{260}/A_{280} ratio of approximately 1.8 and 2.0 respectively. Any protein or phenol contaminants will lower these values. Likewise an A_{260}/A_{280} ratio of less than 2 is considered to be indicative of ethanol or salt contamination (Sambrook *et al.*, 1998).

2.3.4 Analysis of Labelled Target by Optical Density.

For each labelling reaction the A_{260} and A_{550} or A_{650} for Cy3TM and Cy5TM respectively were measured. These measurements were used to calculate the amount of cDNA synthesised and the amount of dye incorporated into the cDNA. See Appendix 2 for calculations.

2.3.5. Slide Gels.

A slide gel was produced by pouring approximately 2.5ml of 1.5% agarose dissolved in TAE buffer (40mM TRIS, 1mM EDTA and 20mM acetic acid, pH8.5) into a slide gel mould (kindly fabricated by Mr John Harris, School of Biosciences, Cardiff University) and carefully placing a cleaned microscope slide over the molten agarose. After the gel had set, the gel on its slide was carefully peeled from the mould and the gel placed in a

normal submerged gel tank under TAE buffer. For RNA analysis the sample was mixed with loading dye and loaded onto the gel. An appropriate molecular weight marker was also loaded to ascertain size. The gel was stained with SYBR® greenII (Molecular Probes) and viewed on a UV transilluminator.

For analysis of Cy™ dye labelled targets, a 1µl aliquot of the sample was mixed with an equal volume of 50% glycerol and was carefully loaded into a well. Electrophoresis was performed at 90V for approximately 30 minutes. For molecular size analysis, either ALFexpress Sizer 50-500(Amersham Pharmacia Biotech) or DNA ladder of the appropriate size range (100bp or 1Kb Plus, Gibco BRL) labelled with Cy3™ or Cy5™ was run in an adjacent slot. Cy™ labelled cDNA was then visualised following scanning at the appropriate wavelength using the GeneTac™ LSIV scanner.

2.3.6. Analysis of RNA by gel Electrophoresis.

RNA samples were resolved by gel electrophoresis using dissociating conditions. The denaturing conditions were provided by NorthernMax™ buffers (Ambion) which are a proprietary mix containing formaldehyde and MOPS. Agarose gels were prepared by melting the required amount of agarose (1-2%w/v) in 1x TAE buffer. When below 50°C, NorthernMax™ denaturing gel buffer (Ambion) was added and the gel poured into a prepared mould. When solid, the gel was put in a gel tank and submerged in NorthernMax™ running buffer (Ambion). The appropriate amount of RNA was mixed with 5x RNA loading buffer (0.25% bromophenol blue, 4mM EDTA, 0.9M formaldehyde, 20% glycerol, 30% formamide, 4x NorthernMax gel buffer), was heated to 65°C for 5min then quenched on ice before loading into gel slots. Millenium™ markers (Ambion) or 0.1-1Kb RNA Marker (Sigma) were run in adjacent slots. The

gels were stained with SYBR™ Green Gold (Invitrogen) RNA stain diluted 1:5,000 in TAE buffer pH 8.5 Nucleic acid bands were then visualised under UV light.

2.4 Microarray

2.4.1. Slides.

The cDNA from the NIA15,000 mouse gene set were printed in a known format on to CMT-GAPS coated slides (Corning) using the Flexys™ High Density Arrayer (Genomic Solutions, Cambridge, UK). The cDNA was immobilised on the slide by first UV cross linking, achieved by placing the printed slides in a UV Stratalinker™ 2400 bench top transilluminator (Stratagene Ltd), then by baking within a lightproof box in an oven at 80°C for 2 hours. The printing and immobilisation of the slides was carried out by the Microarray Facility, School of Biosciences, Cardiff University.

If not used immediately, the microarrays were stored in a desiccator at room temperature, protected from light.

2.4.2. Blocking Of CMT-GAPS™ Coated Slides

Immediately prior to use for hybridisation, the slides were blocked to prevent non-specific binding of target DNA to reactive amine groups present in the Gamma Amino PropylSilane coating. This was achieved by placing the microarray in a 50ml Coplin jar containing blocking solution (1% Bovine Serum Albumin (Fraction V) (Sigma-Aldrich, Poole, UK), 5x SSC (20x SSC consists of 3M NaCl and 0.3M Na Citrate) and 0.1% SDS) pre-warmed to 42°C. Ensuring the slide was completely covered with blocking solution, blocking was performed in a water bath set at 42°C for 45 minutes. The microarray was then dipped briefly (2 seconds each) five times in sterile, filtered Milli-Q™ water. This was followed by a single 5-second dip in room temperature

isopropanol. The blocked slide was then air dried for approximately 10 minutes. After blocking, it was imperative that the microarray be utilised for hybridisation within 1 hour, since hybridisation efficiency decreases rapidly if left for any longer (Hegde *et al.*, 2000).

2.4.3. Using Target Labelled with Cy3/Cy5

The required amount of Cy3TM and Cy5TM labelled targets were combined in a sterile amber microcentrifuge tube and mixed thoroughly. The labelled cDNAs were reduced in volume to 18 μ l using a rotary evaporator at 30°C. To the 18 μ l of target, 1.0 μ l of each Poly-dA³⁰ (1mg ml⁻¹)(MWG Biotech) and human COT-1 DNA(1mg.ml⁻¹) (Gibco BRL) was added to the cDNA followed by denaturation at 95°C for 3 minutes to anneal to any poly-dT sequences. Prior to hybridisation, the volume of the labelled target was adjusted by the addition of hybridisation solution (2x hybridization solution, 50% formamide, 10% SSC, 0.2%SDS) and water. The labelled target was added to the microarray either in the hybridisation station (Genomic Solutions, Cambridge, UK) or under Hybri-slips (Sigma) and hybridised for 18 hours at 42°C.

2.4.4 Post Hybridization.

Following hybridisation unbound label was removed by washing once for 10 minutes at 55°C in Wash Solution 1(1xSSC, 0.2% SDS) then twice for 10 minutes each at 55°C in Wash Solution 2 (0.1% SSC, 0.2% SDS). The slides were finally rinsed in filtered Milli-Q water and dried with compressed air.

2.4.5 Scanning.

Hybridised and washed slides were scanned at the appropriate wavelength using the GeneTAC™ LS-IV scanner (Genomic Solutions, Cambridge, UK) or ScanArray® Express HT (Perkin Elmer)

Chapter 3

Labelling of Target

3. Labelling of Target

3. 1 Introduction

The success of an array experiment rests on the quality of the RNA target labelling. There are two commonly used labelling methods; fluorescent dyes or radioactive isotopes. Labelling can be achieved by three different methods; direct conjugation of label to the RNA, integration of labelled nucleotides into cDNA and introduction of nucleotides with an active side group into cDNA to which labelling molecules can be conjugated.

The most commonly used labelling procedures utilise either RNA polymerases (T7, T3 and SP6), the Klenow fragment of DNA polymerase or reverse transcriptase to incorporate modified nucleotides or deoxynucleotides into RNA or DNA. For example RNA polymerase is used to produce biotinylated RNA for hybridisation with Affymetrix™ chips. The RNA polymerase promoter is attached to the cDNA copy during the reverse transcription of the RNA. This is described in more detail in Chapter 4. Although the use of RNA polymerase has the advantage of amplifying the RNA target, the RNA product is more labile and is at risk of degradation by RNases which are plentiful in the environment.

Reverse transcriptase is probably the most used enzyme for labelling target, particularly for cDNA arrays. Although reverse transcriptase has no intrinsic ability to amplify the target, it has the advantage that it produces a stable cDNA target. The disadvantage of reactions driven by reverse transcriptase is that they tend to copy only the 3' ends of the mRNA species due to limited processivity of the enzyme and secondary structure of the RNA (Yeakley *et al.*, 2002)

3.1.1 Labelling Molecules.

Fluorescent probes are preferable because of their general safety and low toxicity. Radioactive probes are not only hazardous but the signal from them is very prone to spreading, an effect also known as blooming, such that the signal from spots of highly expressed genes will contaminate neighbouring spots.

Most microarray slide scanners use fluorescence as the detection system. Lasers excite fluorescent dyes integrated into the target at their maximum excitation wavelength and the emitted light is measured. The dyes Cy3™ and Cy5™ are most commonly used in microarraying because of their combination of desirable properties (Table 3.1). The Stokes shift for these dyes are 20 and 21 respectively thus simplifying the separation of fluorescent signal from the excitation light. The negatively charged sulphate groups give the dyes good solubility and reduce aggregation. This allows them to function well in the aqueous solutions required by enzymes and hybridisation. They also reduce quenching. The high extinction coefficient is a reflection of the efficiency of these fluorophores to absorb light. The emission maxima are well separated so both dyes can be detected on the same array spot with little cross talk between the two signals. The intensity of a fluorophore is proportional to both the extinction coefficient (ϵ), the measure of efficiency of light absorption and the quantum yield (ϕ), the measure of the efficiency of photon release. Cy3™ and Cy5™ both have high extinction coefficients, $150,000\text{M}^{-1}\text{cm}^{-1}$ and $250,000\text{M}^{-1}\text{cm}^{-1}$ respectively, and quantum yields of 0.14 and 0.15.

The quantum yield of dyes used in microarrays usually lies between 0.1 and 0.5. Fluorescein, which has an extinction coefficient of $67,000\text{M}^{-1}\text{cm}^{-1}$ and a quantum yield of 0.71, would be expected to give intensities similar to Cy5. However, fluorescein is

sensitive to photobleaching and to fluctuations in pH. This has led to the development of the Alexa dyes which have a similar ring structure to fluorescein but different side chains resulting in different excitation and emission spectra and different degrees of solubility. Cy3™ and Cy5™ are derived from the dyes C3 and C5 and differ from them only by the addition of the two sulphate groups which improves solubility.

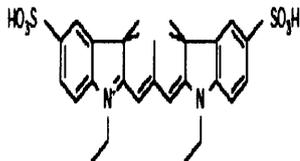
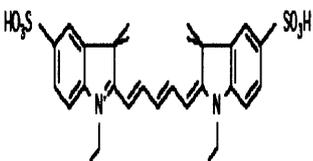
	Cy3™	Cy5™
Structure		
Stoke's Shift	20	21
Quantum Yield(Φ)	0.14	0.15
Extinction Coefficient(ϵ)	150,000M ⁻¹ cm ⁻¹	250,000M ⁻¹ cm ⁻¹
Absorption Maximum(nm)	550 Green	649 Red
Fluorescence Maximum(nm)	570 Yellow	670 red/far-red

Table3.1: Properties of the dyes Cy3™ and Cy5™

In this study Cy3™ and Cy5™ were used as the fluorescent labelling molecules.

3.1.2 Labelling of Target

3.1.2a Labelling of mRNA

Direct chemical labelling is at present the only way of avoiding a reverse transcription step. The method examined here couples fluorescent dyes to RNA or DNA using the ability of PtII compounds to attach to the N₇ of purine residues (Figure 3.1) (van Belkum *et al.*, 1994). After coupling of the Cy3™ or Cy5™ modified guanine to the

cytoplasmic RNA sample, the labelled mRNA is separated by using a standard mRNA selection kit. These kits consist of beads with polyT covalently bound to the surface. The polyA tail of the mRNA hybridises with the polyT and the beads can then be washed clean of free dye and other contaminants. The dye labelled mRNA is then released from the bead.

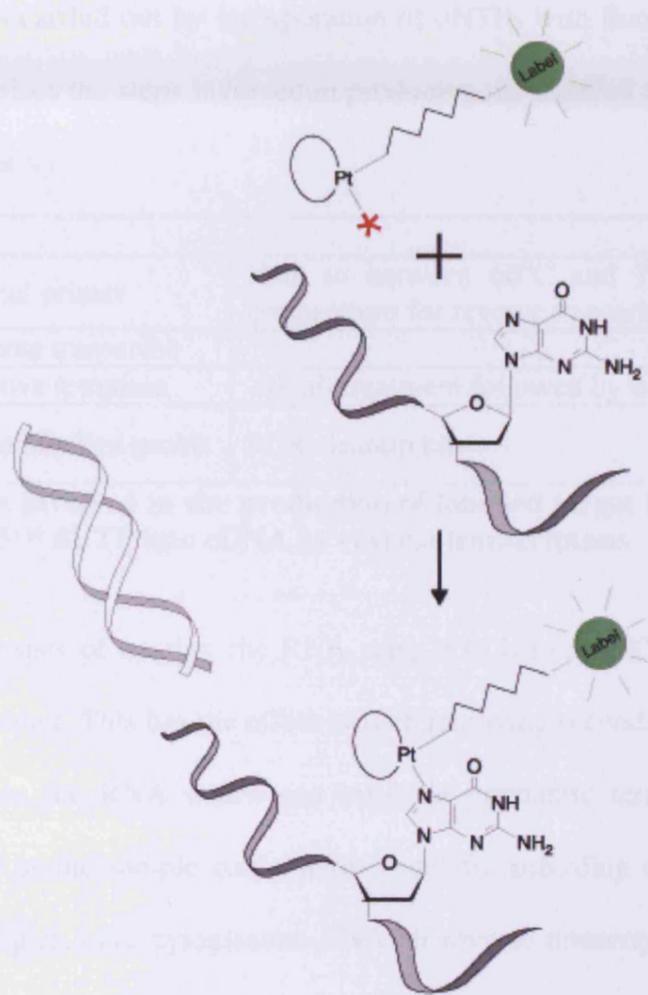


Figure 3.1. Diagram of labelled PtII complex binding to the N₇ of guanine.

Because no enzymes are involved there should be fewer problems with sequence specific variation and premature termination of the reverse transcriptase. Thus both dyes would be expected to be incorporated to the same extent and the problem of 3' bias

would be eliminated. This may be particularly useful when RNA is amplified before labelling; further selection by another round of enzyme treatment could be avoided. However instability of an RNA target remains a problem. Furthermore, as the target is sense, it can only be used with cDNA or antisense oligo arrays.

3.1.2b Direct labelling of cDNA

Direct labelling is carried out by incorporation of dNTPs with fluorescent side groups. Table 3.2 summarizes the steps involved in producing the labelled target by the method used in this laboratory.

Step		
1	Anneal primer	Heat to between 60°C and 70°C and cool to temperature for reverse transcriptase
2	Reverse transcribe	
3	Remove template	Alkali treatment followed by neutralisation
4	Clean labelled probe	PCR cleanup kit.

Table 3.2: Steps involved in the production of labelled target by incorporation of Cy3™ or Cy5™ dUTP into cDNA by reverse transcriptase.

The first step consists of heating the RNA sample to between 60°C and 70°C in the presence of the primer. This has the effect of both removing secondary structure such as hairpin loops from the RNA which can result in premature termination of reverse transcription, and as the sample cools, allows specific annealing of the primer to the RNA. As the samples were cytoplasmic RNA all reverse transcription reactions were primed using oligo-(dT) which anneals preferentially to the poly-A tail of the mRNA. Where the primer was not supplied with a kit, oligo-(dT)₂₃ with a single deoxynucleotide, either dA, dG or dC at the 3' end as the anchor was used. This ensures the hybridisation of the primer at the 5' end of the message.

In the second step, a master mix containing the rest of the reverse transcription reaction components; enzyme, buffer, and a dNTP mix including Cy3™ or Cy5™ labelled dUTP, was added to the RNA and its annealed primer.

For efficient target hybridisation, the unlabelled RNA template must be degraded. This is achieved in Step3 treatment with by alkali followed by neutralisation. The labelled cDNA is then cleaned of enzymes and oligos with a PCR cleanup kit.

3.1.2c Indirect labelling of cDNA.

To alleviate the problem of uneven incorporation of the two dyes commonly seen with direct incorporation of Cy dye conjugated dUTP and to keep the advantages of a stable cDNA product, many workers have turned to indirect labelling of the cDNA.

Although the procedure has more steps and does not avoid the 3' bias inherent in the reverse transcription, the use of indirect labelling with aminoallyl dUTP is reputed to avoid some of the problems associated with the direct integration of Cy™ labelled dUTP. As the aminoallyl group is small, it is believed to be as efficiently incorporated into cDNA as unmodified and therefore integrates into the cDNA at a higher density thus resulting in more labelled target for a given amount of RNA template than for Cy-dUTP. The same molecule, the primary amine aminoallyl dUTP, is incorporated into both the samples being compared so the targets should be more similarly labelled.

The aminoallyl group is attached to the C₅ on the dUTP base and has a 2 carbon spacer. This position ensures it will interfere least with ability of the enzyme to incorporate it and where the fluorescent side groups will be less likely to interfere with hybridisation (Figure 3.2)

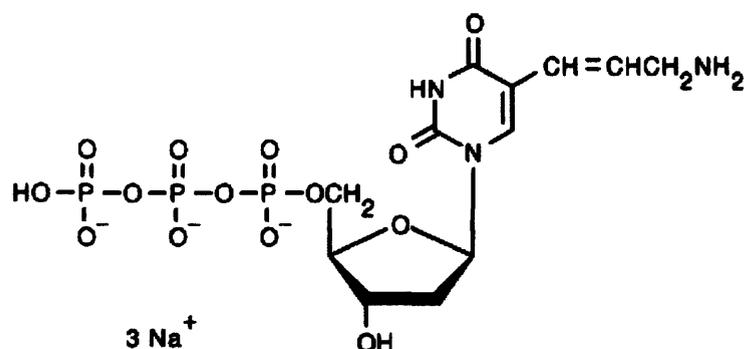


Figure 3.2: Aminoallyl dUTP

The process of indirect labelling can be broken down into a number of discrete steps each of which can be achieved in a number of ways, thus many factors can influence the outcome of labelling target (Table 3.3).

Step		Methods available
1	Anneal primer	
2	Reverse transcribe	Enzyme, ratio of aminoallyl dUTP to dTTP
3	Remove template	RNase H, alkali treatment followed by neutralisation
4	Clean cDNA and exchange buffer	PCR cleanup kit, ultrafiltration, resin + precipitation
5	Bind label	
6	Clean labelled probe	PCR cleanup kit, ion exchange chromatography, desalting, precipitation

Table 3.3: Steps involved in the production of labelled target by incorporation of aminoallyl dUTP into cDNA by reverse transcriptase.

Before the aminoallyl modified cDNA is reacted with the Cy dye it must be cleaned of enzyme, dNTPs, primers and oligos. Also the buffer must be changed from TRIS, which would compete with the coupling of the reactive monofunctional N-hydroxysuccinimide Cy™ dyes to the aminoallyl groups on the cDNA (Figure 3.3).

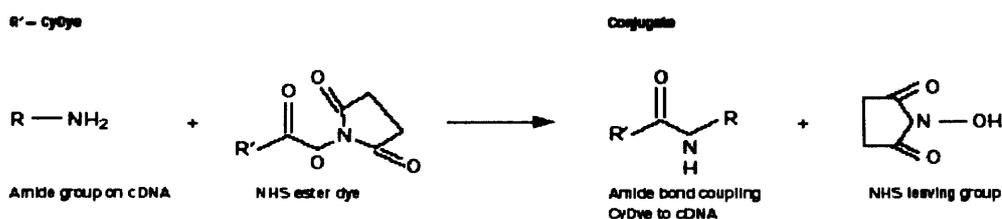


Figure 3.3: Reaction between the primary amine group on aminoallyl dUTP and the monoreactive CyTM dye

One method, the QIAquickTM PCR kit, is designed for purification of double- or single-stranded DNA between 100bp and 10kb and utilises a silica-gel membrane in a spin column. The DNA adsorbs to the silica-membrane in the presence of high salt and the degraded RNA oligomers and unincorporated CyDye-nucleotides are washed away. The labelled cDNA is then eluted in water. Another method uses a resin to remove enzymes followed by ethanol precipitation to remove the primers, oligos and dNTPs. This also achieves the removal of the TRIS buffer. Reactions can also be cleaned and the TRIS buffer removed by applying to a Microcon-30TM ultrafiltration unit. These devices consist of a filter with a nominal molecular weight cutoff of 30,000 Daltons in a centrifuge tube. Sequentially adding water and spinning the tube achieves both desalting and concentration of the sample.

The conjugation of the monoreactive Cy dye to the aminoallyl groups on the cDNA takes place at pH9.0 in a buffer without TRIS at room temperature, usually 0.1M NaHCO₃.

The final clean up of the labelled target to remove the unbound dye can be achieved also by PCR cleanup kit, ion exchange chromatography, desalting or precipitation.

At the time when this project was initiated, there were few well optimised protocols for labelling. Also the development of commercial kits was at early stages. The kit most often used in our laboratory was the CyScribe™ direct labelling kit which incorporates Cy™ dye labelled dUTP into cDNA by means of reverse transcription. Even then it was known that this labelling method suffered from a number of drawbacks the most important of which was that Cy3™ was incorporated at a much higher density into the cDNA than the larger and more hydrophobic Cy5™.

A number of experiments were undertaken to examine the different methods of labelling and to test some of the newly emerging kits as they became available. Fluorescently labelled target was produced by integration of modified nucleotides into cDNA using cytoplasmic RNA as the template for the DNA polymerase reverse transcriptase. A summary of the experiments involving incorporation of label by reverse transcriptase is provided in Table 3.4.

Cytoplasmic RNA has the advantage of being full length and fully processed. In the direct labelling protocol, dUTP conjugated to the fluorescent dyes, Cy3™ or Cy5™ was incorporated into the cDNA using the CyScribe™ Direct Labelling kit. Indirect labelling was achieved by integration of dUTP modified by the addition of an amino group with a short spacer, aminoallyl-dUTP. Monofunctional N-hydroxysuccinimide activated fluorescent dyes were then coupled to the amino groups (Figure 3.3). A number of experiments were conducted based on a method published on the web by Joseph DeRisi (<http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf>). Three commercial kits using integration of aminoallyl dUTP (Atlas™ Glass Fluorescent Labelling Kit, Atlas™ PowerScript™ Fluorescent Labelling Kit and CyScribe™ Post-Labelling Kit) were also assessed.

	Direct Incorporation CyScribe™ First-Strand Labelling Kit	Indirect Incorporation http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf	Indirect Incorporation Atlas™ Glass Fluorescent Labelling Kit	Indirect Incorporation Atlas™ PowerScript™ Fluorescent Labelling Kit	Indirect Incorporation CyScribe™ Post-Labelling Kit
Maximum Input RNA	10ug	10ug	10ug	8ug	5ug
Anneal Primer	70°C, 5min, anchored oligo-(dT)	65°C, 5min, 160pmol anchored oligo-(dT)	70°C, 5min, 70µM anchored oligo-(dT)	70°C, 5min, 70µM anchored oligo-(dT)	70°C, 5min, anchored oligo-(dT)
Reverse Transcribe	1.5h, 42°C	1-2h, 42°C	30min, 42°C	1h, 42°C	1.5h, 42°C
Destroy Template	NaOH	NaOH	RNasin	RNasin	NaOH
Clean cDNA	Not applicable	Microcon-30™ or ethanol precipitation	ethanol precipitation	ethanol precipitation	ethanol precipitation
Bind Label	Not applicable	0.1M NaHCO ₃ , pH9.0	Supplied buffer	Supplied buffer	0.1M NaHCO ₃ , pH9.0
Clean Labelled Probe	QIAquik™ PCR Cleanup Kit	QIAquik™ PCR Cleanup Kit	ethanol precipitation followed by Atlas NucleoSpin™ Extraction Kit	ethanol precipitation followed by QIAquik™ PCR Cleanup Kit	QIAquik™ PCR Cleanup Kit
Full Protocol	Table 3.7	Table 3.8	Table 3.9	Table 3.10	Table 3.11
Table 3.4: Summary of the methods used to integrate Cy dye labelled molecules into cDNA using reverse transcription					

The principle of the DeRisi method and all three kits was similar, with differences only in detail. To check if problems arising were inherent in the method or due to poor optimization, the performance of two kits from the same manufacturer were assessed. The PowerScript™ kit was supposed to be an improvement on the Atlas™ Glass kit.

Direct chemical labelling of RNA was achieved by platinum-linked cyanine dyes (Gupta *et al.*, 2003) using the MICROMAX™ ASAP RNA Labelling Kit.

3.2 Materials and Methods

The ES line IMT-11 (Bierbaum *et al.*, 1994) was cultured as described in Section 2.1. Cytoplasmic RNA was purified from the cells at passage 17 according to the protocol described (Section 2.3.2). The same sample was used for all experiments.

All labelled targets were either used immediately or stored at -20°C protected from light.

Analysis of the probes was carried out by spectrophotometric analysis (Section 2.3.4) and slide gel electrophoresis (Section 2.3.5). The amount of cDNA synthesised and the incorporation of dyes was calculated by application of Beer-Lambert's Law.

<p style="text-align: center;">Beer-Lambert's Law $A = \epsilon lc$ where A is the absorbance ϵ is the extinction coefficient in $\text{cm}^{-1} \text{M}^{-1}$ l is the pathlength in cm c is the concentration in mol l^{-1}</p>
--

The amount of cDNA synthesised and the incorporation of Cy™ dye was quantified by measuring the optical density of the labelled cDNA at 260nm and 550nm for Cy3™ and

at 260nm and 650nm for Cy5™. The amount of cDNA synthesised was calculated as $1 A_{260} \text{unit} = 37 \mu\text{g/ml}$. The number of pmols of dye incorporated into cDNA was calculated using the formulae: $\text{pmol Cy3} = \text{OD}_{550} \times 100/0.15$, where 100 is the volume of the probe in μl and 0.15 is the extinction coefficient for Cy3™ and $\text{pmol Cy5} = \text{OD}_{650} \times 100/0.25$ where 0.25 is the extinction coefficient for Cy5™.

The FOI for each labelled sample, which is the number of Cy-dUTPs per 1,000 nucleotides of cDNA, was also calculated according to the formulae: $\text{FOI Cy3} = \text{OD}_{550} \times 58.5/\text{OD}_{260}$ and $\text{FOI Cy5} = \text{OD}_{650} \times 35.1/\text{OD}_{260}$

See Appendix 2 for details of calculations.

Slide gels (2.3.5) were also run to assess the molecular weight spread and the level of residual free dye.

Kits and reagents were purchased as listed in Table 3.5. All the manufacturers' protocols are reproduced for ease of comparison.

CyScribe™ First-Strand Labelling Kit	Amersham Pharmacia Biotech
Advantage reverse transcriptase	BD Biosciences, Paola Alto, Ca
Aminoallyl dUTP	Sigma
Anchored oligo-(dT)	Sigma
Atlas™ Glass Fluorescent Labelling Kit	BD Biosciences, Paola Alto, Ca
Atlas™ PowerScript™ Fluorescent Labelling Kit	BD Biosciences, Paola Alto, Ca
Cy3™, monoreactive NHS ester derivatized	Amersham Pharmacia Biotech
Cy3™-dUTP	Amersham Pharmacia Biotech
Cy5™, monoreactive NHS ester derivatized	Amersham Pharmacia Biotech
Cy5™-dUTP	Amersham Pharmacia Biotech
CyScribe™ Post- Labelling Kit	Amersham Pharmacia Biotech
dNTP set 100mM each	Invitrogen
Microcon™ YM-30 filters	Amicon
MICROMAX™ ASAP RNA Labelling Kit	PerkinElmer Life Sciences
MMLV reverse transcriptase	Promega
Oligotex™ mRNA Mini Kit	Qiagen
QIAquick™ PCR purification kit	Qiagen
Superscript II™ reverse transcriptase	Invitrogen

Table 3.5: Reagents and kits used.

3.2.1 Direct Labelling

3.2.1a. Production of fluorescently labelled target RNA using the MICROMAX ASAP kit.

The MICROMAX ASAP kit conjugates the Cy dyes directly to the N₇ of the guanine and adenine residues of the RNA through a platinum universal linker. (van Belkum *et al.*, 1994)

Cytoplasmic RNA (10µg) was labelled with Cy dye and then labelled mRNA purified from the sample using the Oligotex™ mRNA Mini Kit as described in Table 3.6.

Instead of combining the Cy3™ and Cy5™ labelled RNA before purifying the mRNA, as in the manufacturer's protocol, the two labelled products were handled separately so that each could be analysed by spectrophotometry and slide gel electrophoresis.

1	Labelling	A10µg sample of cytoplasmic RNA was combined with the provided labelling buffer in a total volume of 20µl. To this, 2µl of the labelling reagent was added and the reaction mixture was incubated at 85°C for 15 min then was cooled on ice. The reaction was stopped by the addition of the ASAP stop buffer.
2	Purification of mRNA	<p>The Qiagen Oligotex™ mRNA Mini Kit was used to purify the labelled mRNA..</p> <p>Prepared Oligotex suspension by heating to 37°C, mixing by vortex and holding at room temperature. Buffer OEB was preheated to 70°C.</p> <p>The volume of the sample was adjusted to 250µl with nuclease free water then 250µl of buffer OBB and 15µl of Oligotex suspension were added.</p> <p>The sample was incubated at 70°C for 3min then held at room temperature (between 20°C and 30°C) for 10min.</p> <p>The Oligotex-mRNA complex was pelleted by centrifugation at 14000g for 2min in a microcentrifuge and the supernatant was removed.</p> <p>The Oligotex-mRNA was resuspended in 400µl Buffer OW2 and transferred to a small spin column inside a 1.5ml microcentrifuge tube.</p> <p>This was then centrifuged at 14000g for 1min.</p> <p>The tube with the flowthrough was discarded, the column was transferred to a fresh tube, 400µl of fresh Buffer OW2 was added to the column and the centrifugation was repeated.</p> <p>The column was transferred to a fresh amber 1.5ml centrifuge tube and the labelled mRNA was eluted by resuspension of the resin in 40µl of Buffer OEB at 70°C and centrifugation for 1min at 14000g. This elution step was repeated.</p>

Table 3.6. Protocol for the production of labelled mRNA using the MicroMax ASAP kit.

3.2.1b Production of fluorescently labelled target cDNA using the CyScribe™ First Strand cDNA Labelling Kit.

The CyScribe™ First Strand Labelling Kit is designed to incorporate either Cy3™ or Cy5™ labelled dUTP into cDNA by reverse transcription. All components except the NaOH, HEPES and the QIAquick™ purification kit components were supplied with the

CyScribe™ First strand cDNA Labelling Kit. The QIAquick™ kit was supplied with all components except the nuclease free water.

Labelling of 10ug of target was carried out using the protocol supplied with the kit as described in Table 3.7. The probes were analysed by spectrophotometry and the amount of cDNA synthesised, the amount of dye incorporated into the cDNA and the number of dye molecules incorporated per 1000 nucleotides, the FOI, were calculated.

1	Anneal primer	Mixed in PCR tube: 10µg cytoplasmic RNA 1µl anchored oligo-(dT) Nuclease free water to 11µl Heated to 70°C for 5min. Cooled to room temperature 10min.
2	Reverse transcribe	The following master mix was assembled on ice and added to the sample 4µl of 5x CyScribe™ buffer, 2µl 100mM DTT, 1µl dUTP nucleotide mix, 1µl of 1mM Cy3™ or Cy5™ labelled dUTP 1µl CyScript™ reverse transcriptase The 20µl reverse transcription reaction was incubated for 1.5h at 42°C.
3	Remove template	2µl of 2.5M NaOH was added and the mixture was incubated at 37°C for 15min, then neutralised with 10µl of 2M HEPES.
4	Clean labelled probe	QIAquick™ PCR Cleanup Kit Five volumes of Buffer PB were added to the sample and the whole volume was applied to a QIAquick spin column in a 2ml collection tube. The cDNA was bound to the membrane by spinning at 13,000g for 1 min in a microcentrifuge and the flowthrough discarded. The column was washed once with 700µl, and then twice with 400µl Buffer PE discarding the flowthrough after each 1min centrifugation at 13,000g. The column was dried of residual ethanol by centrifugation for an additional minute. The probe was then eluted from the column into a sterile, amber microcentrifuge tube by adding 50µl of nuclease free water to the centre of the membrane and incubation for 1min at room temperature before centrifugation for 1min at 13000g. A repeat of this elution step completed the release of the purified probe.

Table 3.7: Protocol for production of labelled cDNA with the CyScribe™ Direct labelling kit.

3.2.2. Indirect Labelling

3.2.2a. Protocol to test parameters of fluorescently labelled target production by aminoallyl labelling.

The protocol used and described in Table 3.8 was adapted from one published by Joseph DeRisi on the web <http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf>.

All the experiments carried out using this protocol were performed with 10 μ g of cytoplasmic RNA. Labelled target was cleaned with the QIAquick™ cleanup kit using a modified washing protocol (Table 3.7, Step 6)

1	Anneal primer	Mixed in PCR tube: 10µg cytoplasmic RNA 150pmol anchored oligo-(dT)(Sigma) Nuclease free water to 20µl (Ambion) Heated to 65°C for 5min Cooled to the reaction temperature being tested
2	Reverse transcribe	A master mix was assembled on ice and added to the sample: 8µl of 5x First Strand Buffer (as provided with the enzyme under test), 10mM DTT 0.5mM each of dG, dA and dC mix. Aminoallyl-dUTP and dTTP, each at 5mM were added in varying proportions such that the final concentration of aminoallyl-dUTP + dTTP was 0.5mM, the same as for each of the other dNTPs. Rnasin (Invitrogen) 40U Reverse transcriptase, 400U, was added. The complete reaction volume was 40µl. The reverse transcription reactions were incubated at 42°C for between 1 and 2h.
3	Remove template	Added: 4µL of 50mM EDTA 2µL of 10N NaOH Incubated at 65°C for 20min. Samples were neutralised with 4µL of 5M Acetic Acid.
4	Clean cDNA and exchange buffer	450µl of nuclease free water was added to the reaction tube and applied to a Microcon-30™ ultrafiltration unit Spun for 8min at 12,000g in a microcentrifuge. 500µl of water was added to the ultrafiltration unit Repeated centrifugation for 8min at 12,000rpm. This wash step was repeated twice. The sample reservoir was inverted over a fresh collection tube and spun at 4,000g to transfer the sample. The sample was then reduced in volume in a rotary evaporator to 1-2µl.

Table 3.7: Protocol for the production of labelled cDNA by the incorporation of aminoallyl dUTP using a modification of the method of Joseph DeRisi.
<http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf> continued on next page

4*	Alternative clean cDNA and exchange buffer	Alternatively; removal of oligonucleotides, primers and exchange of buffers was achieved by ethanol precipitation (2.3.2.)
5	Bind label	A 12.5mM stock of mono-functional Cy3™ and Cy5™ in DMSO was made and 1.25µl was added to the cDNA which was dissolved in 9µl of 0.1M NaHCO ₃ pH9.0. The coupling reaction was incubated for 45min at room temperature in the dark. The reactions were quenched by the addition of 4.5µl of 4M hydroxylamine and incubation in the dark, at room temperature for 15min.
6	Clean labelled probe	Samples were then purified with a Qiagen QIAquick™ PCR purification kit with a modified washing procedure. The whole protocol was executed at room temperature. All centrifugation steps were carried out at 13,000g in a microcentrifuge. The column was washed 3 times with 500µl of the wash buffer PE before elution with the buffer EB. Elution was carried out by incubation of the column with 50µl of EB for 1min followed by centrifugation for 1min. The elution step was repeated once.

Table 3.7 contd: Protocol for the production of labelled cDNA by the incorporation of aminoallyl dUTP using a modification of the method of Joseph DeRisi. <http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf>

3.2.2b Comparison of three commercially available kits.

Three commercially available kits were tested.

3.2.2b.1 Production of Fluorescently Labelled Target cDNA using the CLONTECH Atlas™ Glass Fluorescent Labelling Kit

The CLONTECH Atlas™ Glass Fluorescent Labelling kit was used according to the manufacturer's protocol described in Table 3.8. The NucleoSpin™ purification kit was supplied with the Atlas™ Glass Fluorescent Labelling Kit. This kit is similar in function and in use to the QIAquick™ Cleanup kit used in the other protocols.

This kit also recommends a maximum of 10µg input of total RNA. In this case 10µg of cytoplasmic RNA as in the other protocols was used.

3.2.2b.2 Production of fluorescently labelled target cDNA using the CLONTECH Atlas™ PowerScript™ Fluorescent Labelling Kit.

The CLONTECH Atlas™ PowerScript™ Fluorescent Labelling kit was used according to the manufacturer's protocol (Table 3.9). For this kit a maximum of 8µg of cytoplasmic RNA was combined, in a PCR tube, with anchored oligo-(dT) (Sigma) and the reverse transcription was carried out in exactly the same way as for the Atlas™ Glass Fluorescent Labelling Kit. As no purification kit is supplied with this kit, the QIAquick™ Cleanup Kit was purchased separately. A modification of the QIAquick™ Cleanup Kit protocol as described in Table 3.6 Step 6 was used.

Instead of using RNase H to degrade the template, in some cases NaOH was used (Table 3.9, Step3*).

1	Anneal primer	Cytoplasmic RNA (10µg) was combined with 70µM anchored oligo-(dT) (Sigma) to give a concentration of 5µM in the final 50µl synthesis reaction and the volume was made up to 25µl with nuclease free water. The tube was then heated to 70°C for 5min then cooled to the temperature being tested.
2	Reverse transcribe	A master mix was assembled from 10µl of 5x cDNA Synthesis Buffer, 5µl of 10x dNTP Mix, 7.5µl of nuclease free water 2.5µl (200units/µl) MMLV Reverse Transcriptase provided in the kit, total volume 25µl. This was added to the RNA/oligo-(dT) and the reaction was allowed to proceed for the 30min. At the end of the incubation period, the reaction was heated to 70°C for 5min to inactivate the enzyme then was spun briefly in a microcentrifuge to collect the tube contents
3	Remove template	The tube was then equilibrated at 37°C, 0.5µl RnaseH (10units/µl) were added and incubation continued for 15min to degrade the template.
4	Clean cDNA and exchange buffer	Reactions were stopped, cleaned and the TRIS buffer removed by the addition of 0.5µl EDTA (pH8.0) and 5µl QuickClean™ resin. After mixing by vortexing for 1minute, the tube was spun in a microcentrifuge and the supernatant transferred to a 0.45µm spin filter in a collection tube. The filter assembly was then spun at full speed for 1minute. The cDNA was precipitated from the eluate by ethanol (Section 2.3.2) at -20°C for a minimum of 1h. The cDNA was then pelleted by centrifugation for 20min at full speed in a microcentrifuge. The pellet was washed once in 70% ethanol then airdried.
5	Bind label	The cDNA pellets were dissolved in 10µl of 2x Fluorescent Labelling Buffer. A 5mM solution of mono-reactive fluorescent dye, Cy3 or Cy5, was prepared by adding 45µl of DMSO to a tube of dye. To the cDNA solution, 10µl of the dye was added and coupling was allowed to proceed for between 30min and 1h in the dark.

Table 3.8: Protocol for the production of fluorescently labelled Target cDNA using the CLONTECH Atlas™ Glass Fluorescent Labelling Kit. Continued on next page.

6	Clean labelled probe	<p>The dye-coupled cDNA was precipitated ethanol (Section 2.3.2) at -20°C for a minimum of 2h. The cDNA was then pelleted by centrifugation for 20 minutes at full speed in a microcentrifuge. The pellet was washed once in 70% ethanol then air-dried. The dried pellet was dissolved in 100µl nuclease free water and cleaned as described in the manufacturer's protocol with the Atlas NucleoSpin™ Extraction Kit provided with the labelling kit.</p> <p>The whole process was carried out at room temperature. All centrifugation steps were carried out at 13,000g in a microcentrifuge. 400µl of Buffer NT2 was added to the sample and the thoroughly mixed sample was then loaded on to a column. The column was spun at for 1min. The column was washed 3 times with 700µl of Buffer NT3, spinning as before. The column was dried of residual Buffer NT3 by spinning for a further 2min. The sample was eluted into a fresh amber microcentrifuge tube by applying 50µl of Buffer NE to the column, incubating for 1min and spinning for 1min. This elution step was repeated once.</p>
---	----------------------	--

Table 3.8 Contd: Protocol for the production of fluorescently labelled Target cDNA using the CLONTECH Atlas™ Glass Fluorescent Labelling Kit.

1	Anneal primer	A maximum of 8ug of total RNA was combined, in a PCR tube with 70µM anchored oligo-(dT) (Sigma) to give a concentration of 5µM in the final 20µl synthesis reaction and the volume was made up to 10µl with nuclease free water. The tube was then heated to 70°C for 5min.
2	Reverse transcribe	A master mix was assembled from 10µl of 5x cDNA Synthesis Buffer, 5µl of 10x dNTP Mix, 7.5µl of nuclease free water and 2.5µl (200units/µl) MMLV Reverse Transcriptase provided in the kit, total volume 25µl. This was added to the RNA/oligo-(dT) and the reaction was allowed to proceed for 1h. At the end of the incubation period, the reaction was heated to 70°C for 5min to inactivate the enzyme then was spun briefly in a microcentrifuge to collect the tube contents
3	Remove template	The tube was then equilibrated at 37°C, 0.5µl RnaseH (10units/µl) were added and incubation continued for 15min to degrade the template.
3*	Remove template	The RNA template and other RNA species were degraded by the addition of 2µl of 2.5M NaOH and incubation at 37°C for 15min, followed by neutralisation with 10µl of 2M HEPES.
4	Clean cDNA and exchange buffer	Reactions were stopped, cleaned and the TRIS buffer removed by the addition of 0.5µl EDTA (pH8.0) and 5µl QuickClean™ resin. After mixing by vortexing for 1min, the tube was spun in a microcentrifuge and the supernatant transferred to a 0.45µm spin filter in a collection tube. The filter assembly was then spun at full speed for 1min. The cDNA was precipitated from the eluate by ethanol (Section 2.3.2) at -20°C for a minimum of 1h. The cDNA was then pelleted by centrifugation for 20min at full speed in a microcentrifuge. The pellet was washed once in 70% ethanol then airdried.
5	Bind label	The cDNA pellets were dissolved in 10µl of 2x Fluorescent Labelling Buffer. A 5mM solution of mono-reactive fluorescent dye, Cy3 or Cy5, was prepared by adding 45µl of DMSO to a tube of dye. To the cDNA solution, 10µl of the dye was added and coupling was allowed to proceed for between 30min and 1h in the dark.

Table 3.9: Protocol for the production of fluorescently labelled target cDNA using the CLONTECH Atlas™ PowerScript™ Fluorescent Labelling Kit. Continued on next page.

6	Clean labelled probe	The dye-coupled cDNA precipitated from by ethanol (Section 2.3.2) at -20°C for a minimum of 2h. The cDNA was then pelleted by centrifugation for 20min at full speed in a microcentrifuge. The pellet was washed once in 70% ethanol then airdried as before. The dried pellet was dissolved in 100µl nuclease free water and cleaned with the QIAquick™ kit (Table 3.6 Step 6).
---	----------------------	--

Table 3.9 contd: Protocol for the production of fluorescently labelled target cDNA using the CLONTECH Atlas™ PowerScript™ Fluorescent Labelling Kit.

3.2.2b.3 Production of fluorescently labelled target using the CyScribe™ cDNA**Post Labelling Kit**

The CyScribe™ cDNA Post Labelling kit was used according to the manufacturer's protocol (Table 3.11). Labelling was performed with 5µg of RNA using the CyScribe™ cDNA Post labelling kit. Purification of the labelled targets was carried out using the QIAquick™ cleanup kit with the modified protocol described in Table 3.8 Step 6. The labelled targets were analysed by spectrophotometry and quantified as before.

1	Anneal primer	Mixed in PCR tube: 5µg cytoplasmic RNA 1µl anchored oligo-(dT) Nuclease free water to 11µl Heated to 70°C for 5min Cooled to room temperature 10min.
2	Reverse transcribe	A master mix was assembled from 4µl of 5x CyScribe™ buffer 2µl 100mM DTT 1µl dUTP nucleotide mix 1µl AA-dUTP 1µl CyScript™ reverse transcriptase Total reaction volume 20µl. The reverse transcription reaction 1.5h at 42°C.
3	Remove template	Added 2µl of 2.5M NaOH, incubated at 37°C for 15min, and neutralised with 10µl of 2M HEPES.
4	Clean cDNA and exchange buffer	Ethanol precipitated the cDNA (Section 2.3.2) at -20°C for a minimum of 2h. The cDNA was then pelleted by centrifugation for 20min at full speed in a microcentrifuge. The pellet was washed once in 70% ethanol then airdried.
5	Bind label	The cDNA pellet was dissolved in 15µl of nuclease water and mixed with the contents of one vial of mono-reactive Cy3™ or Cy5™ (as supplied in the kit) dissolved in 15µl 0.1M Sodium Bicarbonate pH9.0. The dye coupling reaction was incubated at room temperature for 1h in the dark. Unreacted Cy dye NHS ester molecules were inactivated by the addition of 15µl of 4M Hydroxylamine.
6	Clean labelled probe	The cDNA was then purified using a QIAquick™ kit as in Table 3.6 Step 6.

Table 3.10: Protocol for the production of fluorescently labelled target using the CyScribe™ cDNA Post Labelling Kit

3.2.3 Comparison of Arrays Hybridised with Target prepared by the Atlas™ PowerScript™ Fluorescent Labelling Kit, the Template being Destroyed either by RNase H or NaOH.

Fluorescently labelled target was prepared using the Atlas™ PowerScript™ Fluorescent Labelling Kit and 5µg of cytoplasmic RNA (Table 3.9). The RNA template was destroyed (Step3) by either RNaseH or by the alternative NaOH (Step 3*)

Slides printed with the NIA 15K mouse gene set were prepared and hybridisations carried out with 15pmol of dye incorporated into the targets according to the protocols in Section 2.4. One slide for each pair of targets were hybridised. The targets were hybridised with array slides according to the plan in Table 3.12.

Slide Number	Experiment Number	Cy3	Cy5
1	1	RNase	RNase
2	1	NaOH	NaOH
3	2	RNase	RNase
4	2	NaOH	NaOH

Table 3.11: Hybridisation plan

The hybridised and washed slides were scanned in a ScanArray Express HT scanner (Perkin Elmer) at 65pmt (photo multiplier tube) and 90% laser power gain. The scan resolution was 10µ and was run at full speed. The pmt gain was derived empirically to maximise sensitivity and minimise both signal to noise ratio and saturation. The value recommended by the manufacturer is 70-80%, however with these arrays this led to saturation of the highly expressed spots. The laser power and scan resolution were set according to the manufacturer's recommendations.

Image analysis of the slides was carried out using the software ImaGene5.5™ (Biodiscovery). Spots were detected using automatic spot finding with the grid constraints enforced and spot detection set with local flexibility of 20 pixels and global flexibility set just above half (Fig 3.4). These settings maximise the spots found while minimising the false detection of contamination as spots. As there was no evidence of negative spots (spots which show on the slides as holes with signal less than the background) this was left unmarked. A spot was designated as empty (Flag 2) if both channels were so flagged and as poor (Flag 3) or negative (Flag 4) if either channel was flagged. The threshold for the designation of a spot as empty was adjusted to 0.6 by using an unflagged Blank spot, and increasing the threshold till it was flagged .

The data from the two experiments was combined such that only spots which were good in both experiments (unflagged) were included. For NaOH this was 1746 spots and for RNase 1219 spots. These spots were further analysed using the linest function; which fits a linear regression line to the data by the least squares method and calculates the test statistics; in Microsoft Excel™ 2003 forcing the intercept to go through the origin. Scatter plots were generated of Cy3™ against Cy5™ and a trendline produced. Statistics from the linest output were examined.

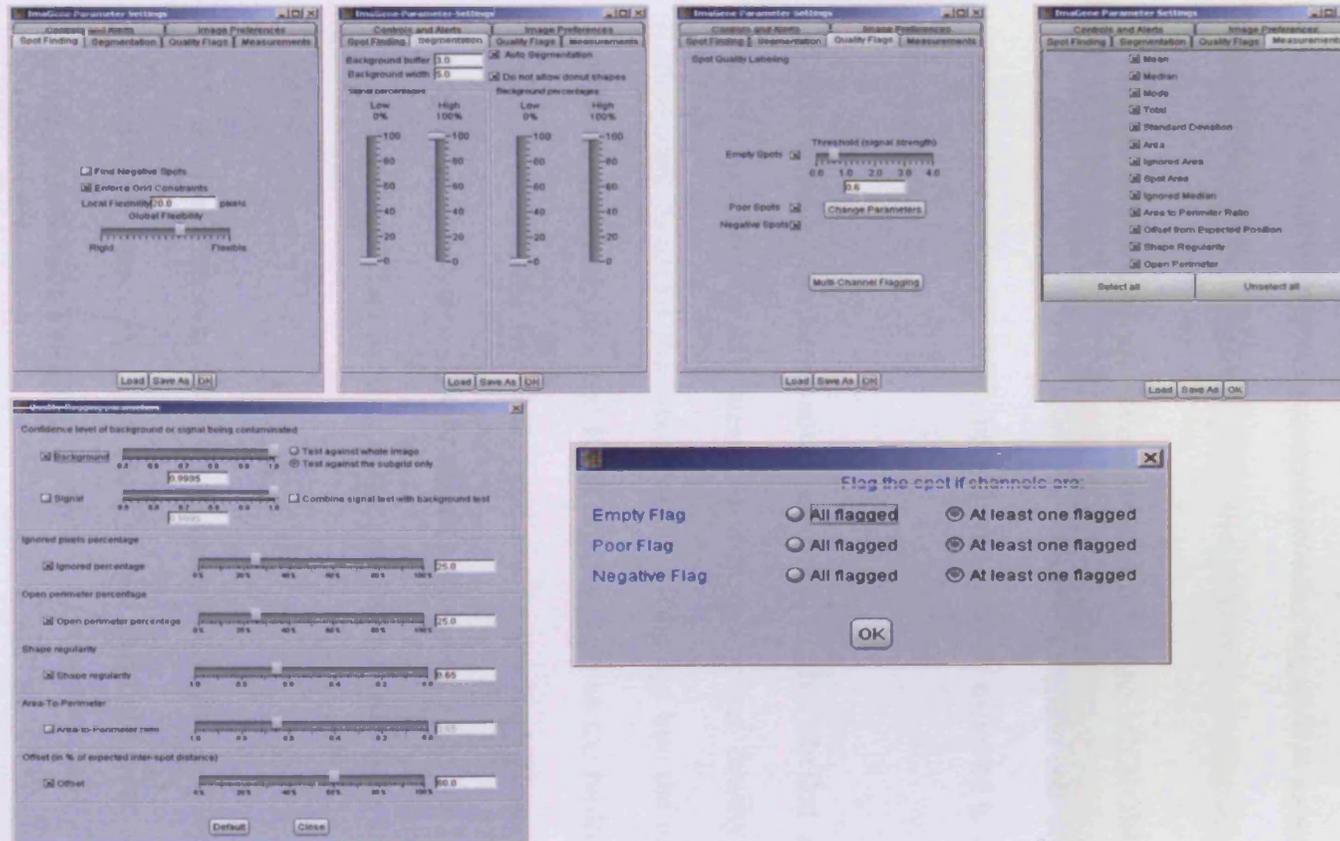


Figure 3.4: Screen shots of the settings used in the image analysis of the slides with RNase or NaOH treated aminoallyl cDNA

3.3 Results and Discussion

3.3.1 Direct Labelling

3.3.1a. Production of fluorescently labelled target RNA using the MICROMAX ASAP kit.

Because no enzymes are involved when using the MICROMAX ASAP kit, there should be fewer problems with sequence specific variation and premature termination of the reverse transcriptase. Also, both dyes would be expected to be incorporated to the same extent.

Unfortunately, the fines sometimes left in the labelled mRNA by the Oligotex™ mRNA purification kit interfered with the optical density of the samples making it impossible to assess the incorporation of the dye into the mRNA. Also, it proved very difficult to run the labelled RNA on slide gels so the molecular weight spread and the amount of residual free dye could not be assessed. However, when hybridised to an array slide, the degree of hybridisation was good with little background problems. There is no numerical data for this as the array slides used were only partially printed (Figure 3.5).

It was also discovered while using this kit that the Cy5™ labelling reagent was considerably less stable than the Cy3 and that both were very unstable in the summer months due to high humidity.

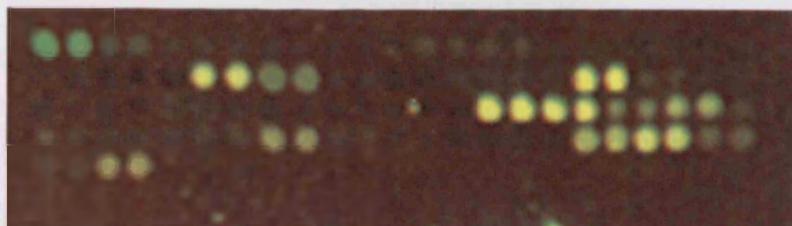


Figure 3.5: Example area of array hybridised with mRNA labelled with the MICROMAX™ ASAP kit (PerkinElmer). Green spots are genes which are more highly expressed in the Cy3™ channel, red are more highly expressed in the Cy5™ channel. Yellow indicates equal expression in both channels.

3.3.1b Production of fluorescently labelled target cDNA using the CyScribe™ First Strand cDNA Labelling Kit.

From the result in Table 3.13 it can be seen that approximately double the amount of Cy5™ labelled target would be required to achieve similar levels of fluorescence in both channels which would be expected to result in loss of sensitivity, particularly with highly expressed genes. This is despite the CyScribe™ enzyme which was developed to overcome the problems of incorporating the large bulky groups and to reduce the problem of uneven incorporation due to the larger size of the Cy5™.

	cDNA ng	pmol dye	FOI
CyScribe™ direct kit Cy3™	616.05	27.33	14.41
CyScribe™ direct kit Cy5™	455.10	13.00	9.27

Table 3.13: Amount of cDNA and dye incorporation using CyScribe™ direct labelling kit. 10µg of cytoplasmic RNA was labelled using the CyScribe™ labelling kit. The amount of cDNA produced was calculated from the OD₂₆₀ and the pmol of dye incorporated was calculated from the OD₅₅₀ for Cy3™ and the OD₆₅₀ for Cy5™. The FOI, which is the number of Cy-dUTPs per 1,000 nucleotides of cDNA, was also calculated.

The protocol supplied with this kit suggests between 10 and 15pmol of dye integrated into the cDNA is required for adequate detection on the microarrays. It required 10µg of cytoplasmic RNA to provide enough labelled target for two array slides in the case of Cy3™ and one slide in the case of Cy5™.

3.3.2 Indirect Labelling: Testing parameters of fluorescently labelled target production by aminoallyl labelling.

3.3.2a Comparison of different Reverse Transcriptase Enzymes and their activity at various temperatures.

Reverse transcriptases vary in their ability to incorporate dNTP analogues into cDNA. They also vary in their optimum working temperature which can be important when the RNA template has secondary structure as this can cause the reverse transcriptase to detach from the template leading to premature termination of cDNA synthesis. Genetically modified reverse transcriptases have been produced which work at higher temperatures and are more efficient at incorporating modified dNTPs. SuperScriptII™ reverse transcriptase is a modified MMLV made by the introduction of point mutations in the RNase H activity centre resulting in reduced RNase H activity while retaining full activity at 42°C. These modifications should result in greater first-strand cDNA yields and more full length copies of the message. The Clontech Advantage™ enzyme is also a modified MMLV which is more active at higher temperatures. PowerScript™ is also MMLV with a point mutation in the RNase H activity centre.

Reverse transcriptase from different sources were tested for their ability to incorporate aminoallyl dUTP into cDNA at various temperatures (Table 3.14). The protocol in Table

3.8 with the Microcon™ ultrafiltration method to achieve step 4, the cleaning of the aminoallyl cDNA, was used in this experiment.

	cDNA ng		pmol dye		FOI	
	Cy3™	Cy5™	Cy3™	Cy5™	Cy3™	Cy5™
Superscript II™ 37°C	170.94	248.64	4.40	15.60	8.36	20.37
MMLV 37°C	130.98	68.82	6.40	5.04	15.86	23.78
Advantage™ 37°C	210.90	193.14	8.80	9.12	13.55	15.33
Superscript II™ 42°C	231.12	233.10	7.20	15.12	10.97	21.06
MMLV 42°C	208.68	142.08	7.6	4.80	11.82	10.97
Advantage™ 42°C	117.66	93.24	2.40	9.12	6.62	31.76
Superscript II™ 48°C	108.78	148.74	2.80	3.84	8.36	8.38
MMLV 48°C	126.54	73.26	2.80	2.88	7.18	12.76
Advantage™ 48°C	111.00	66.60	2.80	4.56	8.19	22.23
Powerscript™ 37°C	228.66	173.16	8.40	4.08	11.93	7.65
Mean	162.73	144.08	5.36	7.42	10.28	17.43
Standard Deviation	48.67	67.83	2.58	4.68	3.02	7.22

Table 3.14: Testing of various reverse transcriptases at different temperatures: 10µg of cytoplasmic RNA was labelled using reverse transcriptase at the temperature indicated in the table for 1.5h. The ratio of aa-dUTP to dTTP was 2:1. Aminoallyl modified cDNA was cleaned using a Microcon™30. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA.

There does not appear to be any significant differences in the synthesis of cDNA between the different reverse transcriptases or the different reaction temperatures. There is considerable variation in the amount of cDNA recovered at the end of the procedure which can be due to many factors. Although some variation could be due to pipetting errors, a particular problem when small volumes are being dispensed, reactions used the same sample of total RNA and the same master mix between Cy3™ and Cy5™. A particular problem with this protocol is the use of the Microcon™30 concentration ultrafilters. The recovery from these devices can be highly variable. Variation in the amount of dye incorporated could be due to the sensitivity of the reactions to pH. Any

TRIS buffer left after the Microcon™30 treatment would both disrupt the pH of the coupling buffer and compete for the binding of the mono-reactive dye. Another factor is the degree of hydration of the dye stock which must be stored with extreme care. Furthermore, the DMSO used to dissolve the dye must be completely dry. These factors have complicated the interpretation of the results. However only twice did 10mg of cytoplasmic RNA yield enough labelled target to hybridize to an array.

3.3.2b Ratio of aminoallyl d-UTP to d-TTP.

The ratio of aminoallyl dUTP to unmodified dTTP is also important; too high a concentration can inhibit the reverse transcriptase enzyme. Furthermore, if the modified side groups are incorporated too frequently and are too close together on the cDNA product, fluorescence quenching can occur or steric hindrance can prevent efficient hybridisation of cDNA with the probes on the array slides.

As the published protocol (<http://derisilab.ucsf.edu/pdfs/amino-allyl-protocol.pdf>) was optimised using yeast DNA and DNA from different species vary in their A/T ratios, experiments were performed using the same protocol with various ratios of aminoallyl d-UTP to unmodified d-TTP. Excess aminoallyl dUTP can inhibit the synthesis of cDNA in the reverse transcription reaction. As Cy5™ is the dye which causes most problems due to its larger size and greater hydrophobicity, initial experiments were carried out with Cy5™ only.

The results in Table 3.15 show all ratios of aminoallyl-dUTP to dTTP produced acceptable amounts of cDNA although the highest amount of cDNA was obtained when using a ratio of 2aa-dUTP: 3dTTP. As expected, the highest labelling index was obtained with a ratio of 2aa-dUTP:1dTTP. Only the dNTP mix supplied with the kit failed to yield enough labelled target for an array.

Ratio of aminoallyl dUTP to dTTP	cDNA ng	pmol dye	FOI
aa-dUTP:dTTP 1:3	188.70	10.08	17.34
aa-dUTP:dTTP 1:2	259.74	13.68	17.10
aa-dUTP:dTTP 1:1	217.56	17.04	25.43
aa-dUTP:dTTP 2:3	348.54	18.24	16.99
aa-dUTP:dTTP 2:1	213.12	19.20	29.25
Kit dNTP mix	255.30	8.88	11.29
Mean	247.16	14.52	19.57
Standard Deviation	56.49	4.34	6.55

Table 3.15: Testing the optimal ratio of aminoallyl dUTP to dTTP in labelling. 10 μ g of total RNA was labelled using the above ratios of aa-dUTP to dTTP and SuperScriptIITM reverse transcriptase. A final reaction was labelled using the dNTP mix from the PowerScript kit whose composition is proprietary information. The aminoallyl modified cDNA was cleaned with a MicroconTM30. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA.

Further high ratios were tested by labelling both with Cy3TM and Cy5TM. Ethanol precipitation (Table 3.8, Step4*) was also tested as a means of both cleaning up the aminoallyl cDNA and changing the buffer from TRIS.

Ratio of aminoallyl dUTP to dTTP	cDNA ng		pmol dye		FOI	
	Cy3 TM	Cy5 TM	Cy3 TM	Cy5 TM	Cy3 TM	Cy5 TM
aa-dUTP:dTTP 3:2	97.68	275.28	0.80	12.00	2.66	14.15
aa-dUTP:dTTP 3:1	202.02	270.84	11.60	30.24	18.64	36.25
aa-dUTP:dTTP 4:1	299.70	299.70	20.80	28.56	22.53	30.94
Mean	199.80	281.94	11.07	23.60	14.61	27.11
Standard Deviation	101.03	15.54	10.02	10.08	10.53	11.54

Table 3.16: Testing further ratios of aminoallyl-dUTP to dTTP: 10 μ g of total RNA was labelled using the above ratios of aa-dUTP to dTTP and Superscript II reverse transcriptase. Buffer exchange and aminoallyl cDNA cleanup was achieved by EtOH precipitation. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA.

Table 3.16 shows that even a ratio of 4 aminoallyl dUTP to 1 dTTP gives a good yield of cDNA. The variation in cDNA yield between the Cy3 and Cy5 samples with the 3:2

aa-dUTP:dTTP ratio possibly reflects loss of the pelleted cDNA from ethanol precipitation as the pellets are small, hard to see and easy to dislodge from the wall of the tube. The uneven labelling is less easy to explain. If the higher channel had been the Cy3, then it could have been explained on the grounds of the incorporated aminoallyl groups being too close together and steric hindrance or quenching of the larger Cy5 side chains. The most probable explanation is a difference in the degree of hydration of the dye stocks used in the labelling.

3.3.2c Purification of Aminoallyl modified cDNA.

To try and improve the reliability of recovery of the cDNA, the aminoallyl cDNA was cleaned and the buffers exchanged by either using the QIAquick™ PCR purification kit at Step 4 Table 3.8 or ethanol precipitation. The QIAquick™ protocol used is exactly as in Step 6 Table 3.8. The results are presented in Table 3.17.

Although both ethanol precipitation and QIAquick™ purification appear to be more reliable than the Microcon™30 for cleaning and buffer exchange, major losses can still occur.

	Aminoallyl cDNA cleanup	cDNA ng		pmol dye		FOI	
		Cy3™	Cy5™	Cy3™	Cy5™	Cy3™	Cy5™
SuperScriptII™	QIAquick™ PCR purification kit	124.32	164.28	8.40	6.72	21.94	13.28
PowerScript™	QIAquick™ PCR purification kit	175.38	122.10	6.00	4.80	11.11	12.76
PowerScript™	Ethanol precipitation	64.38	215.34	0.00	18.48	0.00	27.86
Mean		121.36	167.24	4.80	10.00	11.02	17.97
Standard Deviaion		55.56	46.69	4.33	7.41	10.97	8.57

Table 3.17: Testing methods of aminoallyl cDNA cleanup. 10µg of total RNA was labelled using a ratio of aminoallyl dUTP to dTTP of 4:1. The aminoallyl modified cDNA was cleaned of enzymes, unincorporated nucleotides and fragments of the RNA template and the buffer exchanged as indicated. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA.

3.3.3 Kits

3.3.3a Comparison of the Atlas™ glass and Atlas™ PowerScript™ labelling kits.

The performance of the Atlas™ glass and Atlas™ PowerScript™ labelling kits, both produced by Promega, were compared. The PowerScript™ kit has a maximum input of 8µg of RNA while the Atlas™ kit has a maximum input of 10µg of RNA. Labelling was carried out according to the protocols in Table 3.7 and 3.8 and the probes were analysed as before (Table 3.18).

These kits use a resin, followed by ethanol precipitation to remove the enzymes, primers and fragments of mRNA from the aminoallyl modified cDNA. This should

result in better, more consistent recovery of cDNA and indeed the amount of cDNA in the Cy3™ and Cy5™ reactions, with one exception, is more reproducible.

	cDNA ng		%cDNA/total		pmol dye		FOI	
	Cy3™	Cy5™	Cy3™	Cy5™	Cy3™	Cy5™	Cy3™	Cy5™
Atlas kit 37°C	4280.90	4539.90	42.81	45.40	42.00	47.60	3.19	3.40
Atlas kit 42°C	3537.20	358.90	35.37	3.59	38.67	7.60	3.55	6.88
Atlas kit 48°C	4206.90	4221.70	42.07	42.22	29.33	19.20	2.26	1.48
PowerScript kit 37°C	3888.70	3936.80	48.61	49.21	23.33	20.40	1.95	1.68
Mean	3978.42	3264.32	42.22	35.10	33.33	23.70	2.74	3.36
Standard Deviation	339.81	1952.55	5.42	21.20	8.56	16.95	0.76	2.50

Table 3.18: Comparison of Atlas™ Glass and PowerScript™ kit. 10µg of total RNA and 8µg of total RNA was labelled with the Atlas™ kit and PowerScript™ kit respectively. The Atlas kit was tested with the reverse transcription carried out at the temperature indicated. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA

The picomoles of dye incorporated compares well with the best conditions described above. However, the amount of cDNA recovered with these two kits is unreasonably high. The proportion of mRNA in cells is between 1 and 5% of the total RNA. Although ES cells are believed to have high levels of mRNA, even if reverse transcription was 100% efficient, this is an unlikely amount of cDNA. The only reaction which produced the expected amount of cDNA, the Atlas kit at 42°C with Cy5™ conjugated also had much less dye incorporated suggesting the lower amount of cDNA was due to loss of the cDNA pellet. Moreover the reverse transcription reactions between Cy3™ and Cy5™ are identical.

These two kits rely on RNaseH to degrade the RNA template. The RNA fragments are perhaps not efficiently removed by the purification protocol resulting in artificially high values for cDNA recovery. This is further suggested by the amount of cDNA produced

by the both CyScribe™ Direct kit and the by the experimental results reported in Tables 3.13, 3.14 3.15 and 3.16, where NaOH was used to hydrolyse the RNA template.

3.3.3b Comparison of Powerscript Kit with Modified protocol and the Cyscribe™ cDNA Post Labelling kit.

To test whether the RNase H degradation of the mRNA template is responsible for the excess amount of cDNA, the protocol for the Powerscript™ kit was modified. Step 3 of the Atlas™ PowerScript™ Fluorescent labelling kit protocol (Table 3.10) was replaced hydrolysis with NaOH (Step 3*) and compared with the CyScribe™ cDNA Post Labelling kit which uses NaOH degradation of the template as standard protocol. The amount of cDNA synthesised still is more than might be expected (Table 3.19). However, the methods used for the removal of oligo nucleotide fragments are not 100% efficient, some will coprecipitate with the cDNA. The amount of cDNA is similar in both channels as is the amount of dye. This is enough target to hybridise to 2 array slides.

	cDNA ng		%cDNA/total		pmol dye		FOI	
	Cy3™	Cy5™	Cy3™	Cy5™	Cy3™	Cy5™	Cy3™	Cy5™
PowerScript kit 42°C	547.60	573.50	10.95	11.47	29.33	28.80	17.39	16.30
CyScribe Indirect kit 42°C	440.30	458.80	8.81	9.18	29.33	28.40	21.63	20.10
Mean	493.95	516.15	9.88	10.32	29.33	28.60	19.51	18.20
Standard Deviation	75.87	81.10	1.51	1.62	0	0.28	3.00	2.69

Table 3.19: Comparison of the PowerScript™ kit with the CyScribe™ Indirect kit. 5µg of total RNA was labelled with the above kits. Reverse Transcription was carried out at 42°C. In both cases the RNA template was destroyed using NaOH. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA

3.3.3c Comparison of the method of template removal and the effect on Microarrays.

Discussion with colleagues (Ngoc Nga Vinh, Victoria Workman) suggested that the arrays hybridised with targets labelled using the Atlas™ Glass and Powerscript™ kits with RNA template destruction by RnaseH were not as bright as those labelled by the direct labelling method. This might be expected if the unlabelled nucleic acids in the targets competed with the labelled cDNA,

Five micrograms of cytoplasmic RNA from the same IMT-11 sample as for all the experiments described in this chapter was used to prepare targets using the Powerscript kit. The RNA template was destroyed by either RnaseH or NaOH. Reverse transcription was carried out at 42°C for 1h.

The labelled target was analysed as before (results are presented in Table 3.20).

Experiment Number and Dye	Template removed by RNase			Template removed by NaOH		
	cDNA synthesises(ng)	Dye incorporated (pmol)	FOI	cDNA synthesises(ng)	Dye incorporated (pmol)	FOI
1 Cy3™	2002.0	20.0	3.24	469.9	43.3	29.94
2 Cy3™	1357.9	15.3	3.72	758.5	46.3	19.99
1 Cy5™	1491.0	20.0	4.35	384.8	38.8	32.74
2 Cy5™	1311.5	11.2	2.75	455.1	20.0	18.04
Mean	1540.60	16.62	3.52	517.08	37.1	25.18
Standard Deviation	316.87	4.24	0.68	165.18	11.81	7.25

Table 3.20 Labelled cDNA 5µg of cytoplasmic RNA was labelled per reaction. In Experiment 1 enough labelled cDNA was produced by one 5µg reaction for the experiment. In Expt2, two reactions were carried out so the figures represent the mean of the two labelling reactions. The FOI is the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA

The amount of cDNA produced and the level of dye incorporation was consistent with the earlier results. The FOI (the number of Cy-dUTP molecules per 1,000 nucleotides of

cDNA) for each experiment was very similar between the two channels, however was much lower in the RnaseH treated targets. When run out on slide gels the targets prepared by both methods looked very similar (Figure 3.6).

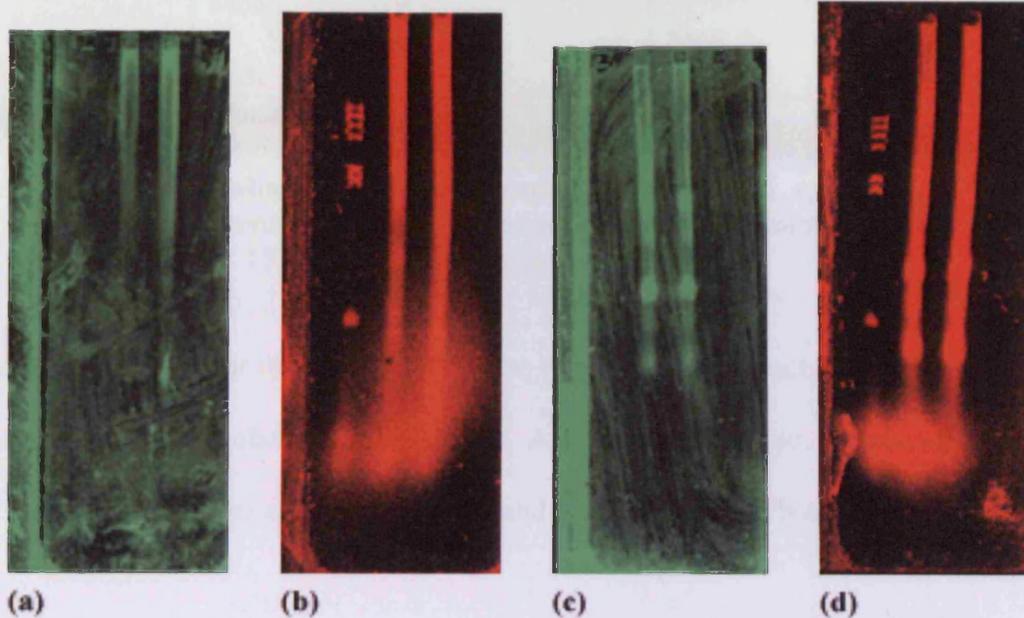


Figure 3.6. Examples of slide gels of Cy3TM and Cy5TM labelled targets.

Gel(a) Cy3TM labelled target, RNA template degraded with RNase H.

Lane 1 λ DNA HindIII fragment ladder (Invitrogen)

Lane 2 Sample Experiment1 (see Table 3.19)

Lane 3 Sample Experiment2 (see Table 3.19)

Lane 4 Empty

Gel(b) Cy5TM labelled target, RNA template degraded with RNase H.

Lane 1 λ DNA HindIII fragment ladder (Invitrogen)

Lane 2 Sample Experiment1 (see Table 3.19)

Lane 3 Sample Experiment2 (see Table 3.19)

Lane 4 Empty

Gel(c) Cy3TM labelled target, RNA template degraded with NaOH

Lane 1 λ DNA HindIII fragment ladder (Invitrogen)

Lane 2 Sample Experiment1 (see Table 3.19)

Lane 3 Sample Experiment2 (see Table 3.19)

Lane 4 Empty

Gel(d) Cy5TM labelled target, RNA template degraded with NaOH.

Lane 1 λ DNA HindIII fragment ladder (Invitrogen)

Lane 2 Sample Experiment1 (see Table 3.19)

Lane 3 Sample Experiment2 (see Table 3.19)

Lane 4 Empty



The total number of good quality spots (flagged 0) were counted on each slide and the mean was calculated (Table 3.21). No significant difference in the number of spots was detected.

	RNase	NaOH
Mean Number of Spots per slide	3055	3168.5
Standard Deviation	1571	963.5

Table 3.21 Mean number of unflagged spots identified from two experiments. Total number of 15247 gene spots on each slide.

The total intensity for the good spots in the two channels on each slide was acquired and the mean was calculated (Table 3.22). Again there was no significant difference between targets produced using RNaseH and NaOH treated cDNA.

	RNase		NaOH	
	Cy3 TM	Cy5 TM	Cy3 TM	Cy5 TM
Mean Total Intensity (n=4)	1,435,606	1,812,615	1,426,739	1,610,123
Standard Deviation	452,036.6	882,522.3	155,978.5	338,497.5

Table 3.22 Total spot intensity of the unflagged spots for the Cy3 and Cy5 channels for two experiments.

The data from the two experiments was combined such that only spots which were unflagged in both experiments were included. For NaOH this was 1746 spots and for RNase H 1219 spots. These spots were further analysed using the linest and Pearson coefficient function in Microsoft ExcelTM 2003 (Table 3.23). Scatter plots were

generated of Cy3 against Cy5 and a trendline produced forcing the intercept to go through the origin. (Fig 3.7).

Treatment of aminoallyl cDNA	Slope	Pearson Coefficient R	F-statistic	Degrees of Freedom
NaOH	0.479±0.004	0.955	14616.17 p<0.0001	1745
RNase	0.538±0.004	0.964	15406.78 p<0.0001	1217

Table 3.23 Statistics from the linest output for the correlation of fluorescence intensity of Cy3 against Cy5 for NaOH treated cDNA and RNase H treated cDNA.

The model produced is highly significant as indicated by both the Pearson Coefficient and the F-statistic. The slope of the line shows that despite the very even labelling of both targets as described by the FOI (the number of Cy-dUTP molecules per 1,000 nucleotides of cDNA), and the very similar total intensities for each channel, the fluorescence intensity of the Cy3 channel is about half that of the Cy5 channel suggesting dye dependant differences in hybridisation.

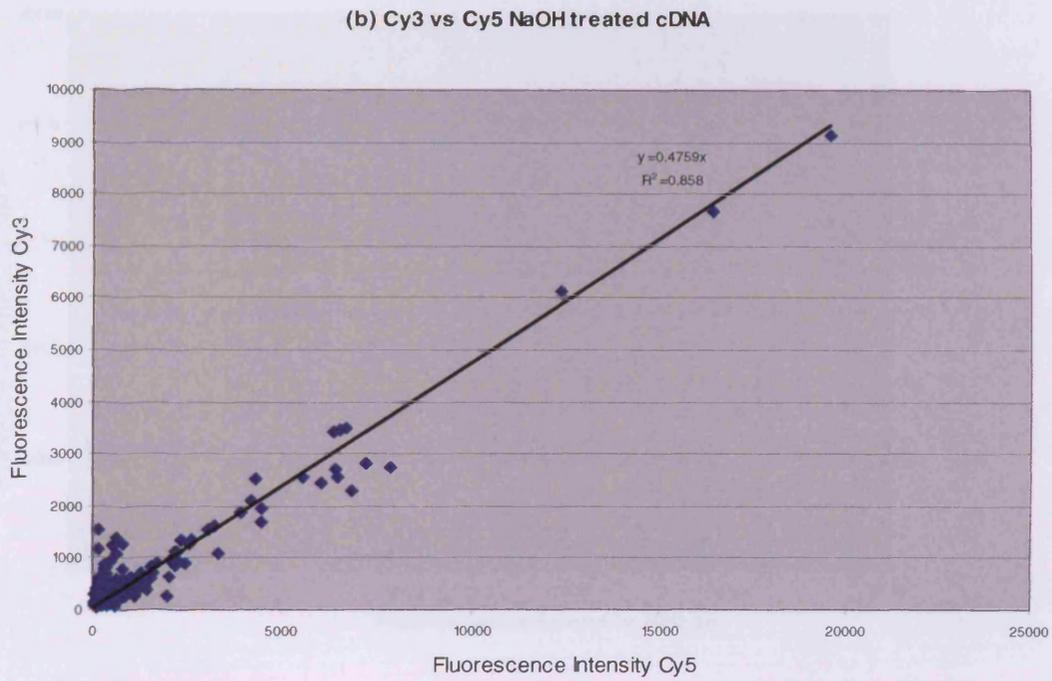
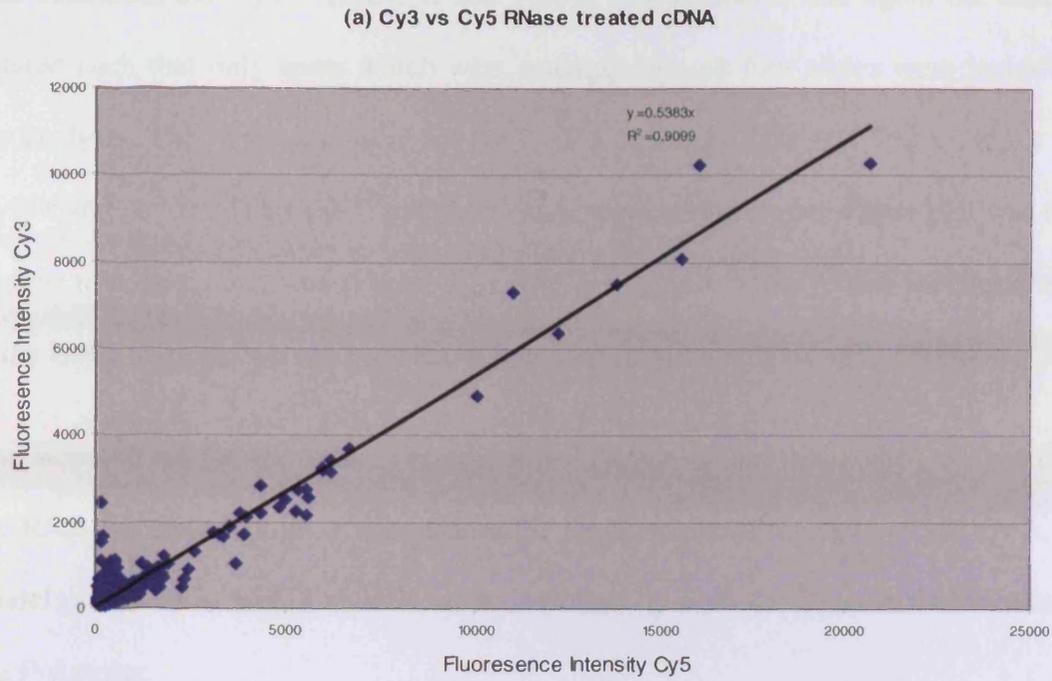


Fig 3.7: Scatter plots of the spot fluorescence intensity of Cy3™ against Cy5™ for (a) RNase and (b) NaOH treated aminoallyl cDNA.

The data from the Cy3™ Rnase H and NaOH was combined and again the data was filtered such that only spots which were unflagged on all four slides were included in the analysis. This was also done for the Cy5™ channels. This left 760 spots for both Cy3™ and Cy5™. The Cy3™ and Cy5™ data was combined, the scatter plot was made and the trendline calculated (Figure 3.8). The significance of the model was again tested using linest and the Pearson coefficient functions in Microsoft Excel™ 2003.

The slope of the the trendline, very close to 1, suggests that the method of destroying the RNA template has no consequences for the fluorence intensity of the spots. The model produced is highly significant as indicated by both the Pearson Coefficient and the F-statistic.

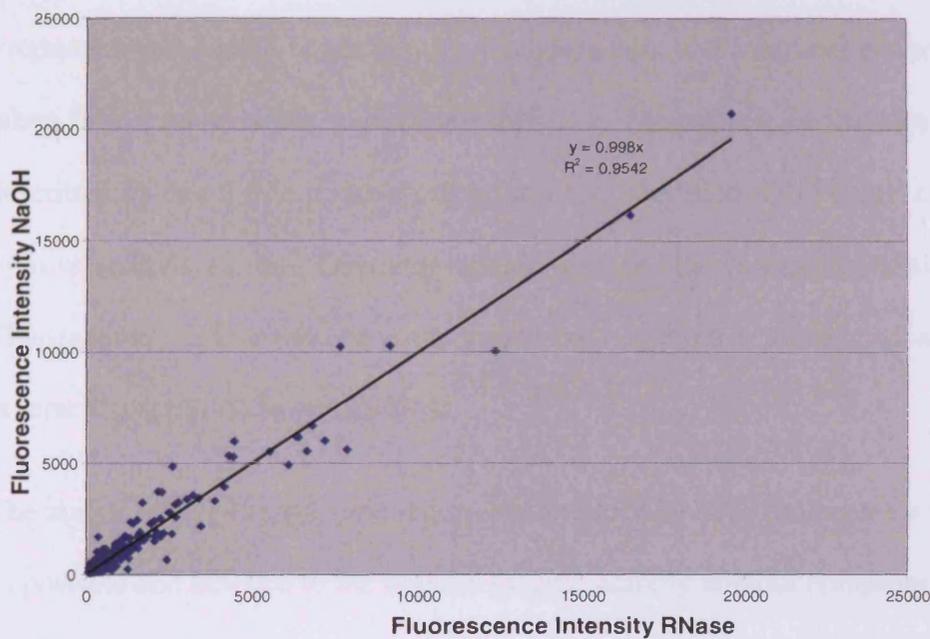


Figure 3.8: Scatter plot of the spot fluorescence intensity for NaOH against RNase H.

However, the NaOH treatment method can produce significantly more labelled target, enough for three arrays whereas the RNase treatment method generally only produces enough for one array (Table 3.16).

Slope	Pearson Coefficient R	F-statistic	Degrees of Freedom
0.998±0.005	0.977	31674.67 p<0.0001	1519

Table 3.23 Statistics from the linest output for the correlation of fluorescence intensity of RNase H treated template and NaOH treated template.

3.4 Conclusions

Production of labelled target is not a trivial process and a number of factors have to be taken into account when performing this task. All but one of the labelling procedures described in this thesis involve incorporation of modified dNTPs into cDNA produced by reverse transcription. The other method involved the chemical modification of RNA. This method, had it worked well, would have offered a number of advantages over reverse transcription based methods.

The aim of labelled target production is to produce labelled molecules which are as long as possible and labelled to the highest specific activity without compromising the ability of the molecules to hybridise to the glass bound probes. As two differently labelled samples are being compared it would be advantageous for the two targets to label to the same extent. Furthermore sequence specific distortions should be avoided.

The direct chemical labelling of the RNA seemed to tick many of these boxes. By avoiding reverse transcription, the length of the target is not affected. As the labelling is

a direct chemical process the two targets being compared should label to the same extent irrespective of the dye molecules being incorporated and sequence specific anomalies should be minimal. In practice, although the labelling could be shown to perform as expected, there were problems. The most difficult was the very short shelf life of the kit in combination with the problems with quality control. The production of the samples is both very expensive and time consuming and the level of unreliability was unacceptable.

Labelling by direct incorporation of labelled dNTP via reverse transcription has the drawback that due to the difference in size and hydrophobicity of the two dye molecules, the specific activity of the two targets differ. In this study three times the amount of Cy5™ labelled target was required to achieve even labelling in both channels. Furthermore sequence dependant differences are more common. For this reason it was decided to concentrate on the indirect labelling method of incorporation of aminoallyl modified dUTP and secondary binding of the fluorescent tag.

Reverse transcriptase has the disadvantage of being a poorly processive enzyme resulting in shortening of the target produced that could lead to non-detection of some probes on the array slides. This is less of a problem with the arrays being used in this study as the majority, but not all, of probes are derived from the 3' end of the genes. However, by careful choice of enzyme and optimisation of the reverse transcription reaction this problem can be reduced.

The length of the cDNA product can be increased by conditions which reduce the secondary structure of the mRNA. This can be achieved by using reverse transcriptases which are active at higher temperatures. This was used as a starting point in optimisation. However, no particular enzyme appeared to function significantly better at

any particular temperature. The actual amount of dye incorporated also seemed low and highly variable. Thus although in many cases, the intensity of labelling as shown by the FOI (frequency of incorporation per 1000 nucleotides) was adequate, there was not enough from one reaction to hybridise on an array slide.

As aminoallyl dUTP can be inhibitory to reverse transcriptase, different ratios of aminoallyl dUTP to dTTP were tested and again no particular condition appeared to perform better than any other. In these experiments, with the exception of the dNTP mix provided with the kit, enough pmoles of dye were incorporated to label an array slide. What was noted was that in some cases, significantly less cDNA was present in Cy3™ labelled product than Cy5™. As both reverse transcription reactions were set up identically using the same master mix, it is likely that the cDNA was being lost in the clean up and buffer exchange steps between incorporation of the aminoallyl dUTP and coupling of the dye. The Microcon-30™ ultrafiltration units can leak leading to cDNA losses. Cleanup of the aminoallyl cDNA was attempted using column based technology, the QIAquick™ PCR purification kit (Qiagen) and ethanol precipitation. In this experiment large cDNA losses were observed in the Cy3™ labelled sample when the cDNA was ethanol precipitated.

The Atlas™ kit was the forerunner of the PowerScript™ kit. They both seemed to perform equally well although loss of the of aminoallyl cDNA can still be a problem. One anomaly noted with these kits was the apparent large amount of cDNA synthesis. Although the amount of dye incorporated into the target is good (enough for between two and three array slides) the FOI is very low suggesting the bulk of the cDNA is unlabelled. However, unlike the non-kit method and the direct incorporation of Cy™ labelled in the CyScribe™ (Amersham Pharmacia) kit, the Atlas™ and Powerscript™

kits (BD Biosciences) use Rnase H to degrade the mRNA template. Consequently, it was decided to replace the Rnase H degradation with NaOH in the PowerScript™ and compare the characteristics of the labelled target with those of target prepared using the CyScribe™ Post-Labeling kit which uses NaOH to degrade the mRNA template as the standard protocol. This demonstrated that there was indeed a difference in the target prepared by the two methods.

The next step was to assess whether the method of mRNA template destruction made a difference to the quality of the hybridisation and to the differential gene expression results. Target was prepared using the PowerScript™ kit and either destroying the template by Rnase H or NaOH. The number of unflagged spots on each slide was taken as a measure of hybridisation in combination with the mean total fluorescence of these spots which showed that there was no apparent difference in the slides hybridised with target produced by either method. Scatter plots of the Cy3™ fluorescence intensity against the Cy5™ fluorescence intensity for each method showed that both methods gave target which behaved in a similar manner; the slope of each plot is similar as is the deviation from the line. When the fluorescence intensity in each channel of each spot detected with both Rnase H and NaOH target was plotted a slope approaching 1 was obtained indicating no difference between the two labelling methods. The disadvantage in using Rnase H to degrade the template is the inability to assess the level of synthesis of cDNA and thereby calculate the FOI of the labelled bases and also less dye incorporated target is produced.

The CyScribe™ Indirect kit and the PowerScript™ kit with the modification of NaOH degradation of the RNA template seem to perform equally with regard to incorporation of the dyes. They both produce between 2 and 4 times the amount of labelled target

from a given amount of cytoplasmic RNA. Notably, labelling in both channels was also more even when compared to the direct incorporation of Cy dye modified dUTP or the method of deRisi or the variants tried. Together with the advantages of kit shelf life, which was very poor in the MICROMAX ASAP kit, and a stable cDNA target, it was obvious that one of these two kits would be the best choice for routinely labelling sample. The PowerScript kit with its modified protocol was chosen because of slightly shorter reverse transcription times and no quenching step.

As the Powerscript™ labelling kit with the modified removal of the RNA template gave such good results, it was decided to use this in all subsequent experiments.

Recently, the method has been further improved by the addition of 0.5µl of PelletPaintNF™ (Novagen, Merck Biosciences Ltd, Nottingham, U.K.) a non-nucleic acid, non-fluorescent, dye-labelled carrier to the ethanol precipitation of the amino-allyl cDNA. This additive colours the pellet blue reducing losses when removing the supernatant. It does not appear to interfere with the analysis of the probe or with the hybridisation to the array.

Chapter 4

Amplification of Target

4. Amplification of Target

4.1 Introduction

We are interested in the pluripotent compartment of the early mouse embryo which contains very few cells. The pluripotent compartment of the pre-implantation blastocyst, the inner cell mass (ICM), at the early blastocyst stage has only about 12 cells rising to about 20 just before implantation (Handyside and Hunter, 1986). The whole blastocyst at about the 64 cell stage contains only 2ng of total RNA (Nagy, 2003) and single cells are generally believed to contain 10-30pg of RNA. Even after implantation on day 5.5, the number of cells in the expanding embryonic ectoderm is only about one hundred. This rises to about 550 on day 6.5 of gestation (Snow, 1976). In the previous chapter, we decided to use the Atlas™ PowerScript™ Fluorescent Labelling Kit which requires 3µg of total RNA. The earliest microarray paper described the use of 5µg of mRNA for target preparation (Schena *et al.*, 1995), the equivalent of 165-500µg of total RNA, assuming 1-3% of total RNA consists of mRNA. It was therefore necessary to use some form of amplification. This can be achieved in two ways; through increasing the fluorescent signal or by amplification of the input RNA.

4.1.1 Signal Amplification.

New labelling methods which can amplify the fluorescent signal post hybridisation have been introduced. The tyramide signal amplification method uses antibodies to introduce horse radish peroxidase (HRP) molecules into the array spots. The highly reactive HRP then oxidizes fluorescent tyramide molecules which couple to the array spots. This procedure can be carried out using a kit provided by Perkin Elmer and requires between 0.5 and 1.0µg of RNA.

The kit produced by Genisphere, which is based on dendrimer technology (Stears *et al.*, 2000) claims that as little as 0.25µg of total RNA is enough to label an array. Both these methods suffer from the drawback that it is difficult to monitor the labelling process as the fluorescent molecules are applied post hybridisation. Furthermore, to acquire this amount of RNA, more than 100 embryos would be required for each sample. We therefore decided that amplification of the RNA sample would be more suitable for our purposes.

4.1.2 Sample Amplification.

Sample amplification has been achieved by conversion of the mRNA transcript to cDNA using reverse transcription followed by either the polymerase chain reaction (RT-PCR) (Saiki *et al.*, 1985) or by *in vitro* transcription (IVT).

4.1.2a Amplification by PCR.

Reverse transcription of RNA followed by PCR (polymerase chain reaction) is a powerful method of amplifying rare transcripts. However, primers are required at both ends of the sequences to be amplified. This requires sequence information for all the probes on the array (Frohman *et al.*, 1988) or a means of introducing a primer to the 5' end of the cDNA. A primer can be added to the 5' end of the cDNA by tailing with an oligonucleotide such as oligo-(dA) using terminal deoxynucleotidyl transferase (TdT) (Belyavsky *et al.*, 1989; Tam *et al.*, 1989). Alternatively, the inherent terminal transferase activity of MMLV reverse transcriptase, which on reaching the 5' cap region of the mRNA will add between 2 and 4 non-template C residues, can be used to add a template switch oligonucleotide (Chenchik, 1998). This oligonucleotide consists of a heel sequence with an oligo-(dG) sequence on the 3' end which base pairs with the d(C)

sequence on the cDNA. The reverse transcriptase then switches template and continues replication to the end of the template thus adding the complementary template switch oligonucleotide to the 3' end of the cDNA (Figure 4.1). PCR can then be carried out using the sequence of the heel as the forward primer (Matz *et al.*, 1999). This method also selects for full length transcript.

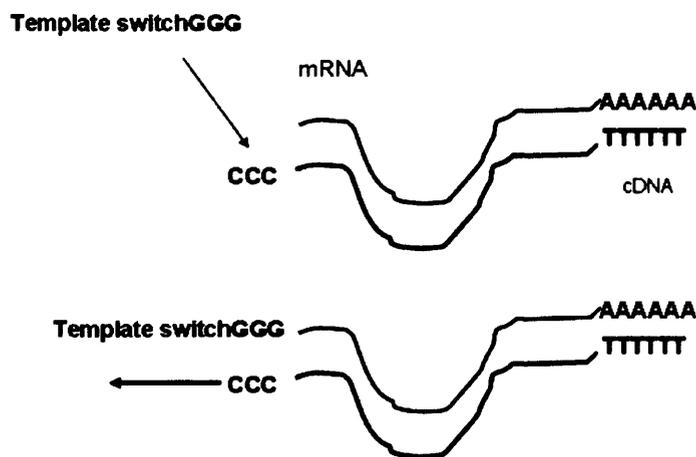


Figure 4.1: Diagram of template switch. When the reverse transcriptase reaches the cap region of the template it adds a string of non-template dCTPs to the cDNA. A template switch oligonucleotide with dGTPs on the 3' end is annealed to the dCTPs and the reverse transcriptase transcribes the template switch.

Although PCR is highly efficient at low RNA input levels, it does suffer from a number of disadvantages resulting in loss of relative transcript levels within samples and of the relative proportions of transcripts between samples. The process is exponential not linear thus high abundance transcripts will soon swamp low. Longer transcripts are less frequently produced than shorter by reverse transcription, and shorter transcripts are

more efficiently amplified than long resulting in loss of maintenance of the relative transcript levels (Belyavsky *et al.*, 1989; Frohman *et al.*, 1988). Furthermore, GC rich sequences are less efficiently amplified by *Taq* DNA polymerase. To some extent the problem of non-proportionality can be reduced by adjusting the amount of input template (Seth *et al.*, 2003) and by careful control of the number of PCR cycles carried out. Over-amplification contributes to serious distortion of the native ratios and it has been suggested that amplification should be halted one cycle before saturation of highly expressed genes (Endege *et al.*, 1999). However the number of cycles suggested seem to vary widely from 65 cycles down to 15 (Iscove *et al.*, 2002; Nagy *et al.*, 2005).

Most recently, a method using a single primer for amplification (SPA) (Smith *et al.*, 2003) has been proposed and on the basis of identification of outliers seems to work well. This has been further modified to give isothermal amplification using a single primer (Dafforn *et al.*, 2004)

4.1.2b Amplification by IVT.

The method of Eberwine (Van Gelder *et al.*, 1990), now used extensively in array analysis, was designed in the pre-microarray era to examine gene expression in single neural cells. The DNA dependant T7 RNA polymerase is highly specific for its promoter (Figure 4.2) (Chamberlin and Ryan, 1982), and is highly processive (Sousa *et al.*, 1992).

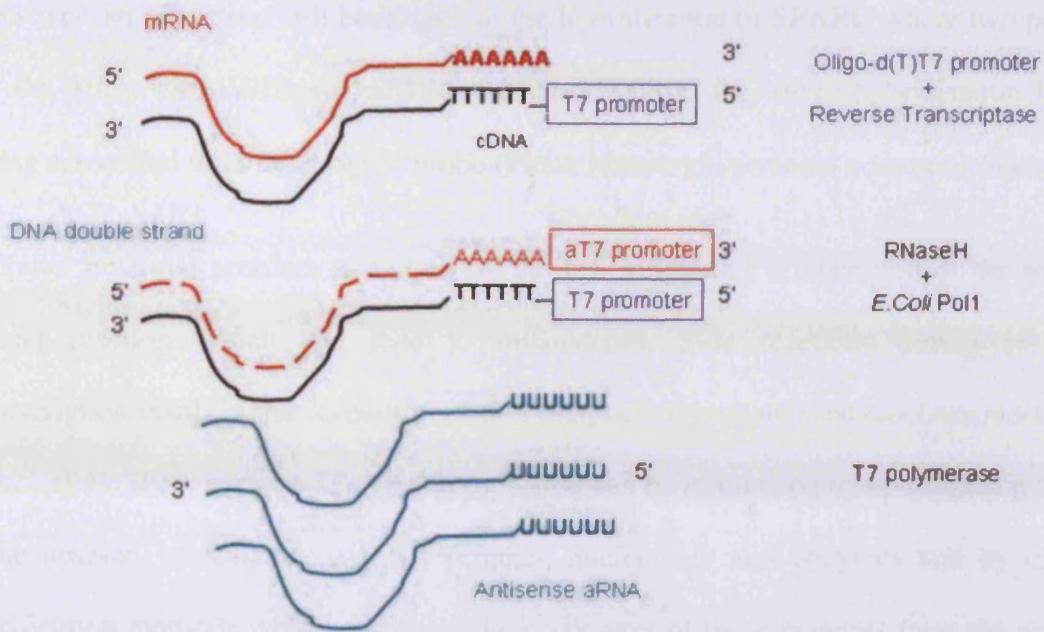
+1

TAATACGACTCACTA TAGGGAGA

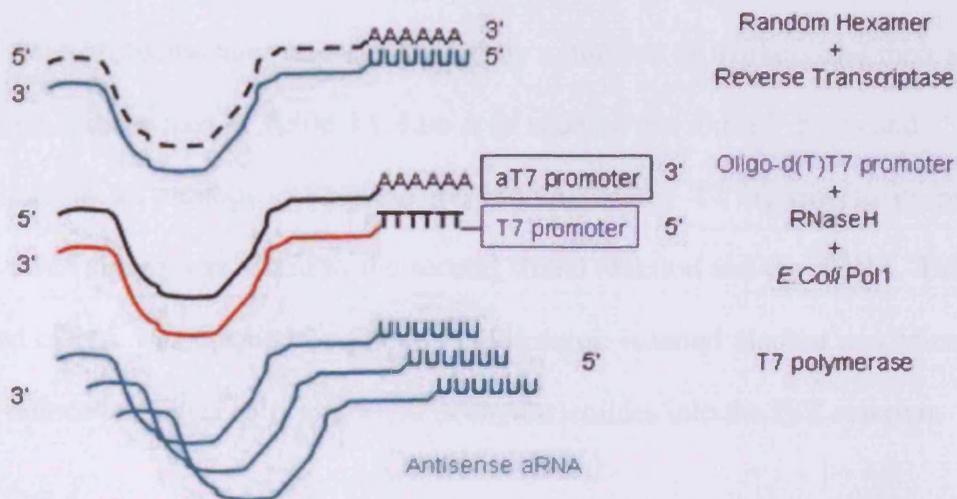
Figure 4.2: Sequence of the T7 promoter. The +1 base, indicated in red, is the first base incorporated into RNA during transcription. The underlined sequence is the minimum required for efficient transcription.

Moreover, the amplification kinetics are linear. In this method the T7 RNA polymerase promoter is attached to cDNA transcribed from total RNA by fusing it to oligo-(dT) and using this as the primer to initiate reverse transcription (Figure 4.3). RNase H is used to degrade the mRNA and the resulting oligonucleotides are used to prime the formation of the second DNA strand. The second strand is polymerised by the enzyme *E.Coli* Polymerase I and the cDNA blunt ended by T4 DNA polymerase. The addition of T7 RNA polymerase and nucleotides with an appropriate buffer allows the creation of a library of mRNAs without inserting the cDNA into plasmids. The enzyme binds to the double stranded promoter, separates the DNA strands and uses the 3'-5' strand as the template for synthesis of antisense RNA. Where a second round of amplification is required, the amplified first round RNA is reverse transcribed using random hexamers as primers. The RNA template is once more degraded with RNase H and a second strand synthesised in the presence of the T7-oligo-(dT) primer which anneals to the polyA tail on the second strand. In this way the T7 promoter is added to the cDNA allowing further synthesis of antisense RNA.

Although the original Eberwine method is used routinely to label RNA for Affymetrix™ hybridisation by introduction of biotinylated NTPs into the first round amplified RNA (also called cRNA or complementary RNA) thus providing limited amplification of the target, it has the disadvantage of loss of 5' complexity introduced by the reverse transcription step. This can be particularly problematical with oligonucleotide arrays. The NIA gene set used in this study is mostly made up of ESTs (expressed sequence tags) derived from the 3' end of the mRNA transcripts so this is less of a problem.



Round 1 of Amplification ready for Labelling or second round of amplification



Round 2 of Amplification ready for labelling

Figure 4.3: Diagram of mRNA Amplification derived from the Eberwine Procedure. The T7 RNA polymerase oligo-dT promoter anneals to the poly-A tail of the mRNA and on reverse transcription produces cDNA with the T7 promoter attached. The mRNA template is destroyed with RNase H and the cDNA double stranded with *E. Coli* Pol I. Multiple copies of antisense RNA is produced by a T7 polymerase reaction. Where more antisense RNA is required a second round of amplification can be carried out by making cDNA with a poly-A tail using random hexamers as the primer. After destruction of the antisense RNA template with RNase H, the oligo-dT-T7 promoter is used to prime the second strand reaction again using *E. Coli* Pol I. The T7 reaction will then produce more antisense RNA.

However, an effect has still been seen in the hybridisation of SPARC where two probes on the array show different degrees of hybridisation, the lower hybridisation levels being associated with the more 5' probe (Fiona Mansergh, personal communication).

Another potential problem is caused by the low stringency temperature of the second strand reaction which can lead to mismatches. Side reactions during *in vitro* transcription result in the formation of non template dependant products (Arnaud-Barbe *et al.*, 1998; Biebricher and Luce, 1996). These can be minimised by optimization of the concentration of reagents such as primers, nucleotides and enzymes and by cDNA purification methods which minimize the carry over of these reagents from the reverse transcription and double stranding reactions.

All of these problems have been addressed by a number of workers and their solutions have been summarized in Table 4.1. Luo *et al* added a few extra 5' bases and 3' d(T)s to the primer in an attempt to stabilise the binding of the T7 enzyme to its promoter. *E.Coli* DNA ligase was added to the second strand reaction and the cDNA. The double stranded cDNA was cleaned by phenol: chloroform: isoamyl alcohol and MicroconTM-100 to reduce carryover of primers and deoxynucleotides into the IVT reaction.

Wang approached the 3' bias problem by selection for full length cDNA using the template switch strategy in the same way as for PCR amplification described above. The second strand extension is carried out at 75°C using *Taq* DNA polymerase which improves stringency of the second strand formation.

	Van Gelder R.N. et al, (1990)	Luo et al, (1999)	Wang et al, (2000)	Baugh et al, (2001)	Scheidl et al, (2002)
*Primers	Oligo ₂₀ -T7-(dT) ₁₅	Oligo ₂₃ -T7-(dT) ₂₁	Oligo ₂₀ -T7-(dT) ₁₅	Oligo ₁₃ -T7-(dT) ₂₄	Oligo ₂₀ -T7-(dT) ₂₄ as Wang
1 st Strand	Reverse Transcription	Reverse Transcription	Reverse Transcription	Reverse Transcription with T4gp32 protein	Reverse Transcription
Second Strand Primer	Fragments from RNaseH treatment	Fragments from RNaseH treatment	Template Switch (in 1 st strand reaction)	Fragments from RNaseH treatment	Template Switch (in 1 st strand reaction) as Wang
Template Removal	<i>E. coli</i> RNaseH	<i>E. coli</i> RNaseH	<i>E. coli</i> RNaseH	<i>E. coli</i> RNaseH	<i>E. coli</i> RNaseH
Second Strand Reaction	DNA Polymerase I T4 DNA Polymerase	DNA Polymerase I <i>E. coli</i> DNA ligase T4 DNA Polymerase	<i>Taq</i> polymerase	DNA Polymerase I <i>E. coli</i> DNA ligase T4 DNA Polymerase	<i>Taq</i> polymerase
Cleaning	Dialysis on a 25μ nitrocellulose filter	Phenol/Chloroform Microcon 100	Phenol/Chloroform Ethanol precipitation Gel filtration P6 column	Phenol/Chloroform Microcon 100	Phenol/Chloroform Microcon 100
Amplification	T7 <i>In Vitro</i> Transcription	T7 <i>In Vitro</i> Transcription	T7 <i>In Vitro</i> Transcription	T7 <i>In Vitro</i> Transcription	T7 <i>In Vitro</i> Transcription DNase treatment of aRNA
Cleaning of amplified RNA		Microcon 100	Trizol	Microcon 100	GeneElute RNA Cleanup kit

Table 4.1. Steps in the process of linear amplification and the methods used by different authors.

* Primer details on Table 4.3.

Baugh *et al* (2001) addressed the problem of 3' bias by addition of the phage protein T4gp32 which improves the processivity of reverse transcriptase.

The published methods were tested for their suitability for the amplification of our experimental material. When commercially produced kits became available, they were also tested for their ability to produce adequate amounts of amplified RNA from our samples.

A particular problem with all the amplification methods based on T7 polymerase is they are all complex and very time consuming. It takes at least three days to complete the two rounds of amplification minimally required by the tiny RNA samples in this project. Moreover, the samples are invisible and it is difficult to assess success during the procedure. For this reason, part of this project was devoted to trying to develop a method which would make use of a solid phase: magnetic beads or latex beads. The simplified and more effective washing procedures would reduce the carryover of primers and nucleotides which contribute to the non-specific products from T7 polymerase side reactions. Not only would washing and exchange buffers be simplified but potentially, a reusable source of target could be produced (Figure 4.4).

Magnetic or latex beads with a spacer, to reduce steric hindrance, and oligo-(dT) conjugated to the surface were used to capture mRNA from the total RNA samples. Buffer exchange and washing steps were achieved by placing the tubes, in the case of magnetic beads, in a magnetic stand which allowed easy removal of the supernatant. Where latex beads were used, the solid and liquid phases were separated by centrifugation. The oligo-(dT) used to capture the mRNA was then used as the primer for reverse transcription to synthesise the cDNA strand. Addition of a template switch oligonucleotide bearing the T7 polymerase promoter, RNase H treatment and second

strand synthesis followed by *in vitro* transcription should have produced amplified RNA in the sense orientation. The amplified RNA produced should have a poly A tail which can further bind to the unsaturated oligo-(dT) on the surface of the beads. The bead saturated with double stranded DNA may be reusable.

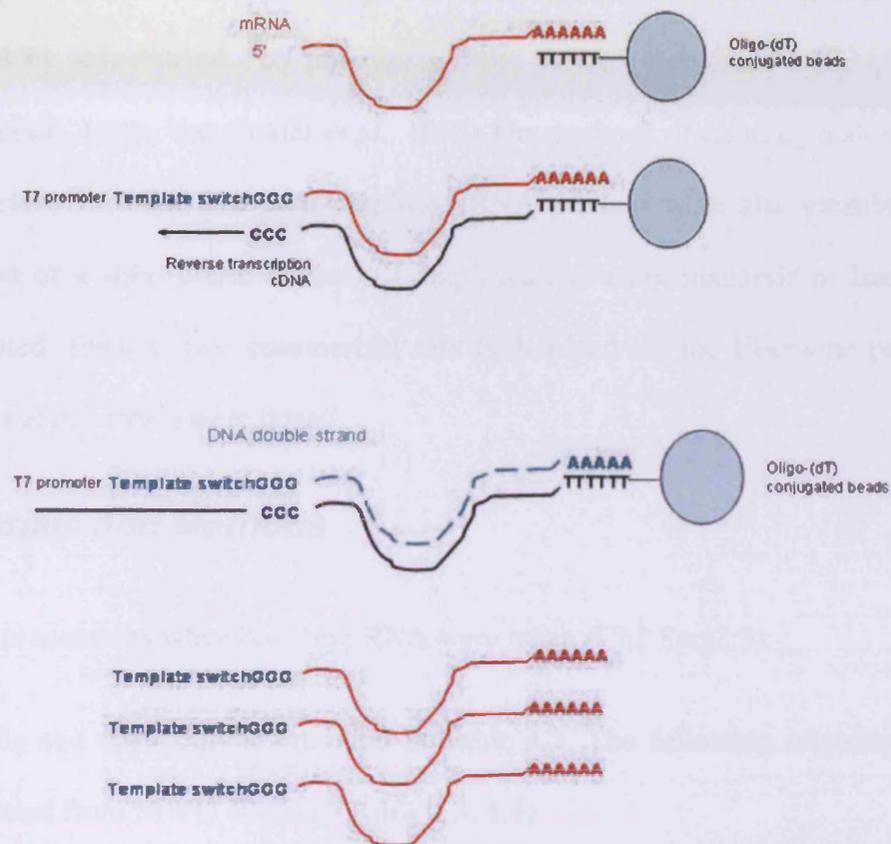


Figure 4.4: Diagram of amplification using beads and template switch. The polyA tail of the mRNA hybridises to the oligo-(dT) conjugated to latex or magnetic beads. Reverse transcription with MMLV reverse transcriptase adds a string of 3-4 Cs to the end of the synthesised cDNA. An oligo with a string of Gs, the template switch, is then used to add a RNA T7 polymerase to the sequence. Double stranding of the cDNA allows the production of amplified sense RNA in an *in vitro* transcription reaction.

In an attempt to reduce steric hindrance of the bead on T7 enzyme, a streptavidin conjugated bead was used to bind a biotinylated oligonucleotide which was used as a spacer arm.

Experiments were performed to try and duplicate the results of Wang *et al* (2000) The efficiency of the reverse transcription reaction was investigated and the addition of the phage T4gp32 protein was used (Baugh *et al.*, 2001) The second strand reaction was investigated by substituting *Taq* polymerase with DNA Polymerase I (Baugh *et al.*, 2001; Luo *et al.*, 1999; Van Gelder *et al.*, 1990). The methods of cleaning both the DNA template before IVT and the final amplified RNA product were also examined. The development of a solid phase method of amplification using magnetic or latex beads was attempted. Finally, two commercial kits both based on the Eberwine procedure (Van Gelder *et al.*, 1990) were tested.

4.2 Materials and Methods

All normal precautions when handling RNA were taken (Ch2 Sect2.3)

A list of kits and chemicals is provided in Table 4.2. The following oligonucleotides were purchased from MWG Biotech (Table 4.3, 4.4).

Where appropriate, the published protocols are reproduced in tabular form for ease of comparison.

7.5M ammonium acetate	Sigma
Advantage cDNA polymerase mix	Promega
Anchored oligo-(dT) ₂₃	Sigma
Dithiothreitol (DTT)	Invitrogen
dNTP 100mM Set	Invitrogen
Dynabeads® Oligo(dT) ₂₅	Dynal
Dynabeads®M270 Streptavidin	Dynal
GenElute™ Mammalian Total RNA Miniprep Kit	Sigma-Aldrich
Linear Acrylamide (0.1 µg/µl)	Ambion
MEGAscript® Kit	Ambion
MessageAmp™ Kit	Ambion
Micro Bio-Spin Chromatograph Columns (Bio-gel P6 in TRIS)	Bio-Rad
Microcon® -100 filters	Amicon
Mini RNA Isolation Kit™	Zymo Research
Oligodex™ beads	Qiagen
Phage T4gp32 protein	USB
Phase Lock® gel tubes	5'-3' Inc
Phenol:Chloroform:Isoamyl alcohol (25:24:1)	Boehringer Mannheim
Poly(dI-dC) Poly-deoxy-inosine-deoxy-cytidylic acid	Roche
QIAquick® PCR purification kit	Qiagen
Random Hexamer	Promega
RiboAmp™ Kit	Arcturus
RNase H (2U/µl)	Invitrogen
RNeasy®	Qiagen
RNAlater®	Qiagen
RNasin (Recombinant)40U/□l	Invitrogen
Superscript II™ reverse transcriptase	Invitrogen
Taq DNA polymerase 5U/µl	Promega
TRIzol®	Invitrogen

Table 4.2: Reagents and kits used

Van Gelder <i>et al</i> (1990) Wang <i>et al</i> (2000)	AAACGACGGCCAGTGAATTG TAATACGACTCACTATAGGGCGT ₁₅
Luo <i>et al</i> (1999)	TCTAGTCGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGT ₂₁
Baugh <i>et al</i> (2001)	GCATTAGCGGCCGCGAAATTAATACGACTCACTATAGGGCGT _{21v}

Table 4.3: Oligonucleotides used to prime the first strand reactions in amplification experiments. All are written in the 5' to 3' direction. Red letters indicates the minimal T7 promoter and the bold indicates the +1 nucleotide i.e. the first transcribed nucleotide.

Template Switch	AAGCAGTGGTATCAACGCAGAGTACGCGGG
20-T7Min	AAACGACGGCCAGTGAATTG TAATACGACTCACTATAGGGCGGGG
9-T7Min	AGTGAATTG TAATACGACTCACTATAGGGCGGGG
T7-Min	TAATACGACTCACTATAGGGCGGGG
TSS	TAATACGACTCACTATAGGGCGCCAAAGAGTAGGG
Biotin Oligo	*GTGGTGGTGGT ₂₅
<i>Pou5f1</i> F	CTTGACCTCAGCCTTAAGAA
<i>Pou5f1</i> R	TGGAACCACATCCTTCTCTAGC

Table 4.4: Oligonucleotides used in this study. All are written in the 5' to 3' direction. Red letters indicate the minimal T7 promoter and the bold indicates the +1 nucleotide i.e. the first transcribed nucleotide. Blue indicates a template switch. * indicates a 5' nucleotide modified by biotinylation.

4.2.1 Amplification using the method of Wang

The protocol is described in Table 4.5 and Table 4.6. This protocol was used as the starting point for testing methods both involving template switch mechanisms and without.

1	Anneal primer 1	Mix in PCR tube: cytoplasmic RNA 0.1µg-0.5µg Oligo ₂₀ -T7-(dT) ₁₅ (Table 4.3). Nuclease free water to 10µl Heated to 70°C for 3min. Cooled to room temperature.
2	Reverse transcribe	The following master mix was assembled on ice and added to the sample 4µl of 5x SuperScriptII™ first strand buffer, 2µl 100mM DTT, 1µl Rnasin (20U/µl) 1µl Template switch oligonucleotide (0.1µg/µl-0.5µg/µl) (Table 4.4) 2µl dNTP nucleotide mix (10mM), 2µl SuperScriptII™ reverse transcriptase (200U/µl) The reverse transcription reaction was incubated for 1.5h at 42°C.
3	Second strand cDNA synthesis	Added to reverse transcription reaction: 106µl H ₂ O 15µl 10x Advantage™ PCR buffer 3µl 10mM dNTP 1µl RNase H 2U/µl 3µl Advantage cDNA polymerase mix. Heated to 37°C for 5min to digest RNA template, 94°C for 2min to denature, 65°C for 1min for specific priming followed by 72°C for 30min to extend the 2 nd strand. Reaction stopped by addition of 7.5µl of 1N NaOH, 2mM EDTA and incubation at 65°C for 10min.

Table 4.5: Protocol for the 1st round of amplification (Wang *et al.*, 2000) continued on next page.

4	Cleanup of double stranded cDNA	<p>Added to 2nd Strand reaction: 1µl of linear acrylamide (0.1µg/µl) 150µl Phenol: Chloroform: Isoamyl alcohol (25:24:1) and mixed well. Transferred to phase lock gel tube and spun at 12,000g for 5min in a microcentrifuge. Transferred aqueous phase to a fresh nuclease free, 1.7ml microcentrifuge tube. Precipitated with 70µl of 7.5M ammonium acetate and 1ml EtOH. Mixed and centrifuged immediately at 12,000g for 20min in a microcentrifuge. Washed pellet twice with 800µl of 100% EtOH, centrifuging at 12,000g for 8min to recover the pellet. Airdried pellet and dissolved in 70µl H₂O. Gel-filtration column (Bio-gel P6 in TRIS) was prepared. The column contents were resuspended by inverting several times. The end was snapped off the column and placed over a collection tube. The cap was removed from the column and the buffer allowed to drain by gravity. The column was re-equilibrated by washing four times with water, each time allowing the column to drain by gravity and discarding the flowthrough. The column was then spun at 200g for 2min at room temperature. The 70µl sample was then loaded on to the centre of the column and spun at 700g for 4min. The cleaned and desalted cDNA was then dried in a rotary evaporator and dissolved in 8µl of H₂O.</p>
---	---------------------------------	---

Table 4.5: Protocol for the 1st round of amplification (Wang *et al.*, 2000) continued on next page.

5	<i>In vitro</i> Transcription	Using the MEGAscript® kit a master mix was assembled at room temperature consisting of 2µl of each 75mM NTP (ATP, GTP, CTP and UTP) 2µl of reaction buffer 2µl of enzyme mix (RNase inhibitor and T7 phage RNA polymerase) This was added to the 8µl sample and was incubated at 37°C for 6h.
6	Purification of amplified RNA	To the amplified RNA sample, 1ml of TRIzol reagent was added and mixed well. 200µl of chloroform was added and the tube mixed by inversion. After standing 2-3min at room temperature the tube was centrifuged at 12,000g for 15min at 4°C. The aqueous phase was transferred to a fresh microcentrifuge tube and the amplified RNA was precipitated by the addition of 500µl of isopropanol. The tube was held at room temperature for 10min and then spun at 12000g for 15min. The amplified RNA pellet was then washed twice in 1ml of 70% EtOH spinning each at 12,000g for 10min to ensure recovery of the pellet. The pellet was airdried and dissolved in H ₂ O.

Table 4.5 Protocol for the 1st round of amplification(Wang *et al.*, 2000)continued.

1	Anneal primers	Combined in PCR tube: 0.5-1 μ g of first round amplified RNA with 2 μ g random hexamer in a total volume 11 μ l of nuclease water. Heat to 70°C for 3min, Cooled to room temperature.
2	Reverse transcription	The following master mix was assembled on ice and added to the sample 4 μ l of 5x SuperScriptII™ first strand buffer, 1 μ l Oligo ₂₀ -T7-(dT) ₁₅ (0.5 μ g/ μ l) 2 μ l 100mM DTT, 1 μ l Rnasin (20U/ μ l) 2 μ l dNTP nucleotide mix (10mM), 2 μ l SuperScriptII™ reverse transcriptase (200U/ μ l) The reverse transcription reaction was incubated for 1.5h at 42°C.
3	Second Strand Synthesis	As for first round amplification
4	Cleanup of double stranded cDNA	As for first round amplification
5	<i>In vitro</i> Transcription	As for first round amplification
6	Purification of amplified RNA	As for first round amplification

Table 4.6 Protocol for the second round of amplification.(Wang *et al.*, 2000)

4.2.2 Amplification using Beads.

The protocols described in Table 4.7 and 4.8 were used to make sense RNA on a magnetic (DynaBeads® Oligo-(dT)₂₅) or polystyrene- latex (Oligotex ®).

4.2.3 Amplification using the MessageAmp™ amplified RNA Kit.

Amplification with the MessageAmp™ amplified RNA kit (Ambion) was carried out according to the manufacturer's protocol (Manual Version 0110).

4.2.4 Amplification using the RiboAmp™ Kit (Arcturus).

RNA was amplified according to the protocols published by the manufacturer, Version A and Version B.

4.2.5 Analysis of Amplification Products Using PCR.

Preparative mixtures were contained in 500µl thin walled PCR tubes. Denaturing, annealing and chain extensions were performed in a Biometra® UNO-Thermoblock™ with a heated lid (Biometra, Thistle Scientific, Glasgow, UK) programmed according to the required conditions. Analysis of the cDNA attached to the beads was carried out by adding the beads to the PCR reaction mix. The reaction mix consisted of 1x Advantage™ PCR buffer, 0.2mM dNTPs, 100pmol of each primer (Table 2.7), 1.0µl Advantage™ polymerase mix and nuclease-free water to make up the final volume of 50µl. The PCR reaction was carried out for two cycles in the presence of the beads.

1	Prepare beads	2x Binding Buffer: 40mM Tris-HCl,pH7.5 2.0M LiCl 4mM EDTA An aliquot of Dynabeads® Oligo(dT) ₂₅ was washed once in 2x Binding Buffer then were resuspended in of 2x Binding Buffer.
2	Anneal RNA to poly-(dT) beads	10ng of cytoplasmic RNA in H ₂ O was heated to 90°C for 1min then quenched on ice. The RNA sample was then mixed with the magnetic beads and allowed to bind for 5min at room temperature.
3	Reverse transcription	The beads were washed twice in 1x binding buffer and once in 1x first strand buffer. Added following reverse transcription mix: 1µl oligonucleotide at appropriate concentration 2µl 0.1M DTT 1µl Rnasin 2µl 10mM dNTPs 4µl 5x First Strand buffer 10µl H ₂ O 2µl Superscript II reverse transcriptase Incubated at 42°C for 90min.
4	Second strand synthesis	Added PCR mix: 106µl H ₂ O 15µl Advantage PCR buffer 3µl 10mM dNTP 1µl RNase H 3µl Advantage <i>Taq</i> polymerase Heated to 37°C for 5min to digest RNA template, 94°C for 2min to denature, 65°C for 1min for specific priming followed by 72°C for 30min to extend the 2 nd strand. Reaction stopped by addition of 7.5µl of 1N NaOH, 2mM EDTA and incubation at 65°C for 10min.
5	Clean up of beads and double stranded DNA	Washed beads twice in binding buffer and once with <i>in vitro</i> transcription buffer.
6	<i>In vitro</i> transcription	Added <i>in vitro</i> transcription mix using MegaScript®Kit: 8µl H ₂ O 8µl dNTP mix 2µl 10X <i>in vitro</i> transcription buffer 2µl T7 enzyme Incubated at 37°C for 14h.
7	Elution	Where appropriate RNA was eluted from the beads with 10mM TRIS pH7.5.

Table 4.7: Protocol for the amplification of RNA using Oligo-(dT) beads(Dynal)

1	Prepare beads	<p>Buffer A: 0.1M NaOH, 50mM NaCl Buffer B: 100mM NaCl Buffer B&W: 5mM Tris pH7.5, 0.5mM EDTA, 1M NaCl</p> <p>25µl of beads at 10mg/ml were washed: 1x for 3min with Buffer A 2x for 3min with Buffer B 1x for 1min with Buffer B&W</p> <p>Suspended beads in 50µl in Buffer B&W Added 50µl of biotinylated oligonucleotide at 5pmol/µl. Incubated at room temperature for 10min. Washed 2x Buffer B&W</p>
2	Anneal RNA to poly-(dT) beads	<p>10ng of total RNA in H₂O was heated to 90°C for 1min then quenched on ice. The RNA sample was then mixed with the magnetic beads and allowed to bind for 5min at room temperature.</p>
3	Reverse transcription	<p>The beads were washed once in 1x first strand buffer. Added following reverse transcription mix: 1µl oligonucleotide at appropriate concentration 2µl 0.1M DTT 1µl Rnasin 2µl 10mM dNTPs 4µl 5x First Strand buffer 10µl H₂O 2µl Superscript II reverse transcriptase Incubated at 42°C for 90min. Removed the reaction from the beads.</p>
4	Second strand synthesis	<p>Added PCR mix: 42µl H₂O 10µl Advantage PCR buffer 1µl 10mM dNTP 0.5µl RNase H 1µl Advantage <i>Taq</i> polymerase</p> <p>Heated to 37°C for 5min to digest RNA template, 94°C for 2min to denature, 65°C for 1min for specific priming followed by 72°C for 30min to extend the 2nd strand. Reaction stopped by addition of 7.5µl of 1N NaOH, 2mM EDTA and incubation at 65°C for 10min.</p>
5	Clean up of beads and double stranded DNA	<p>Washed beads 3x in T7 reaction buffer.</p>

Table 4.8. Amplification of RNA using streptavidin coated beads(Dynal).

For the switch-*Pou5F1* oligonucleotides, the cDNA was denatured at 94°C for four minutes, followed by two cycles of 94°C for 2 minutes, 65°C for 1 minute to anneal the primers to the DNA and 75°C for 1 minute to extend the DNA chain. The synthesis was allowed to extend at 75°C for 5 minutes. The beads were then removed and 30 further cycles followed by a 5 minute extension period were carried out.

Reverse transcription followed by PCR (RT-PCR) was used to analyse amplified RNA. A 1µl sample of amplified RNA was annealed to either 0.5µg of oligo-(dT) or 2µg of random hexamer by heating to 70°C then holding on ice. The primer annealed amplified RNA was input to a reverse transcription reaction containing 1x Reverse transcription buffer, 100mM DTT, 8U RNasin and 1mM dNTPs. The total reaction volume was 20µl and the reaction was carried out at 42°C for 1.5 hours. A tenth of the reverse transcription reaction was subjected to PCR in a mix consisting of 1x PCR buffer without MgCl, 200µM dNTP, 2mM MgCl, and 50pmol primer. For *Pou5F1* the optimum cycling conditions were 94°C for 2minutes, 58°C for 1minute and 75°C for 1minute.

4.3 Results and Discussion

4.3.1 Amplification using Oligo-(dT)-T7.

The method published by (Wang *et al.*, 2000) aimed to reduce the 3' bias of the target and to improve specificity of the second strand. This was done by using a template switch oligo to select for full length mRNA and by using *Taq* polymerase which functions at 72°C to synthesise the second strand of DNA. Initially the applicability was tested with 10ng of cytoplasmic RNA purified from the embryonic stem cell line IMT-

11 diluted from the sample described in Section 3.2. The reverse transcription reaction was primed with 0.1µg of T7-oligo-(dT). Also 0.1µg of the template switch oligonucleotide was used. After the first round of amplification, 0.8µg of amplified RNA was obtained, an estimated 1:2500 fold amplification. However, after the second round of amplification no amplified RNA was detected either by spectrophotometric analysis or by agarose slide gel electrophoresis.

The initiation of reverse transcription is dependant on the concentration of primers in the reaction and on the amount of mRNA. Reactions were carried out with 0.1, 0.2 and 0.5µg of each primer. To confirm that reverse transcription had occurred, an aliquot of the cleaned, double stranded DNA was subjected to PCR using primers for the *Pou5F1* gene (Table 4.4.) Agarose gel electrophoresis of the PCR products showed a band of the expected size (259 base pairs) confirming that reverse transcription had taken place (Figure 4.5). The remainder of the double stranded DNA was *in vitro* transcribed; however no amplified RNA product was detected either by spectrophotometric analysis or by agarose slide gel electrophoresis (2.3.3 and 2.3.5).

The *in vitro* transcription reaction was tested by carrying out the protocol with 0.2µg of the T7-oligo-(dT) and template switch primers. The formation of cDNA was confirmed by PCR as before (not shown) and *in vitro* transcription was carried out. Two control reactions, the first with 0.5µg of pTRI-Xef1α and one which had a mix of 0.25µg of pTRI-Xef1α and double stranded cDNA. The control plasmid has an insert which transcribes the *Xenopus* elongation factor α.

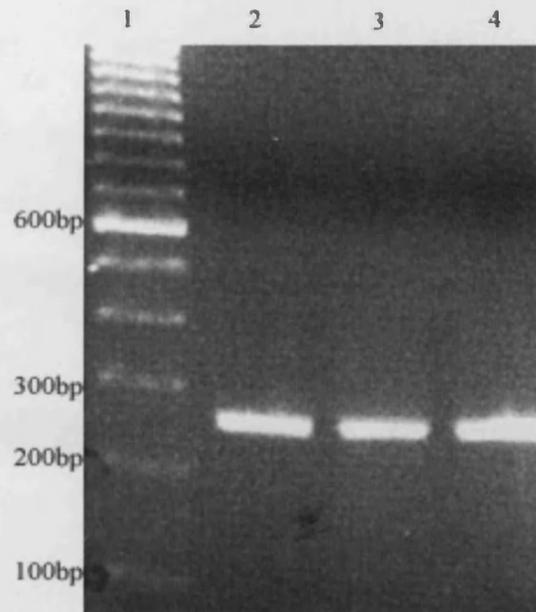


Fig4.5: Agarose gel of PCR product *Pou5F*: PCR of double stranded cDNA from reverse transcription of 10ng of cytoplasmic RNA (round 1 of amplification). The first strand was primed with T7-oligo-(dT) and the second strand initiated from the Template Switch primer.

Lane 1: 100bp ladder

Lane 2: 0.1 μ g of each primer

Lane 3: 0.2 μ g of each primer

Lane 4: 0.5 μ g of each primer

The reactions were carried out at 4°C, 16°C and 37°C for 6h and 26h (Figure 4.6) and the products were run on a TAE agarose gel. The control plasmid seems to amplify well at 37°C and is not prevented from amplifying by addition of the sample. There is a faint band in lane 17 which is the IMT-11 sample amplified for 26h at 16°C. This defined band is not what would be expected from amplified RNA but resembles a template independent product produced by certain batches of T7 RNA polymerase (Biebricher and Luce, 1996).

Time	6h									26h											
Temperature	4°C			16°C			37°C			4°C			16°C			37°C					
Sample	C	E	M	C	E	M	C	E	M	C	E	M	C	E	M	C	E	M			
Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

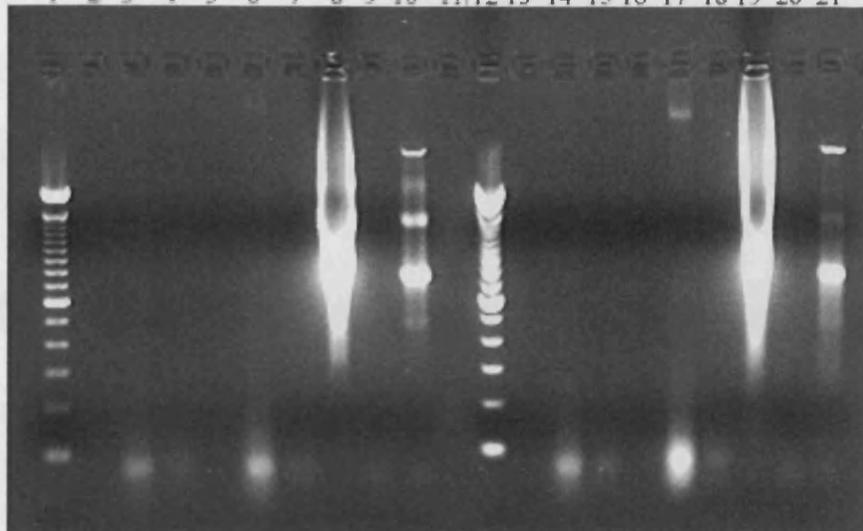


Figure 4.6: Test of *in vitro* transcription reaction: Agarose gel of RNA product. *In vitro* transcription was tested using different conditions. Time indicates the number of hours for which the reaction was carried out and Temperature the reaction temperature. C is the amplification product from 0.5 μ g of the control plasmid pTRI-Xef α . E is the experimental sample of double stranded DNA derived from 10ng of totalRNA. M is a mixture of 0.21 μ g of the control plasmid + one quarter of the cDNA from an experimental sample. Lane1 and 12:100bp ladder. Lanes 2-10 6h amplification, lanes 13-21 26h amplification. Lanes 8, 10, 19 and 21 shows that the *in vitro* transcription reaction only occurred at 37°C and is probably complete after 6h. Although no product is detected from the experimental sample (Lanes 9 and 20) it does not appear to be due to an inhibitory contaminant as the mixture of control and experimental sample has product (lanes 10 and 21). Lane 11 is empty.

Although T7 polymerase will use single or double stranded template, the promoter must be double stranded. Thus, the second strand reaction must copy the whole mRNA molecule for *in vitro* transcription to take place. *E.Coli* DNA polymerase 1 (Pol1) is believed to be a more processive enzyme than *Taq* polymerase and was used by both (Baugh *et al.*, 2001; Luo *et al.*, 1999). The amount of cytoplasmic RNA used was also tested. Reactions were carried out with 10ng, 100ng and 900ng of total RNA (Figure

4.7). The amount of T7-oligo-(dT) primer was fixed at 0.1 μ g while the amount of template switch primer was varied at 0.1 μ g, 0.2 μ g and 0.5 μ g. One quarter of the resulting amplified RNA was run out on a 0.8% agarose gel. Only 900ng of input RNA appears to give any detectable amplified RNA product suggesting amplification from ES cells does occur but requires more starting material than for the material used by Wang *et al* (2000).

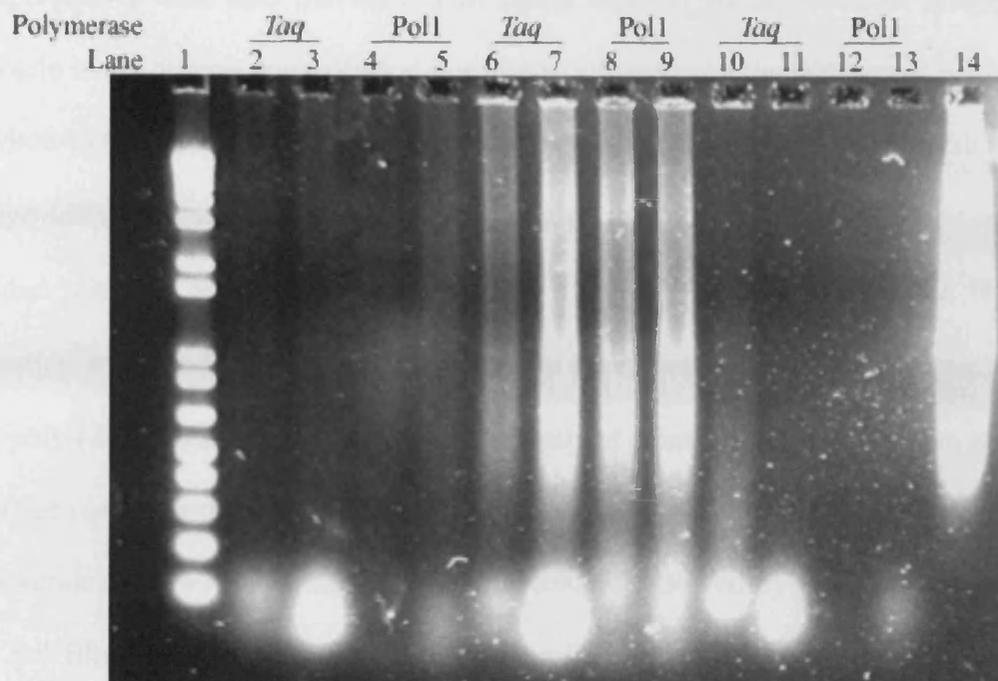


Figure 4.7: Agarose gel of amplification products. 10, 100, and 900ng of cytoplasmic RNA was put into an amplification reaction with 0.1 μ g of T7-oligo-d(T) primer and either 0.1 μ g or 0.5 μ g of template switch primer. The double stranding reaction was carried out with either *E. Coli* DNA Polymerase I or with Taq polymerase for each concentration of primer.

Lane 1: 1kb+ Ladder (Invitrogen)

Lane 2: 100ng total RNA, 0.1 μ g template switch primer, Taq polymerase

Lane 3: 100ng total RNA, 0.5 μ g template switch primer, *E. Coli* DNA polymerase I

Lane 4: 100ng total RNA, 0.1 μ g template switch primer, Taq polymerase

Lane 5: 100ng total RNA, 0.5 μ g template switch primer, *E. Coli* DNA polymerase I

Lane 6: 900ng total RNA, 0.1 μ g template switch primer, Taq polymerase

Lane 7: 900ng total RNA, 0.5 μ g template switch primer, *E. Coli* DNA polymerase I

Lane 8: 900ng total RNA, 0.1 μ g template switch primer, Taq polymerase

Lane 9: 900ng total RNA, 0.5 μ g template switch primer, *E. Coli* DNA polymerase I

Lane 10: 10ng total RNA, 0.1 μ g template switch primer, Taq polymerase

Lane 11: 10ng total RNA, 0.5 μ g template switch primer, *E. Coli* DNA polymerase I

Lane 12: 10ng total RNA, 0.1 μ g template switch primer, Taq polymerase

Lane 13: 10ng total RNA, 0.5 μ g template switch primer, *E. Coli* DNA polymerase I

Lane 14: control plasmid pTRI-Xef α .

Experiments were also carried out to assess whether the addition of phage T4gp32 protein to the reverse transcription reaction would improve the addition of a T7 template switch to the cDNA. For these experiments the first strand was primed with anchored oligo-(dT)₂₃ to ensure binding at the 3' end of the message and the T7 promoter was added with the template switch using the primer TSS (Table 4.4). The first strand reaction contained 6µg of T4gp32 protein in the experimental tube. A carrier of 200ng of poly-(dIdC) was also added to the first strand reaction. Cleaning of the cDNA was carried out using the QIAquick® PCR purification kit. This not only is easier, less time consuming and less toxic than phenol:chloroform followed by Microcon®100 filtration or gel filtration, but also oligonucleotides of less than 40 bases and double stranded DNA of less than 100bp are removed. The T7 polymerase step was carried out for 6 hours at 37°C and the amplified RNA generated was cleaned using the QIAGEN RNeasy® mini kit.

T4gp32 protein does not appear to make a difference to the amount of amplified RNA synthesised (Table 4.9). The control with no RNA template demonstrates the amount of non specific product which is produced in the absence of template. In this experimental design amplified RNA can only be made if the reverse transcription reaches the cap region of the mRNA so more amplified RNA could only be synthesised if more full length cDNA was synthesised.

Amount of total RNA template(ng)	T4gp32 protein	Yield of amplified RNA after the first Round
0	-	1015.2ng
10	+	753.6ng
10	-	799.2ng

Table 4.9: Yield of amplified RNA estimated by spectrophotometry after a single round of amplification.

It has been further suggested that when using the protocol of Wang *et al.* (2000) most of the second strand priming arises from the oligonucleotides produced by RNase H. An experiment was set up to compare the effect of the template switch oligonucleotide by comparing a protocol based on the method describe by (Luo *et al.*, 1999) which has no template switch and that of (Wang *et al.*, 2000).

Reactions were set up with no template, 10ng of IMT-11 RNA and 100ng of an artificial polyA RNA which was 891bases in length as a control. Both the sets of reactions were primed by the T7 oligo described by both Eberwine and Wang (Table 4.3). In the Luo reactions the second strand reaction was carried out using T4 DNA polymerase (Table 4.5 Step*3) and double stranded cDNA was cleaned using the QIAquick PCR cleanup kit. The amplified RNA was purified using the RNeasy Mini kit cleanup protocol. For the template switch protocol the same samples were set up and the first strand set up as for the non template switch. The second strand was carried out using *Taq* DNA polymerase. The cDNA was cleaned using both phenol:chloroform:isoamyl alcohol and the QIAquick PCR cleanup kit with the addition of 1µl of linear acrylamide as a carrier.

Amount of template	Luo Round1	Luo Round2	Wang Round1	Wang Round2
0	520.8ng	388.8ng	256.8ng	139.2ng
10ng	304.8ng	556.8ng	1.17 μ g	362.4ng
100ng control	708.0ng	13.73 μ g	468ng	156.0ng

Table 4.10: Yield of amplified RNA after 1 and 2 rounds of amplification for protocols without a template switch (Luo *et al.*, 1999) and with a template switch (Wang *et al.*, 2000). The control is an artificial polyA tail containing RNA 891 bases in length

Although the Luo method seems to amplify, the level of amplification is not enough and a third round of amplification would be required. The Wang protocol appears to have less after 2nd round. This is explainable if the oligonucleotide products of the RNase H reaction are not being used to form the second strand. Additionally the second round first strand synthesis is initiated by random hexamer thus it is less likely that template switch will be added

4.3.2 Dynabeads

The design of the experiments using oligo-(dT)₂₅ beads requires the T7 promoter to be attached to the free end of the cDNA. This was achieved by adding the T7 promoter with a template switch. The first oligonucleotide tried, T7 Min-S (Table 4.3) consisted of the minimum T7 promoter with the three dG residues on the 3' end which bind to the non-template dC residues on the cDNA. Spectrophotometric analysis suggested amplified RNA was present; however nothing was visible on agarose gel electrophoresis.

In an attempt to stabilize the T7 promoter, oligonucleotides with 9 or 20 residues, 9-T7Min-S and 20-T7Min-S at the 5' end of the switch oligo were compared to the T7Min-S oligonucleotide. The sequence of the 9 and 20 residues added to the T7Min-S

oligonucleotide was random and checked by BLAST to ensure there was no homology to known genes. All three oligonucleotides were set up at both 0.1 μ g and 0.5 μ g per reaction. To check if full length cDNA was present on the bead, the beads with T7Min-S and 9-T7Min-S (the beads with 20-T7Min-S were lost during the procedure) were put into a PCR reaction using the switch oligonucleotide as the upstream primer and *Pou5F1* as the downstream primer. The predicted band size on agarose gel electrophoresis is 853bp. A band of approximately this size was detected on the gel (Figure 4.8).

Because it proved impossible to get reliable measurements with spectrophotometry due to residue from the beads, the presence of amplified RNA was checked by PCR using *Pou5F1*. As the input RNA is destroyed by RNase H any band generated on the gel must come from either the cDNA or amplified RNA. Reverse transcription was carried out with two control reactions, one with no reverse transcriptase which controls for reverse transcriptase activity in the *Taq* polymerase. The second control has enzyme but no oligonucleotide primer in the reverse transcription reaction. The absence of a band in both of these reactions after PCR is indicative of no cDNA contamination in the amplified RNA sample. The band expected would be 261bp in size and a single band of about this size was detected on agarose gel electrophoresis of the RT-PCR products (Figure 4.9). However no RNA was detectable by gel electrophoresis.

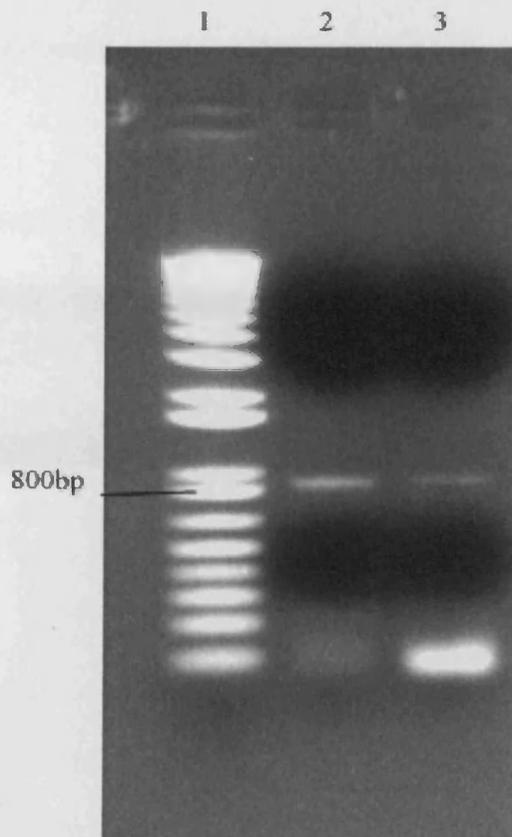


Figure 4.8: Test of attachment of T7 template switch to 3' end of cDNA. Agarose gel of PCR products. The T7Min-S and the 9-T7-Min-S template switch oligonucleotides were used in the formation of a first strand primed from an oligo-(dT) bead. The presence of full length cDNA was tested by PCR using the template switch as the forward, and *Pou5F1* as the reverse primers. A band of 853bp was expected.

Lane 1: 100bp ladder

Lane 2: T7-Min-S to *Pou5F1*.

Lane 3: 9-T7-Min-S to *Pou5F1*.

A new oligonucleotide T7SS (Table 4.4) was designed; the random sequence between the T7 promoter and the switch sequence should allow easier diagnostics as it will be transcribed by the T7 polymerase enzyme. Experiments were performed to check if the newly synthesised amplified RNA, which should have a polyA tail, is binding directly to the beads on synthesis or staying in the liquid phase. To ensure the removal of all traces of LiCl and primers before the *in vitro* transcription, extra wash steps were inserted with 10mM Tris, pH7.5, 150mM NaCl, 1mM EDTA. To try to ensure no beads

were left to interfere with spectrophotometric analysis, the beads were not only processed on the magnetic stand but were also spun in a microcentrifuge at 12000g for 10 minutes.

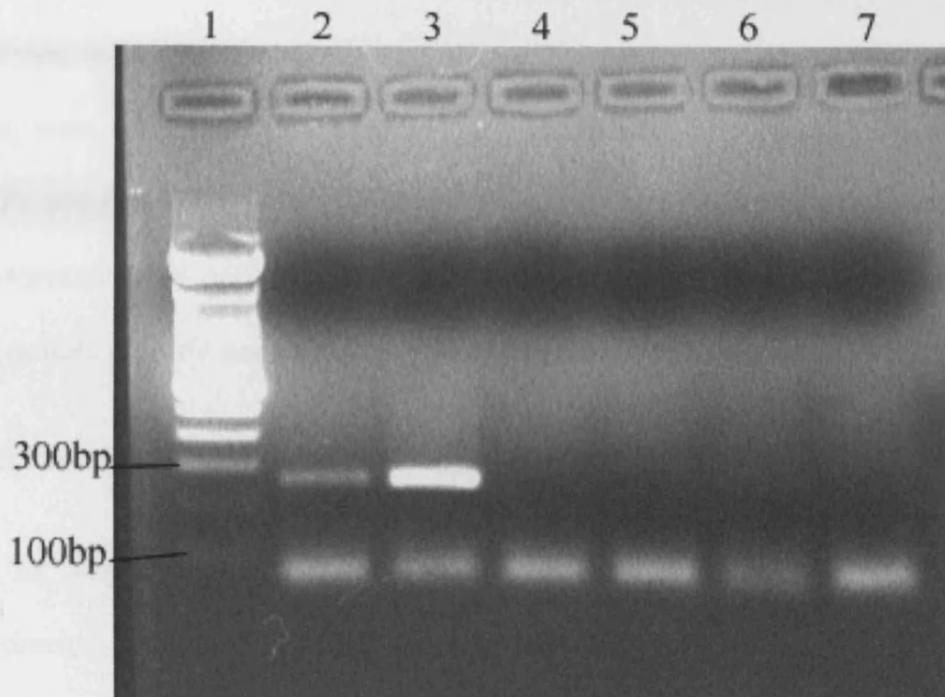


Figure 4.9: Test for presence of amplified RNA product after IVT of bead bound cDNA. Agarose gel of PCR products of amplified RNA for the *Pou5F1* gene.

Lane 1: 100bp ladder

Lane 2: Reverse transcription primed with oligo-dT

Lane 3: Reverse transcription primed with dN6,

Lane 4 and 5: as 2 and 3 without reverse transcriptase.

Lane 6: no primer in reverse transcription.

Lane 7: no template in PCR

All supernatants and eluates from the beads were checked for the presence of amplified RNA by spectrophotometry and although large amounts of amplified RNA were apparently in the liquid phase, none was detected by gel electrophoresis. The polystyrene-latex oligo-dT conjugated beads, Oligotex®, which are a component of the

QIAGEN mRNA isolation kits, were tested to see if they were more suitable than the Dynal magnetic beads. The results with these beads was identical to the Dynabeads® Oligo (dT)₂₅.

Finally, as it was not clear if there was inhibition of full length amplified RNA due to hindrance caused by the spacer and oligo-(dT), streptavidin coated beads were tried. These were conjugated to a biotinylated oligonucleotide consisting of a sequence of dNTPs and oligo-(dT)₂₅ (BiotinOligo Table4.3). The mixed oligonucleotide sequence was checked on BLAST for sequence homologies and was biotinylated at the 5' end. On PCR neither *Pou5F1* nor Switch to *Pou5F1* sequence was detected.

4.3.3 MessageAmp™ (Ambion) amplified RNA Kit.

This kit was used to amplify the same sample of cytoplasmic RNA as the other experiments. After one round of amplification the yields shown in Table 4.11 were obtained.

Sample No	Amount of Total RNA input to first round	Amount of amplified RNA input to second round	Yield of amplified RNA after two rounds of amplification
1	10ng	1.5µg	3.3µg
2	10ng	1.1µg	1.1µg
3	100ng	1.7µg	2.9µg
4	100ng	2.4µg	7.4µg

Table 4.11: Yield of amplified RNA from MessageAmp™ Kit

The whole of the amplified RNA from the first round was input into the second round and the amplified RNA was again analysed by spectrophotometry. The amplified RNA

was precipitated and 1 μ g from sample 1 and 3 was run on an agarose slide gel (Figure 4.10). Sample 1, 10ng of input total RNA, appears to have the expected amount of amplified RNA. Nothing was detected from the 100ng input sample. This may be due to loss of the sample from the slide gel, a problem often seen. Labelling the remainder of this amplified RNA was attempted without success.

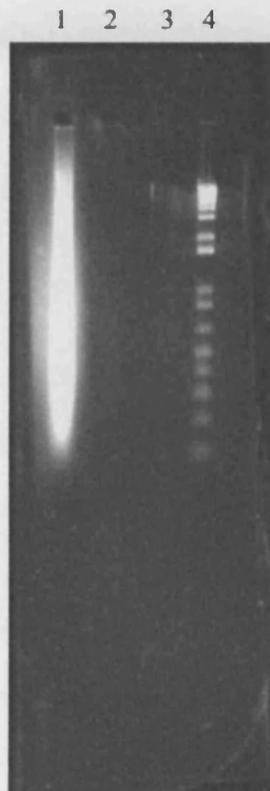


Figure 4.10: Agarose slide gel of 1 μ g of amplified RNA: RNA was amplified from 10ng or 100ng of cytoplasmic RNA from the ES cell line IMT-11 using the MessageAmp™ Kit to carry out 2 rounds of amplification.

- Lane 1: 10ng of RNA in amplification
- Lane 2: 100ng of RNA in amplification
- Lane 3: blank
- Lane 4: 1kb+ ladder

4.3.4 RiboAmp™ (Arcturus) Kit

The RiboAmp™ kit (Arcturus) was used to amplify 10ng and 100ng of the control RNA supplied by the kit, 10ng of IMT-11 cytoplasmic RNA and a no template control using the manufacturer's protocol described in Version A. The major difference between this kit and the other protocols and kits tried is that the second strand is primed by exogenously added primer rather than by the products from the RNase H. The amplified RNA is eluted in a volume of 60µl after the first round of amplification. The volume of the amplified RNA was reduced to 20µl by vacuum centrifugation and a 1µl sample was diluted 1:50 for spectrophotometric analysis.

Slide gel electrophoresis was used to visualise the amplified RNA. Only 1µl of each sample was run on the gel (Figure 4.11). Although the slide gels are very thin (approximately 1mm), the amount of amplified RNA loaded was below the detection limit of Sybergreen Gold in the case of the samples derived from 10ng of total RNA.

The lane with 100ng of total RNA is very clear. Since these gels are non-denaturing, it is not possible to estimate the size spread of the amplified RNA. The "no template" control lane shows no evidence of side reactions.

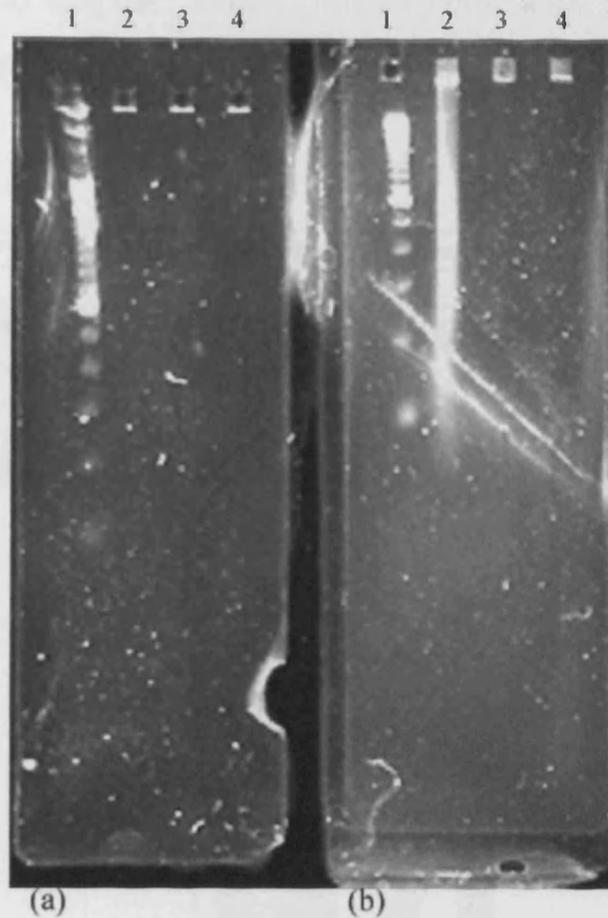


Figure 4.11: Slide gels of amplified RNA from the first round of amplification using the RiboAmp™ kit: 10ng of IMT-11 cytoplasmic RNA and 10ng and 100ng of the control RNA supplied with the kit were subjected to one round of amplification and 1/20th of the product was run on a gel

Gel (a) Lane 1: 100bp ladder
 Lane 2: 10ng IMT-11
 Lane 3: 10ng control
 Lane 4: no template control

Gel (b) Lane 1: 100bp ladder
 Lane 2: 100ng control.
 Lanes 3 and 4: empty

All of the amplified RNA from the 10ng samples and the no template control were put into a second round of amplification and the amplified RNA was again analysed both by spectrophotometry (Table 4.12) and by slide gel electrophoresis (Figure 4.12). In both cases a good amount of amplified RNA was synthesised and no side reactions were

detected in the “no template” control. This amplified RNA labelled well with the Clontech amino-allyl labelling kit.

Template	Yield after Round 1	Yield after Round 2
0	0	0
10ng IMT-11 total RNA	200ng	42.6 μ g
10ng kit control	600ng	37.08 μ g
100ng kit control	52.8 μ g	Not done

Table 4.12: Yield of amplified RNA after amplification using the RiboAmp™ kit.

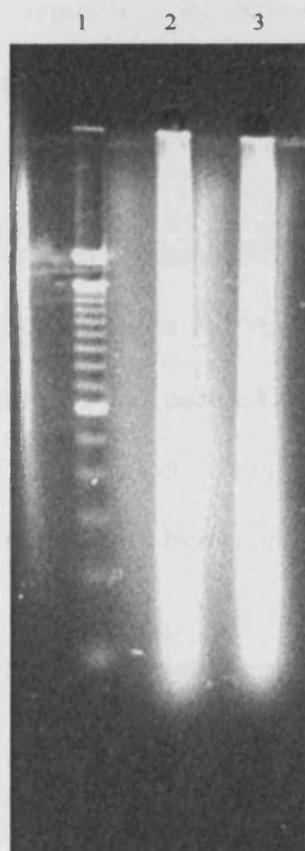


Figure 4.12: Slide gel of amplified RNA: 10ng of input RNA and 2 rounds of amplification using the RiboAmp™ kit.

Lane 1: 100bp ladder,
 Lane 2: 10ng IMT-11 input,
 Lane 3: 10ng kit control.

To confirm that the RiboAmp™ kit (Arcturus) could not only amplify RNA diluted from bulk preparations, but could indeed amplify the type of material we were interested in, the RNA from 19 ICMs was purified using TRIzol® with the addition of 0.1µg of linear acrylamide. After the first round, the yield of amplified RNA was 220ng and after the second round 7.5µg; enough for 2 labelling reactions.

Although TRIzol® purification of total RNA seemed compatible with the amplification procedure; the Mini RNA Isolation Kit™ produced by Zymo Research has several advantages. Not only is the procedure fast, simple and elutes the RNA in a very small volume which can be directly input into the RiboAmp™ amplification procedure without the need for precipitation, but there is no risk of residual phenol inhibiting downstream enzyme reactions.

A sample of 5 ES colonies picked from a plate of IMT-11p15 and a sample of 19 ICM's were purified using the Mini RNA Isolation Kit™. The remainder of the plate of ES cells was harvested and the totalRNA was purified using the GenElute™ Mammalian Total RNA Miniprep Kit and 10ng was used as template in the RiboAmp® procedure. The yield of amplified RNA obtained after two rounds of amplification is presented in Table 4.13

Sample	Yield
19 ICM	5.22µg
5 ES colonies	1.8µg
10ng total ES RNA	89.1µg

Table 4.13: Yield of amplified RNA after two rounds of amplification using the RiboAmp™ kit.

A 0.5 μ g aliquot of each sample was run on a slide gel Figure 4.13. The remainder of the ICM and ES colony amplified RNA was labelled with Cy3™ and Cy5™ respectively using the PowerScript™ labelling kit (Promega)

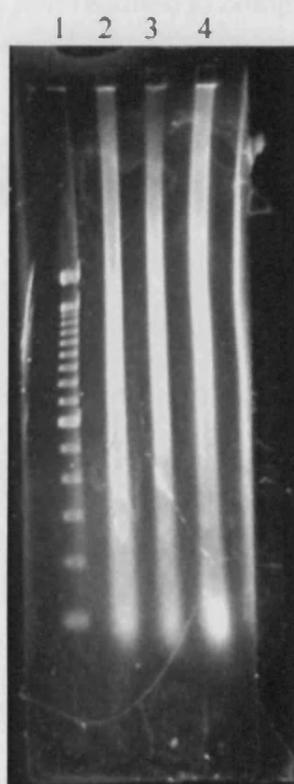


Figure 4.13: aRNA from two rounds of amplification. A sample of 19 ICM's and 5 ES cell colonies were purified using the Mini RNA Isolation Kit™ (Zymo). The remainder of the plate from which the ES colonies were picked was harvested and the RNA isolated using a GeneElute™ Mammalian Total RNA Miniprep Kit. All the RNA from ICM's and ES colonies, and 10ng of RNA from the plate (bulk) was used as template for 2 rounds of amplification using the RiboAmp™ kit. Each lane was loaded with 0.5 μ g of amplified product.

Lane1: 100bp ladder

Lane2 ICM

Lane3 ES cell colonies

Lane4 diluted RNA from bulk culture

The RiboAmp manufacturer's protocol version B significantly reduced the incubation times for the reverse transcription and second strand reactions without affecting the yield of RNA or the quality. Also adjustment of elution volumes from the purification steps removed the need for concentration of the samples by vacuum centrifugation. However, two and a half days are still required to complete the procedure.

4.4 Conclusions

The purpose of the work in this chapter was to develop reliable methods of linear amplification for the very small total RNA samples available in this study. This chapter demonstrates the difficulties which can be encountered when optimising such complicated, multistep procedures at the limit of the sensitivity of the methods.

The published method of Wang *et al* (2000) was tried using 10ng of total RNA which is about the amount which we might expect to have available. Modifications to the basic method as suggested by other authors were also tried. Analysis of the products of each step in the process suggested that the individual enzymatic reactions were working however, the complete procedure failed to produce amplified RNA from the experimental material. As significant amounts of product were produced by the first round when 900ng of total RNA were processed and the presence of amplified RNA was detected by RT-PCR when 10ng were used, it could be concluded that for some reason the second round of amplification was not functioning. The amount of product after the second round of amplification was less than after the first. This suggested that either only non-specific products were being made as in the no template control or the priming of the second round was not efficiently adding the polyA tail and annealing the

oligo₂₀-T7-(dT)₁₅ primer. The ability to amplify the *Pou5f1* gene from the first round product argues against there being no specific product but if the template switch is added infrequently, only some of the cDNA would be expected to reach the RNA cap region, and there is little or no priming from the RNase H fragments then the amount of template which can be transcribed may be very small and the bulk of the product seen would be the non-specific side products which IVT produces when there is too little or even no template (Biebricher and Luce, 1996). Thus selection for full length cDNA may be impractical with the very small amounts of sample RNA used in this study. This belief was strengthened by the results using the protocol of Luo *et al* (1999) where amplification appears to have taken place. However the amount of amplified product was inadequate and would have required a third round of amplification. This would not have been satisfactory due to the shortening of the amplification product with every round of amplification leading to more severe 3' bias.

The difficulty encountered with the addition of the template switch and the necessity of producing full length template was probably also a major factor in the failure of the protocols that used beads. These protocols required the T7 promoter to be applied with a template switch mechanism.

Each separate enzyme reaction requires optimisation individually and with respect to each other. This is a task which is most probably better undertaken by the commercial companies. However, a measure of the difficulty of this task is the long time it has taken for competing companies to produce their own kits and very few of these are able to amplify from as low as 10ng.

The first MegaAmp® kit produced by Ambion did not work well at the very low RNA input levels necessary for this project (the company has subsequently produced an improved kit, the MegaAmpII® kit, which works with the lower amounts of RNA).

Although the protocol details are proprietary information, the RiboAmp® kit is most similar to the protocols of Eberwine and Luo in that no attempt is made to select for full length cDNA. Furthermore, instead of relying on oligonucleotides generated by the destruction of the mRNA template by RNase H as primers for the second strand reaction, an exogenous primer is used. In combination with column technology which efficiently cleans each reaction product before the next step, this kit produces approximately 50µg of amplified RNA from a 10ng sample. The removal of primers and dNTPs before the IVT is extremely important as they can contribute to non-specific product.

At the time of investigation, as the RiboAmp® kit was the only one which gave adequate levels of amplification with the material relevant to this project; it was used in all subsequent experiments

Chapter 5

DNase Treatment of RNA and Comparison of Analysis Methods

5. Optimisation of Amplification

5.1. Introduction

The process of amplification provides great opportunity for the introduction of systematic errors and DNA contamination of the RNA samples is a potential source of these. Although many workers routinely treat the RNase with DNase before amplification others do not. It has been suggested that reverse transcription from the polyA-T7 primer annealing to DNA is a rare event, and that signal from such will be neutral since the genome is the same in all cells (personal communication Per Lindahl) therefore treatment with DNase is unnecessary. As the tissue samples in this study are at the most only a few hundred cells, it would be advantageous to carry out as few procedures as possible and thus avoid loss of material through multiple purification steps and degradation. However we have compared DNase treatment with untreated samples experimentally in order to check for any contamination effects.

5.1.1 Sample Preparation

RNA can be isolated by disrupting the cells or tissues in guanidine thiocyanate buffer followed by phenol extraction (Chomczynski and Sacchi, 1987). Commercially available solutions such as TRIzol® (Invitrogen), as used in Section 4.3.4 to extract ICMs, combine phenol and guanidine thiocyanate in a mono-phasic solution in which the tissue is disrupted. Addition of chloroform separates the organic and aqueous phases allowing the isolation of the total RNA from the aqueous phase. Alternatively, after lysis in guanidine thiocyanate, the cell extract can be passed over a silica membrane in a column format. RNA longer than 200 bases will bind to the column and impurities are washed away.

The phenol based protocols have been shown to yield amplifiable RNA even with only 19 ICMs (4.3.4). However, traces of phenol in the RNA product could inhibit the downstream enzymatic reactions of amplification. Furthermore, phenol is toxic and produces organic waste for which disposal is difficult. Conversely, disposal of the waste produced from silica columns is much easier and the RNA is eluted in water. Moreover, as RNA less than 200 bases long is not bound to the silica but washed through with other contaminants, there is the advantage of enriching for mRNA: 5.8S rRNA, 5S rRNA and tRNA are selectively excluded. Another advantage of column technology is that DNase treatment can be carried out on the column if desired. Until recently the major difficulty with the columns available was the large elution volumes (30 μ l for the RNeasy® Mini Kit, QIAgen) necessitating concentration of the sample before amplification; another step where material could be lost or degraded. The introduction of the Mini RNA Isolation Kit™ (Zymo Research) circumvents this problem as the RNA can be eluted in as little as 8 μ l of water.

The RiboAmp™ (Arcturus) kit used in this study recommends treatment of the input RNA with DNase. In this experiment we ask whether treatment of our RNA with DNase before amplification, labelling and hybridisation to microarray slides results in a difference in the gene expression profile or data quality. Total RNA was extracted from groups of 10 ES colonies picked from a plate of IMT-11 at p16. The sample was split into two and one half was treated with DNase and the remaining half left untreated. The samples were then amplified, labelled and applied to microarray slides printed with the NIA 15K gene set.

5.1.2 Image Analysis

The aim of image analysis is to identify each spot through its unique address on the slide, to segment the image into foreground and background pixels and to quantify the spot intensities. Spot finding and segmentation was carried out using the package *ImaGene*TM (BioDiscovery). A deformable grid is placed over the array image allowing the spots to be assigned the unique address related to the identity of the gene probe. The pixels at this address are then assigned to either signal or background. First a target mask chosen to be larger than any spot is applied and the pixels allocated to foreground or background from the histogram of pixel values within the area of the mask. The graphics file, corresponding to the surface of the microarray, is a 16bit TIF (tagged image format) thus the intensity of each spot lies between 0 and 65,536 (2^{16}).

5.2.3 Normalisation and Data Analysis

When analysing two colour arrays relative levels are being measured and thus the two channels need normalising. Furthermore, microarray images are rarely perfect, particularly from non-commercial sources. Usually, they suffer from spatial bias, print tip bias and dye bias. Moreover, low intensity spots have more noise than high. Global normalisation by scaling in the same manner as for Northern blotting is the simplest normalisation using all the gene spots on the array. This does not take into account intensity dependant bias which shows on the MA plot as a curve. The most frequently used normalisation procedure is loess (also known as Lowess, locally weighted polynomial regression). Loess normalisation assumes that most of the spots on the array are not differentially regulated and consequently would be unsuitable for arrays where this would not be the case such as the so called boutique arrays. Loess takes a percentage of the spots and fits a curve to these spots. The larger the number of spots

(or window) the smoother the fitted curve. Print tip Loess, where a Loess curve is fitted to each print tip field, is often used as a proxy for spatial normalisation. Loess curves are fitted for each print tip (Yang *et al.*, 2002b). Variation between slides also needs to be considered.

Normalisation and analysis was carried out using both GeneSight™ (BioDiscovery) and Limma (BioConductor). GeneSight™ is a commercial package for the analysis of microarray data. It is generally simple to use providing pull down menu options for processing data. However, it is a “black box” package with limited options for data normalisation and processing. In particular, only global normalisation procedures are provided and considerable bias due to print tip variation and spatial variation can be left. It does have the advantage of being specifically designed for ImaGene data files which produce the two channel data in separate files. There is also flexibility in the spots which are included in the normalisation process; landmarks and other non-data spots such as spiked in controls can be removed from the normalisation process. Limma, on the other hand, is an R based package (<http://www.r-project.org/>) provided by the open source, open development project designed experiments and the assessment of differential expression. It has the disadvantage of being command line software but provides more options for normalisation and visualisation in analysis. In particular Limma produces an MA plot which is similar to a scatter plot but instead of the logged intensities for the two channels of a spot being plotted against each other, the log ratio of each spot (M) is plotted against its total intensity (A). This has the advantage of spreading out low intensity spots allowing better visualisation. However, Limma does not read flag information from ImaGene files thus all spots including landmarks and blanks are included in the normalisation process.

The effect of various background removal procedures was examined using Limma. Background removal by local background subtraction is not generally recommended. Firstly, it tends to increase the variance of the intensity values. Secondly, the background around a spot does not reflect the background under the spot. Finally, simple subtraction of the local background tends to result in many negative intensity or zero values leading to loss of data and interference with normalisation between arrays. Other methods of background correction which do not result in negative signal values are provided in Limma. The most used procedure is 'normexp' which by fitting a normal distribution to the background and an exponential model to the foreground, the expected background removed is calculated on the basis of the foreground observed. This results in all foreground intensities being positive. Adding a constant to all intensities reduces the variation of low intensity spots bringing the log ratios closer to zero.

In addition to within array normalisation, Limma provides methods of between arrays normalisation. Between arrays normalisation is applied after within array normalisation. Variation between arrays often arises due to differences in print quality or from differences in the scanner settings. The simplest and often the most useful method is scale normalisation which simply applies the same scale to all the M-values (Yang *et al.*, 2002b). The method "Aquantile" ensures that the A-values (average intensities) have the same empirical distribution across arrays leaving the M-values (log-ratios) unchanged.

Limma provides a list of differentially expressed genes using the functions "topTable" and statistical function "eBayes" which outputs a number of summary statistics. Instead

of a normal t-statistic a moderated t-statistic is provided. This is interpreted in the same way as the normal t-statistic but borrows information from all the genes on the array.

The B-statistic is the log-odds that the gene is differentially expressed thus a B-statistic of zero records a 50:50 chance that the gene is differentially expressed. The moderated t-statistic and B-statistic should give the same ranking of differentially expressed genes, however as these tests are based on different assumptions about the data, this may not be the case. In particular, the B-statistic depends on an estimate of the expected level of differential expression

A feature of microarray analysis is that many tests are being carried out in parallel. For an experiment where for example 10,000 genes were being tested and the critical p-value was 0.05, then 500 significantly differentiated genes would be detected by chance even if none are actually differentially regulated. Thus a false discovery rate algorithm is applied to the p-values to control the number of false positive results. If an algorithm which is too stringent is applied true positives will be lost. For microarray data, the one most frequently used is 'fdr' (Benjamini and Hochberg, 1995).

The scanned images of the microarray slides were subjected to image analysis using ImaGene™ 5.5.3 (BioDiscovery). The raw intensity data was examined using Excel to provide information about the dynamic range of intensities and the number and quality of features detected on the array. The data was input into both GeneSight™ 3.5 (BioDiscovery) and Limma (BioConductor) and the raw data displayed in a number of ways. Various background correction and normalisation procedures were applied and differential expression between DNase treated and untreated samples examined.

5.2. Materials and Methods

5.2.1 Target Preparation

A plate of the ES cell line IMT-11 at passage 16 was washed twice in DPBS-A and colonies were picked from a plate using a glass capillary. Groups of ten colonies were put into microcentrifuge tubes in a minimal volume of DPBS-A and 200µl of Zymo™ RNA Extraction Buffer was added. Samples were stored at -70°C until extraction. The RNA was extracted from two samples (20 colonies) according to the manufacturer's protocol (2.3.1b). The two purified total RNA samples were then pooled and divided into two aliquots. One aliquot was then treated with DNase (provided in kit) and cleaned using the DNA Free RNA kit™ (Zymo) according to the manufacturer's protocol. The other aliquot was left untreated. The whole amount of both the control and DNase treated samples were then amplified for 2 rounds using the RiboAmp HS™ RNA Amplification kit (Arcturus Bioscience Inc, California, USA). A 500pg kit control sample was also amplified. The amplified RNA was analysed by spectrophotometry and by denaturing agarose gel electrophoresis.

A 5µg aliquot of the amplified RNA, both control and experimental was then labelled according to the modified Atlas™ PowerScript™ Fluorescent Labelling Kit (BD Biosciences) labelling described in Table 3.9 and Step 3*. Precipitation and recovery of the aminoallyl conjugated cDNA was enhanced by the addition of 0.5µl of PelletPaintNF™ (Novagen, Merck Biosciences Ltd, Nottingham, U.K.), Section 3.4.

A 15pmol aliquot of each Cy labelled target was combined and the non specific binding sites blocked with human COT-1 DNA (Invitrogen) and Poly-dA³⁰ (MWG) as described

in Section 2.4.3. After denaturation, an equal volume of two times concentrated hybridisation buffer (Section 2.4.3) was added to the target.

5.2.2 Slide Preparation and Hybridisation

Each slide was blocked as described in Section 2.4.2 and the prepared target applied to the slides under Hybri-slips (Sigma) for 18h at 42°C according to the experimental design described in Table 5.1.

Slide Number	Cy3	Cy5
11	DNase	control
12	DNase	control
13	control	DNase
14	control	DNase

Table 5.1: Design of experiment.

After washing and drying in accordance with the conditions described in Section 2.4.4, the arrays were scanned in a ScanArray® Express HT (PerkinElmer™) equipped with red ($\lambda=633\text{nm}$) and green ($\lambda=543\text{nm}$) lasers. The slides were scanned at a resolution of 10 μ run at full speed. The pmt gain was set at 65% and the laser power at 95%.

5.2.3 Image Analysis

Spot finding and segmentation were carried out using the ImaGene™ 5.5.3 image analysis program (BioDiscovery). Screen shots of the settings used are provided in Figure 5.1. As there were no negative spots, visible as darker spots on the slides, this option was left deselected. The grid constraints were enforced and the global flexibility set as shown such that the spot finding did not misidentify contamination as spots. Automatic segmentation was used with the shown buffer and background settings as these gave the least number of incorrectly flagged spots.

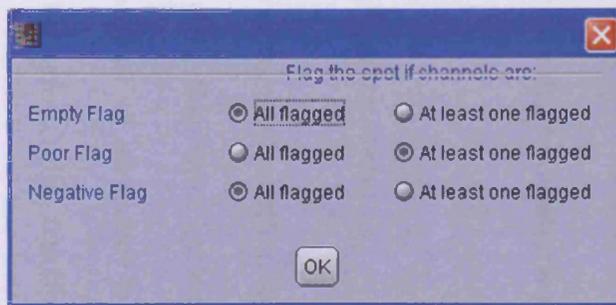
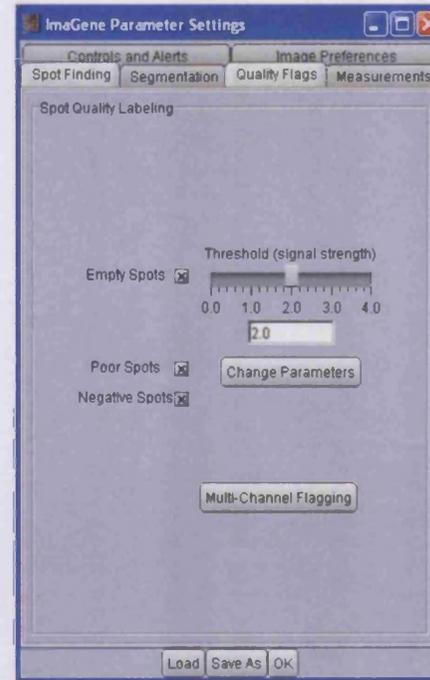
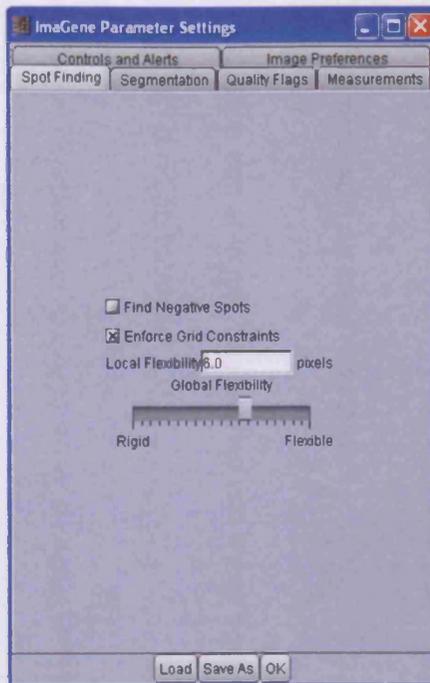


Figure 5.1: Screen shots of the settings for image analysis

Doughnut shaped spots were not allowed although this was not a problem on these slides. The threshold for empty data spots was set using the blank spots printed on the slides and adjusting till all were flagged as empty (Flag2). A spot was flagged as empty only if there was no signal above the threshold in both channels.

Negative spots were flagged using the same constraints. Spots which failed the quality checks were flagged if either channel failed. All non-data spots; Landmarks, controls, blank (printed with buffer only) and empty (not printed) spots were manually flagged as 1. Histograms and scatter plots of the raw data were generated. The text files generated by ImaGene™ were opened in Excel, the data was filtered and statistics for each slide were generated.

5.2.4 Normalisation and Analysis

Normalisation and analysis was carried out using both GeneSight™ 3.5 (BioDiscovery) and Limma (BioConductor).

5.2.4a Normalisation and Analysis using GeneSight™3.5 (BioDiscovery)

Data was loaded into GeneSight™3.5(BioDiscovery) according to the method shown in Figure 5.2. The data was processed in the sequence: omit flag types (selected for individual analyses, floor values to 20 (this raises the value of all intensities less than twenty to twenty avoiding negative values and ratios with zeros), log₂ transform the intensities, normalise with loess then subtract the log₂ green from the log₂ red values (M value) and combine the data for each array. Histograms and scatter plots were generated. Differentially regulated genes were identified by first finding the spots more than 2-fold up or down regulated. A confidence analysis test to identify the statistically significant genes at the 99% level was then carried out.

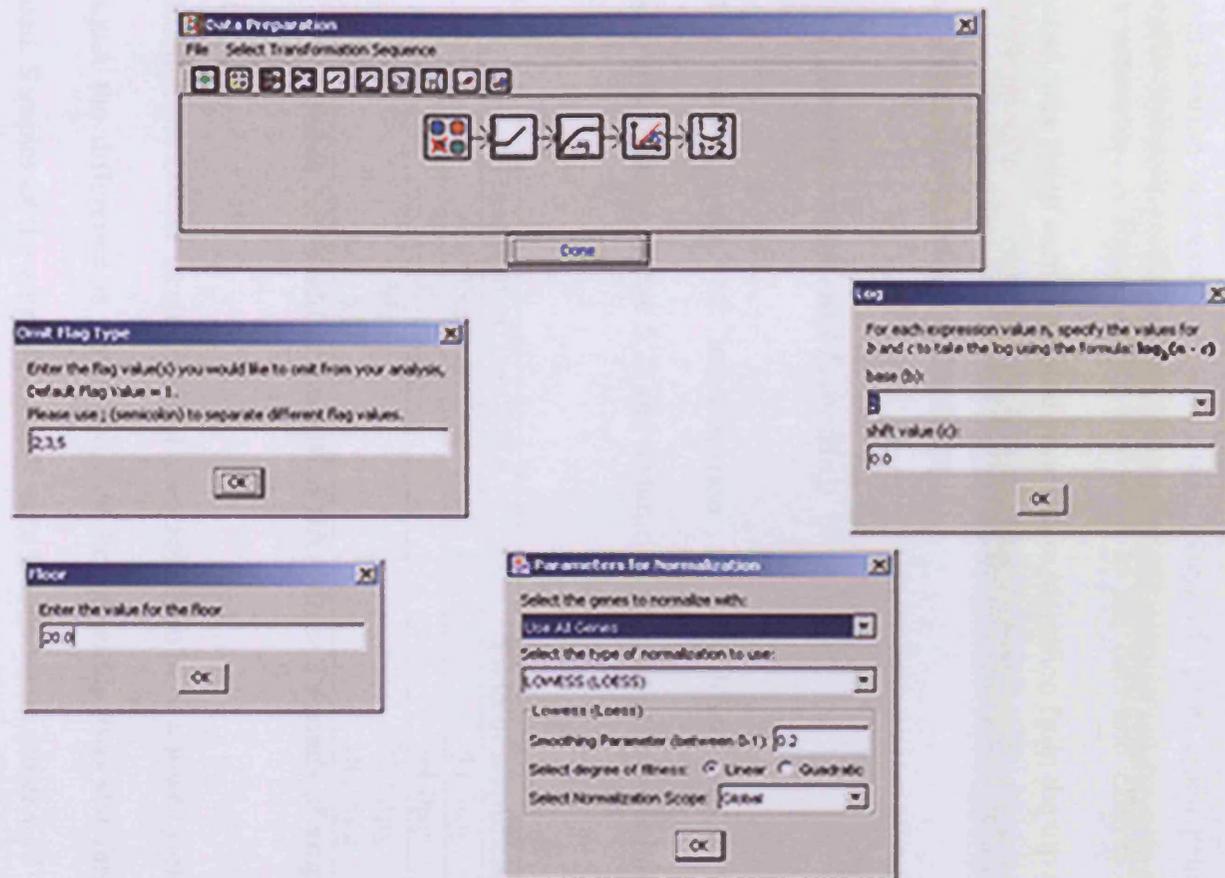


Figure 5.2: Screenshots of settings used for processing the data.

5.2.4b Normalisation and Analysis using Limma

The raw data text files were loaded into Limma using the commands in Appendix 3. Diagnostic plots of the raw data were generated. Various background correction and normalisation protocols carried out and diagnostic plots again generated to assess the consequences. A linear model was fitted to the data and eBayes statistics produced. Statistically significant regulated genes were identified from the top table.

5.3 Results and Discussion

5.3.1 Amplification and Labelling

After two rounds of amplification the amplified RNA was analysed by spectrophotometry (Table 5.2) and denaturing agarose gel electrophoresis (Figure 5.3).

Treatment of TotalRNA	Yield of amplified RNA
DNase +	57.3 μ g
DNase-	94.0 μ g
Mean	75.65 μ g
SD	18.35 μ g

Table 5.2: Yield of amplified RNA after 2 rounds of amplification.

Although the DNase treated aliquot does appear to have a lower yield than the untreated aliquot the difference is not significant being within two standard deviations of the mean. Samples of the amplified RNA were labelled with either Cy3TM or Cy5TM and the labelled products were analysed by spectrophotometry (Table 5.3) and by slide gel electrophoresis (Figure 5.4).

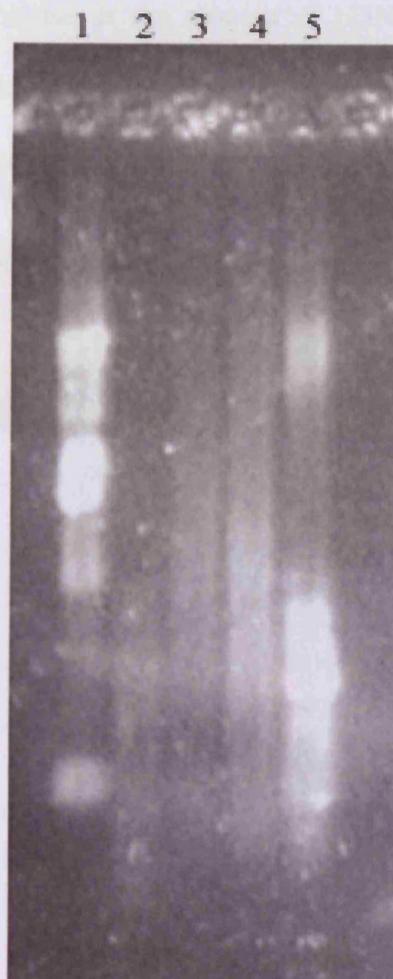


Figure 5.3: Denaturing agarose gel of 1 μ g amplified RNA

Lane1: RNA ladder.

Lane2: Kit control 10ng amplified in parallel with samples

Lane3: DNase treated sample.

Lane4: Untreated sample.

Lane 5: RNA ladder (Sigma)

Treatment of totalRNA	Yield of cDNA(μ g)		pmol of Cy dye incorporated		FOI	
	Cy3	Cy5	Cy3	Cy5	Cy3	Cy5
DNase+	440.3	407	28.0	18.0	20.6	18.0
DNase-	754.8	821.4	28.0	32.4	12.0	12.8

Table 5.3: Yield of labelled target and incorporation of dye molecules.

Again, the difference between the DNase treated and untreated are not significantly different. The apparent increase in the amount of cDNA synthesised with the same amount of dye incorporation suggests inaccuracy in the quantification of the input amplified RNA. Examination of the denaturing gel (Figure 5.3, Lanes 3 and 4) confirms this.

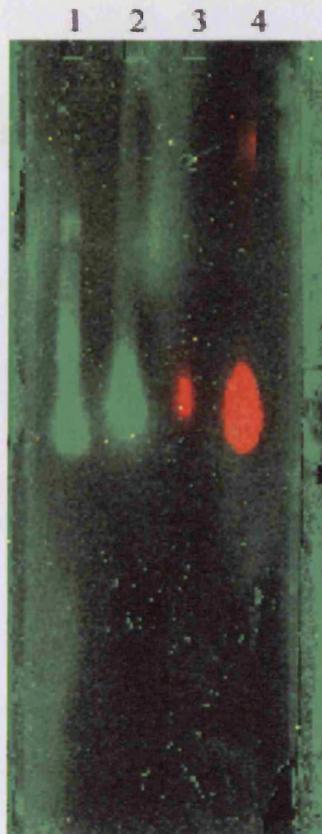


Figure 5.4: Slide gel of labelled amplified RNA target.

Lane 1: 100bp ladder, Cy3™ labelled.

Lane 2: untreated sample, Cy3™ labelled.

Lane 3: DNase treated sample, Cy5™ labelled.

Lane 4: DNase treated, Cy5™ labelled.

5.3.2 Analysis

5.3.2a Image Analysis

After filtering out all the landmarks, empty, blank and scorecard spots, the number of good, empty and poor spots detected were recorded (Table 5.4).

Slide Number	Good Spots		Empty Spots		Poor Spots		Non Data Spots	
	Number	%of total printed	Number	%of total printed	Number	%of total printed	Number	%of total printed
11	11114	68.91	3200	19.84	933	5.78	881	5.46
12	11848	73.46	2517	15.61	882	5.47	881	5.46
13	13049	80.91	1316	8.16	882	5.47	881	5.46
14	12013	74.48	2369	14.69	865	5.36	881	5.46
Mean	12006		2350.5		890.5			
SD	690.75		674.52		25.5			

Table 5.4: Number and quality of spots found by ImaGene on each slide.

All four slides were very similar with no real difference in the dye reversed pair. The maximum and minimum intensity of the spots show that the data is within the linear range of the assay, none of the spots are below zero (negative spots) and none are over 65,536; the maximum detection limit of the scanner 2^{16} (Table5.5).

However, some of the lower intensity spots should be disregarded as they are lower than the mean background intensity + 2 standard deviations.

	11				12			
	Cy3 DNase		Cy5 Control		Cy3 DNase		Cy5 Control	
	Signal	Background	Signal	Background	Signal	Background	Signal	Background
Total	39719697	2874733	31469286	2836929	43531501	2653556	30767081	2695366
Maximum	52429.19	1968.7	49507.96	1013.295	49866.72	7818.427	50975.3	5521.066
Minimum	155.9058	140.7585	159.4688	140.5767	139.7211	130.6407	155.3575	136.6654
Mean Spot	2774.885	200.8337	2198.497	198.1926	3030.386	184.7237	2141.809	187.6343
SD	5442.502	39.12995	4550.162	30.49111	5654.837	87.52242	4392.979	49.67983
Slide Number	13				14			
	Cy3 Control		Cy5 DNase		Cy3 Control		Cy5 DNase	
	Signal	Background	Signal	Background	Signal	Background	Signal	Background
Total	26689105	2924201	65726756	3592284	26596318	2918505	44686411	3337158
Maximum	47101.77	1784.457	53177.62	2644.202	51842.26	8090.993	54010.8	13874.29
Minimum	161.9427	154.9125	193.5648	177.2587	167.0465	146.9656	186.6628	168.9225
Mean	1857.926	203.5643	4575.479	250.072	1849.407	202.9417	3107.323	232.0532
SD	3755.291	33.75438	7517.186	44.64857	3771.753	69.87828	5723.164	119.1988

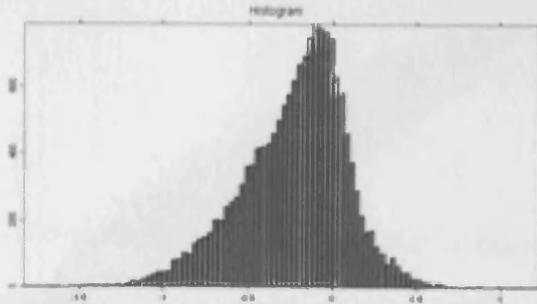
Table 5.5: Foreground and background signal intensities for each array slide.

Examination of the histograms and scatter plots (Figure 5.5 and 5.6) suggests that some spatial and between array slide normalisation is required. The distinct banana shaped scatter plot from Slide 13 shows strong evidence of intensity dependant dye bias. Spots from the scorecard which have hybridised with the target are evident as both a shoulder on the histogram and a low intensity group on the scatter plot. This is also present on slide 14 to a lesser extent. As the dyes are reversed with respect to slides 11 and 12, this is evidence of dye bias. However, the dye swap should be sufficient to counteract this.

5.3.2b GeneSight™ (BioDiscovery)

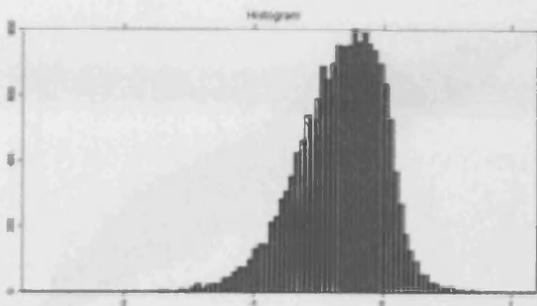
The analysis package GeneSight™ is designed for full integration with the image analysis program ImaGene™ and can be used to remove spots from the normalisation process by utilising the spot quality flags. It provides a number of tools with various options for the processing of microarray data. After processing the data scatter plots were produced. To assess the impact of non-data spots on normalisation and detection of differential expression analysis was carried out using various categories of spots (Figure 5.7).

The slope of the regression line under all three analysis conditions is very close to 1 indicating very little difference in the two samples being examined. The inclusion of empty data, poor or non-data spots seems to have little impact on the relationship between the samples. The R^2 value, also very close to 1, indicating the goodness of fit.



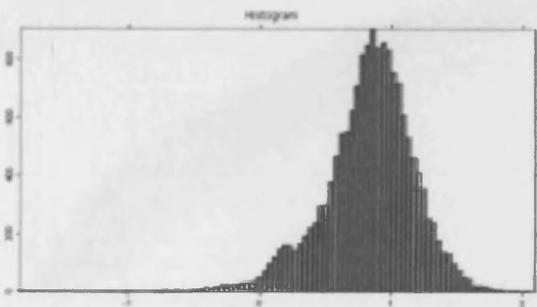
(a) Slide 11: Untreated Cy5™

DNase Cy3™



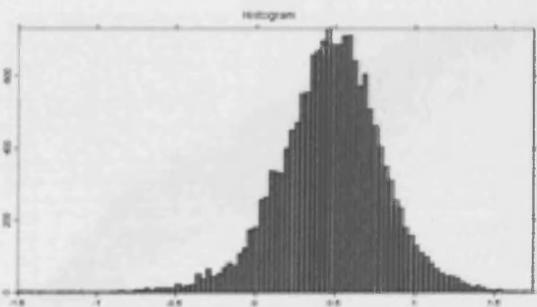
(b) Slide 12: Untreated Cy5™

DNase Cy3™



(c) Slide 13: DNase Cy5™

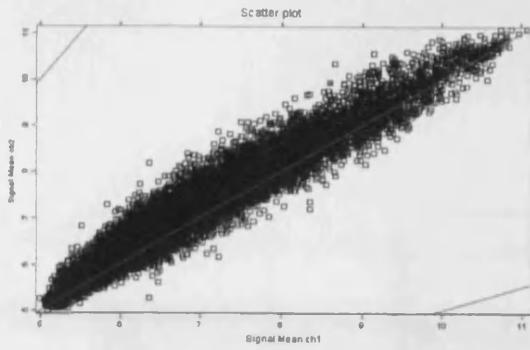
Untreated Cy3™



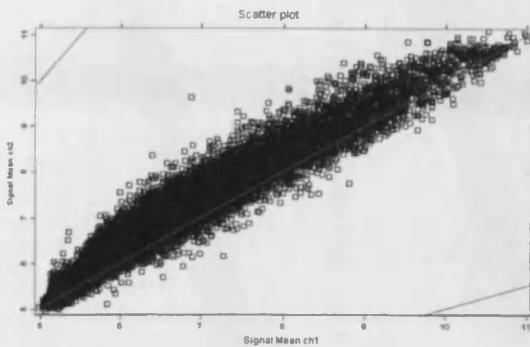
(d) Slide 14: DNase Cy5™

Untreated Cy3™

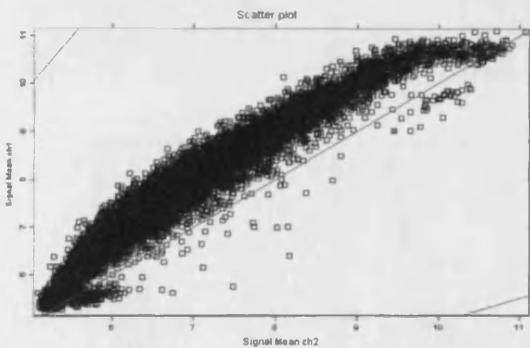
Figure 5.5: Histograms of the raw data $\text{Cy5}^{\text{TM}}\text{intensity}/\text{Cy3}^{\text{TM}}\text{intensity}$ The ratio of two channels is plotted on the x-axis and the frequency plotted on the y-axis.



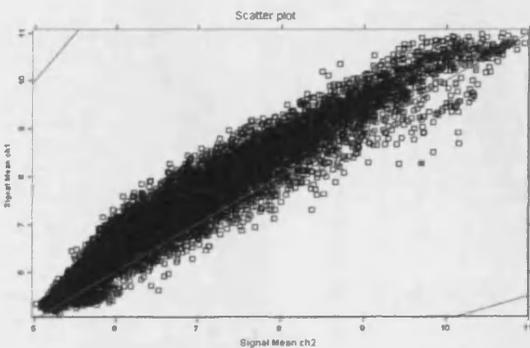
(a) Slide 11: Untreated Cy5TM
DNase Cy3TM



(b) Slide 12: Untreated Cy5TM
DNase Cy3TM



(c) Slide 13: DNase Cy5TM
Untreated Cy3TM



(d) Slide 14: DNase Cy5TM
Untreated Cy3TM

Figure 5.6: Scatter plots of raw data. x-axis=control, y-axis=DNase treated. The spot marked red is a landmark for orientation.

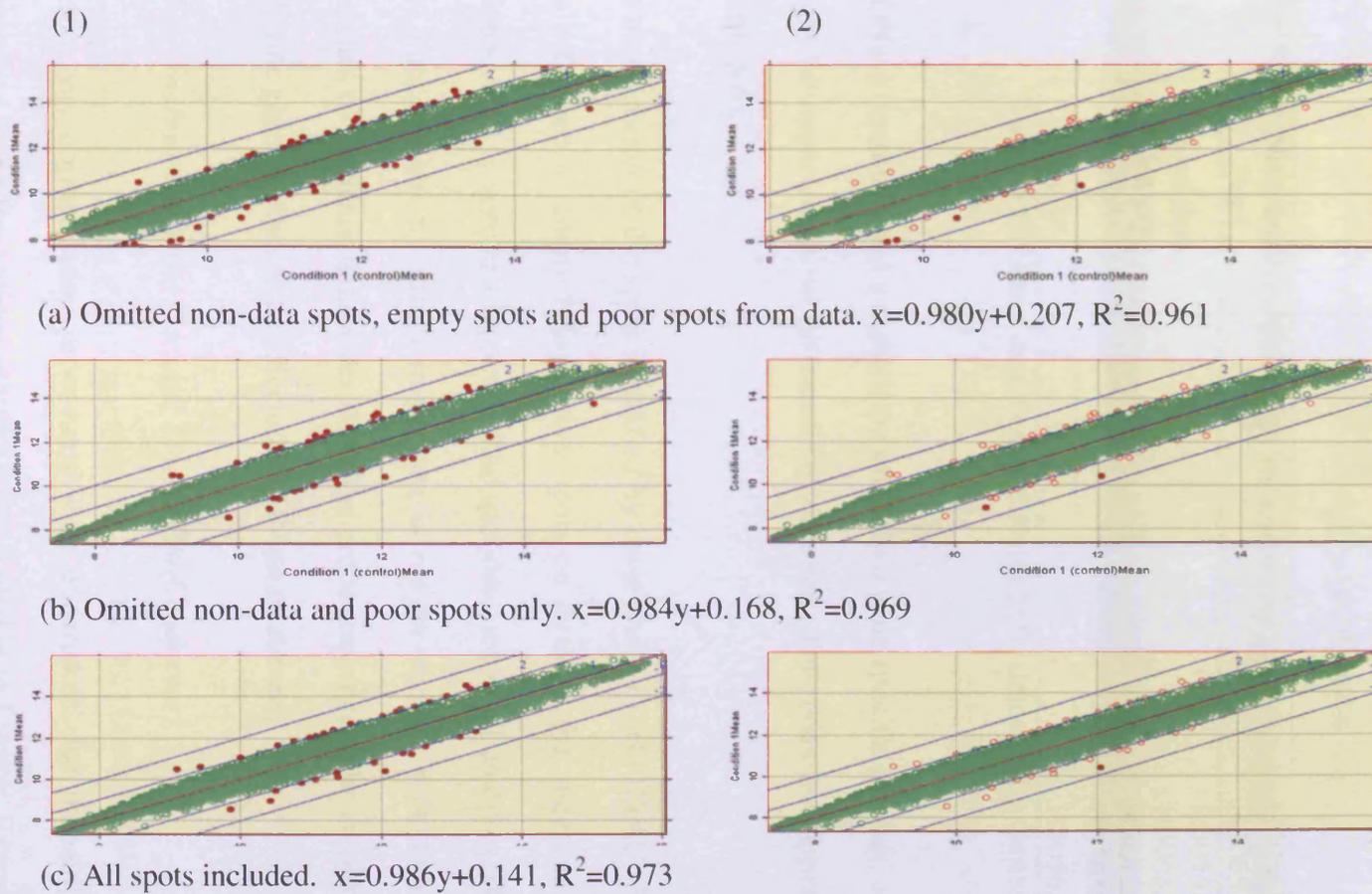


Figure 5.7: Scatter plots of analysis of data after normalisation. Red spots indicated regulated genes. Plots in column (1) show spots more than 2-fold regulated. Plots in column(2) are 2-fold regulated at the 99% confidence limit. x =amplified RNA from untreated total RNA, y = amplified RNA from DNase treated total RNA.

Spot Types Omitted	Genes more than 2-fold different	Genes more than 2-fold different at 99% confidence level	Identity
Non-data, Poor	42	2	H3056D06-3 H3065A01-3
Non-data, Empty Poor	46	4	H3056D06-3 H3065A01-3 H3072A07-3 H3105C08-3
none	33	1	H3056D06-3

Table 5.6: Genes detected as differentially detected by GeneSight™

If DNase treatment had a significant effect in a gene specific manner, a number of genes might be expected to be differentially detected. This does not appear to be the case (Table 5.6).

On examination of the spots identified by these analyses, H3072A07-3 and H3105C08-3 are flagged as empty (Flag2) data spots on three of the slides out of the four and H3065A01-3 is very low intensity. The spot detected in all analyses, H3056D06-3, is of low intensity but by visual inspection is not an artefact such as dirt on the slide. Searches of sequence data bases show the probe sequence to be an unknown EST. Thus only one gene appears to be affected by DNase treatment.

5.3.2c Analysis using the Package Limma (BioConductor)

Before any normalisation procedures were undertaken, the unmanipulated data was scrutinised using the various plotting tools available in Limma. Examination of the red and green background image plots showed that these slides had few background problems (Figure 5.8). There is however, an area of higher background on metagrid

(2,4) and in the corner (12,4). This is not so easy to see in the print tip loess image although the need for normalisation is clear (Figure 5.9(a)).

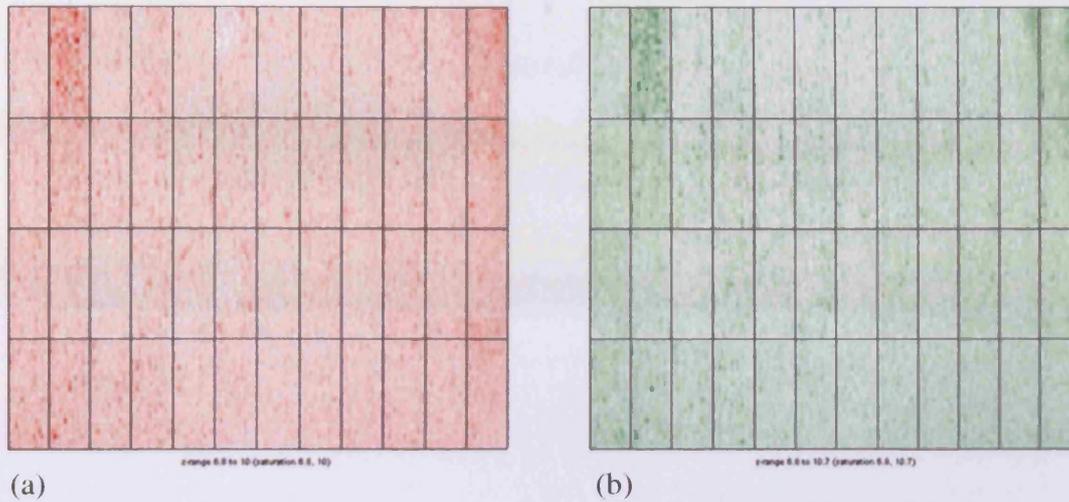


Figure 5.8: Example of image plots of the background obtained from the raw data from slide 11.(a) Cy5 background (b) Cy3 background.

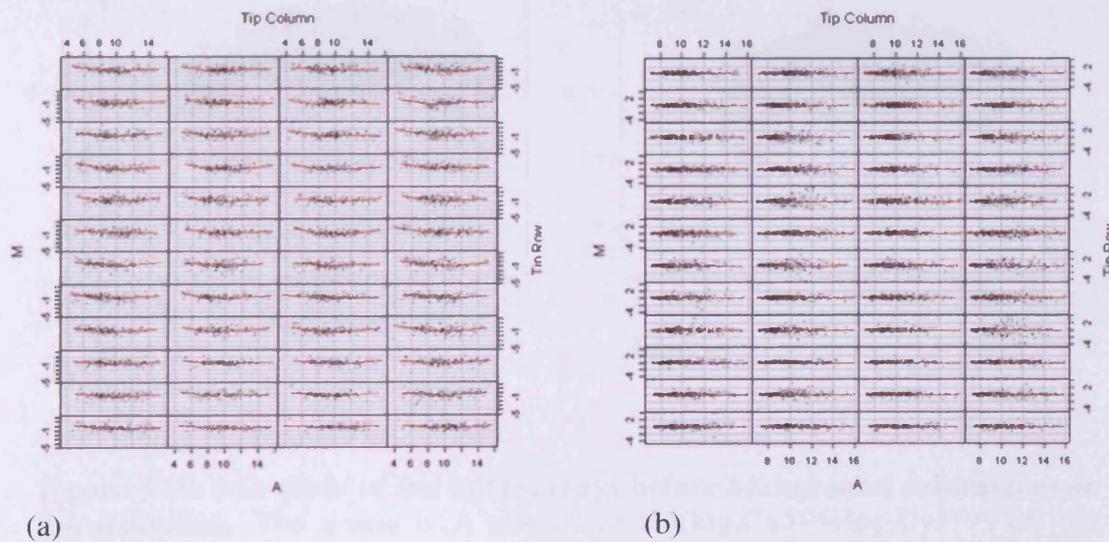


Figure 5.9: Example of print tip loess plot of the data from slide 11 (a) before normalisation and (b) after printtip loess normalisation.

Figure 5.10 shows the MA plot for each slide before normalisation.

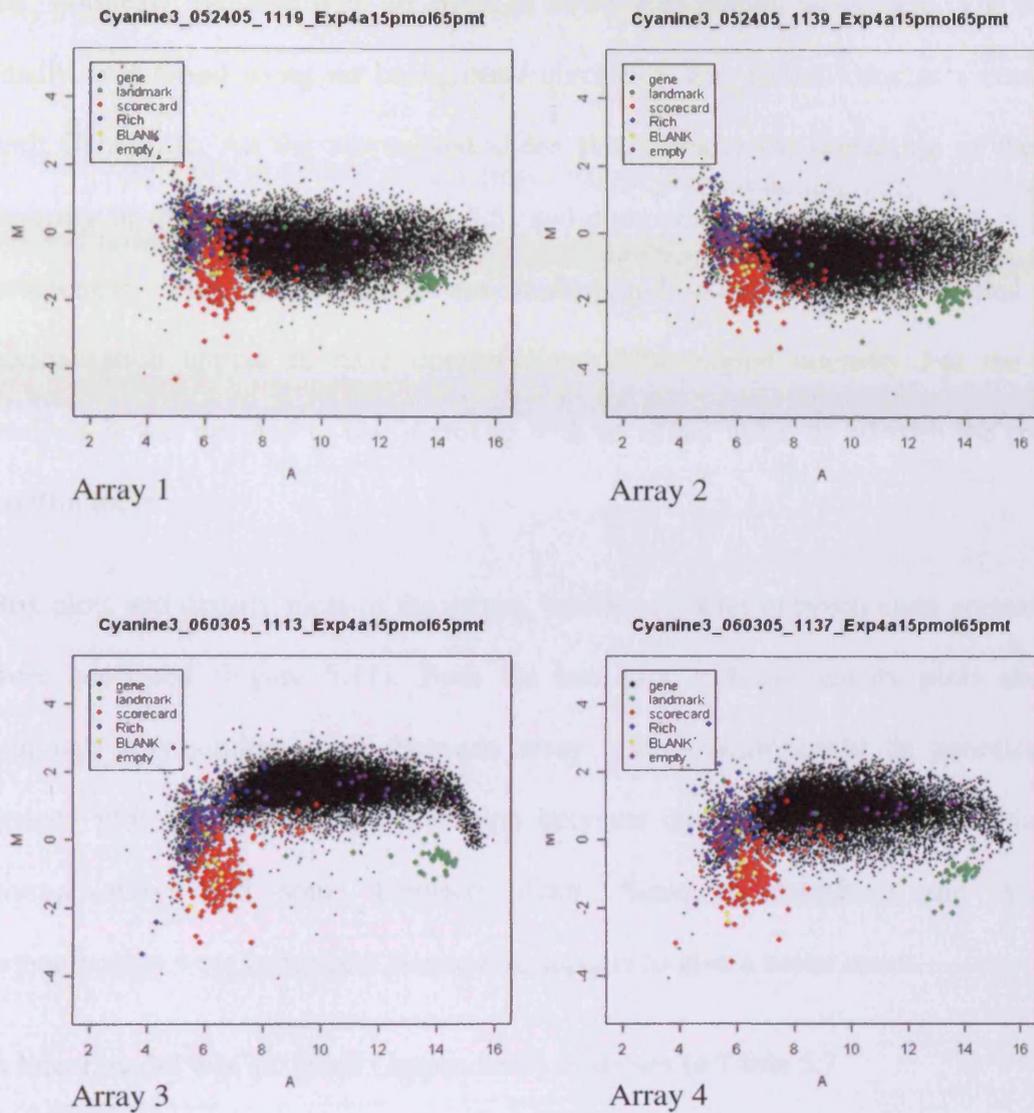


Figure 5.10: MA plots of the microarrays before background subtraction or normalisation. The x-axis is A where $A = 0.5(\log_2 \text{Cy5}^{\text{TM}} + \log_2 \text{Cy3}^{\text{TM}})$ i.e. the total spot intensity and M is the y-axis where $M = \log_2 \text{Cy5}^{\text{TM}} / \text{Cy3}^{\text{TM}}$ i.e. the fold change in intensity.

Background is removed from the foreground intensity before within array normalisation. The default procedure for Limma is to subtract the local background and carry out print tip Loess normalisation. First the data was normalised by print tip Loess

with no background correction. The normalisation procedure was then repeated using the “normexp” function with an offset of 50 for background adjustment. The data was finally normalised using no background correction and global loess as a comparison with GeneSight. All the normalised slides show clearly the correction of the higher intensity in the red channel (Table 5.5) and compensation for the intensity dependant component. The slides where no background correction was applied before normalisation appear to have compression of the signal intensity. For the Limma analysis it was decided to use normexp with an offset value of 50 with the command `printtip loess`.

Box plots and density plots of the arrays, before and after between slide normalisation, were generated (Figure 5.11). Both the box plot and the density plots show that although very similar, some between array normalisation could be beneficial. The density plot shows very little variation between channels (this is after within slide normalisation) and some between slides. Scale normalisation and A-quantile normalisation were compared. A-quantile appears to give a better result.

A linear model was designed (Appendix 3) as shown in Table 5.7.

Array	DNase
1	-1
2	-1
3	1
4	1

Table 5.7: Design matrix of linear model.

Coefficients used to model data for this simple dye swap experiment.

$$\text{Cy5}^{\text{TM}}/\text{Cy3}^{\text{TM}} = -1(\text{Cy3}^{\text{TM}}/\text{Cy5}^{\text{TM}})$$

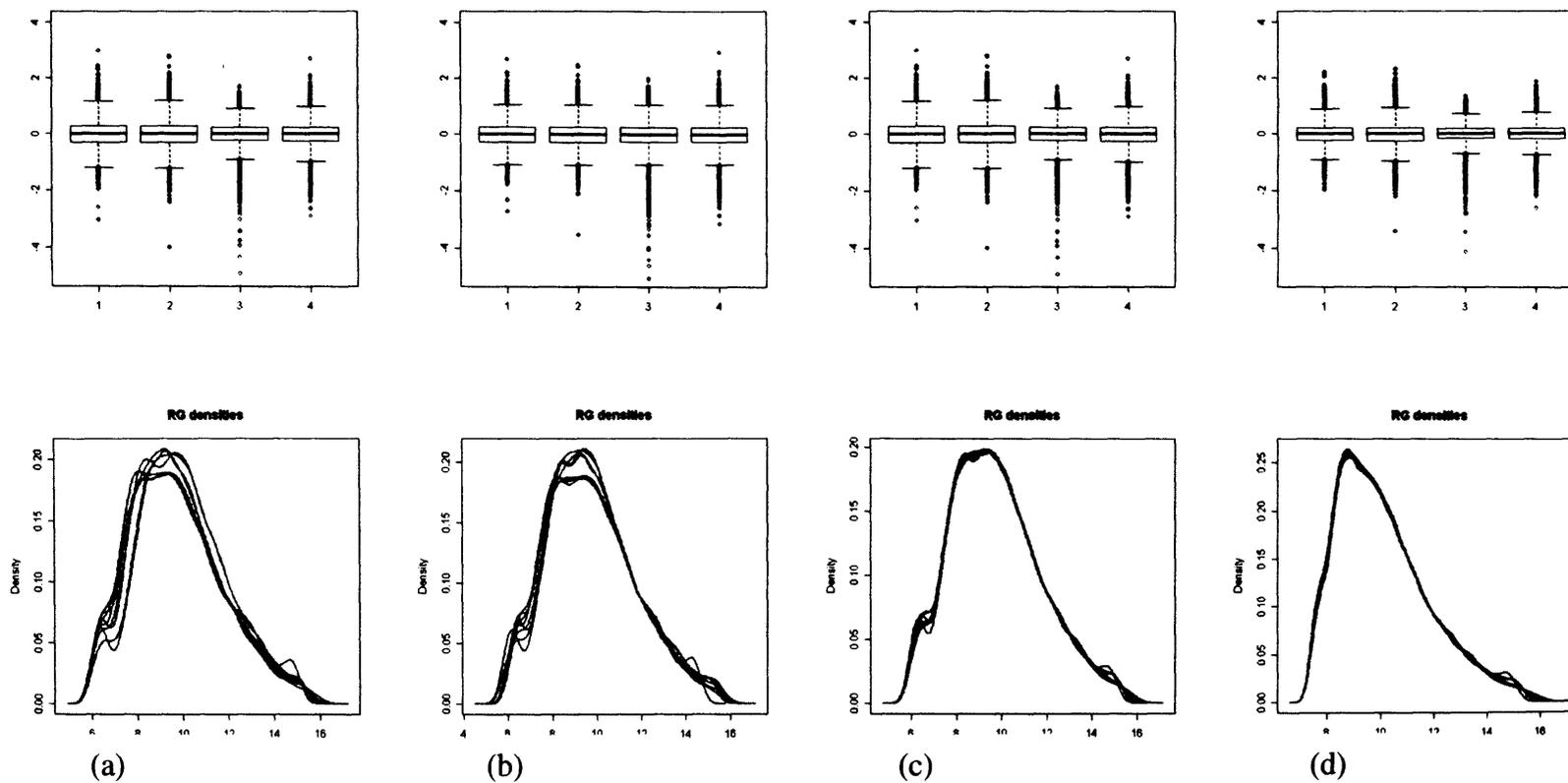


Figure 5.11: Box plots and density plots of (a) the four print tip loess normalised slides before between array normalisation (b) after scale normalisation (c) after A quantile normalisation and (d) with no background correction before print tip Loess A-quantile normalisation.

The linear model was fitted to the data, and the eBayes statistics generated. The TopTable command is used to output the data. When the data was only subjected to global normalisation, no significant genes were found, only landmarks and Scorecard spots were significant.

Method	Background Correction	Within Array Normalisation	Between Array Normalisation	Genes Found
1	None	Printtip Loess	None	6
2	normexp	Printtip Loess	None	9
3	normexp	Printtip Loess	Scale	0
4	normexp	Printtip Loess	A-quantile	9
5	None	Printtip Loess	A-quantile	6
6	None	Global Loess	None	0

Table 5.8: Summary of the analysis of the data processing and the number of genes differentially detected.

The nine genes found as being differentially detected by methods 2 and 4 (Table 5.8) are identical both in rank and identity. The six genes found by Method 1 are the first six genes in the list of nine albeit in a slightly altered rank order. The six genes found by method 5 are the same as those found by method 1 with only the last two genes in reverse order compared to Method 1. Methods 3 and Methods 6 both yield no spots.

As this data is unfiltered and includes spots with a signal less than the background mean plus two standard deviations, intensity data for the nine spots were examined in Excel. Seven of these genes had intensities greater than the background mean + 2 standard deviations in both channels of all four slides. Two of the genes had one slide where one channel was below this cut-off value. Examination of the M-values for these genes reveals only the three were more than 3-fold different. One was just under 3-fold different. The remaining five were only marginally over 2-fold different (Table 5.9)

	Field	Meta.Row	Meta.Column	Row	Column	Gene.ID	Status	M	A	t	P.Value	B
6493	A	5	4	6	4	H3050H04-3	gene	-1.771	8.86	-13.77	0.0169	5.02
8105	A	7	1	2	20	H3019A12-3	gene	1.625	9.52	10.67	0.0488	3.84
10240	A	8	3	8	13	H3079C09-3	gene	-1.429	10.61	-9.94	0.0488	3.48
13256	A	10	4	8	5	H3075D05-3	gene	1.148	13.07	9.29	0.0488	3.12
13494	A	11	1	3	12	H3027A03-3	gene	1.146	11.46	9.16	0.0488	3.04
13040	A	10	3	13	20	H3135C05-3	gene	1.119	12.96	9.08	0.0488	3
9899	A	8	2	8	8	H3073B10-3	gene	-1.13	8.41	-9	0.0488	2.95
5156	A	4	4	6	11	H3056D06-3	gene	-1.857	11.11	-8.87	0.0488	2.87
1355	A	2	1	1	11	H3006A09-3	gene	-1.17	11.72	-8.81	0.0488	2.83
14376	A	11	3	13	12	H3131C03-3	gene	0.983	10.37	8	0.0771	2.29

Table 5.9: Table of top ten differentially expressed genes. The data was processed by removing the background with normexp Offset=50, normalisation within arrays by printtip Loess and normalisation between arrays by A-quantile or with none. Only the first 9 genes are statistically significant. $p < 0.05$. When processed by within array printtip loess and without background removal or between array normalisation only the first six genes, highlighted in yellow are found to be significantly different at $p < 0.05$. Note actual p values of yellow list are different.

The gene H3056D06, which is the only gene detected by GeneSight™ (BioDiscovery), is also on the list of 9 genes and has the highest M-value (-1.857)

5.4 Conclusions

The purpose of the work presented in this chapter was to assess the impact of DNase treatment on the outcome of array analysis and to examine some of the statistical methods available.

Treatment of the RNA samples before amplification with DNase had little effect on the yield of amplified RNA and its subsequent labelling. Furthermore, the detection profile of the genes was only very slightly altered. As two samples, Dnase treated and untreated, were being compared and few differences between the treatments were expected, a direct comparison was made with reversed dyes to allow for any dye bias. The histograms of the raw data demonstrate the necessity for the dye swap design.

Analysis with GeneSight™ was used to assess the impact of the different categories of spots on the finding of differential detection. The non-data spots, in particular the Landmarks might be expected to have the largest impact on identification of differentially detected genes but their removal resulted in the detection of one extra gene (2 in total) in comparison to analysis where no spots were removed from the normalisation process. However, this spot (H3065A01-3) is of very low intensity and would normally be eliminated from the analysis. The further two spots assigned as differentially detected when the empty spots (flagged by ImaGene™) are removed from the normalisation process are also empty spots. This leaves a single gene (H3056D06-3) as potentially affected by DNase treatment.

The extensive plotting tools in Limma allowed closer scrutiny of the slides before any statistical procedures were applied and the affects of these could be assessed. Under the most stringent analysis conditions, a maximum of nine genes are shown to be differentially detected. However most of the genes detected would be described as marginal. Subsequently, the genes detected have been analysed by semi-quantitative RT-PCR (Hunter *et al.*). The gene H3056D06-3, was detected by both GeneSight™ and Limma when background correction was carried out by “normexp”, printtip Loess within array normalisation, and with either no between array normalisation or with “A-quantile” between array normalisation. Although very clear on the arrays, it proved to be a false positive when checked by semi-quantitative RT-PCR. The two genes H3050H04-3 and H3073B10-3 were confirmed as being differentially detected by semi-quantitative RT-PCR. These two genes were not detected at all by GeneSight™ and both were only detected by Limma when “normexp” was used for background correction, with printtip Loess and either no “between array” normalisation or A-quantile “between array” normalisation. Thus methods 2 and 4 (Table 5.8) yield two positive results and seven false +ves, while methods 1 and 5 detect only 5 false positives at the cost of losing one positive result. Methods 3 and 6 yield no false positives but at the expense of losing the positive results.

This data demonstrates how sometimes a decision must be made whether it is more important to detect all differentially regulated genes at the expense of more false positives.

Chapter 6

Comparison of the Transcriptomic Profile of Embryonic Stem Cells and the Pluripotent Compartment of the Embryo

6. Embryonic stem cells and transcriptomic analysis

6.1 Introduction

ES cells are the cells of the pluripotent compartment of the embryo isolated into culture (Evans and Kaufman, 1981; Martin, 1981). Their ability to integrate into an embryo and contribute to chimaeric animals (Bradley *et al.*, 1984; Nagy, 2003) and the expression of the pluripotency associated markers; *nanog* (Chambers *et al.*, 2003; Mitsui *et al.*, 2003), *Pou5f1* (Oct3/4) (Nichols *et al.*, 1998) and *Zfp42* (Rex1) (Pelton *et al.*, 2002); argues identity with the pre-implantation ICM. However, protein expression profiling using two-dimensional electrophoresis indicates a later phenotype, that of the embryonic ectoderm (Lovell-Badge and Evans, 1980). Gardner *et al* (1985) (Gardner *et al.*, 1985) injected ICM cells from blastocysts up to Day 5 early implanting stage and obtained chimaeras. Injection of cells from embryos more advanced than this failed to produce chimaeras. Furthermore cells from the ICM of blastocysts undergoing lactational delay or reactivated delayed blastocysts did not produce chimaeras even though these can implant and develop to term. Lawson *et al* (1991) by injecting the lineage marker HRP into a single cell of the post-implantation ectoderm showed that after *in vitro* culture these cells could contribute to mesoderm thus confirming the ability of these cells to contribute to multiple cell lineages.

Development of the embryo has been described in terms of progressive loss of totipotency. The fertilised egg is fully totipotent, on reaching the blastocyst stage the embryo is subdivided into two restricted tissues, the inner cell mass (ICM) and its trophoctoderm. The trophoctoderm is restricted to contributing to the Reichert's membrane and placenta while the ICM develops further to produce a layer of primitive

endoderm. This primitive endoderm migrates round the inner surface of the trophoctoderm to form the parietal endoderm while the endoderm which remains in contact with the ICM becomes the primitive visceral endoderm. Over this peri-implantation period the ICM is proliferating and is often at this stage referred to as the epiblast. Signals from the visceral yolk sac endoderm, the layer of endoderm which remains in contact with the ICM, trigger the epiblast to form a pro-amniotic cavity and to transform into a pseudo-stratified epithelium; the embryonic ectoderm.

Over this period the ICM, epiblast and embryonic ectoderm have the same potency *in vivo* and will give rise to the whole foetus. They also will all form teratocarcinoma, a tumour consisting of both differentiated and stem cells (Section 1.1.1) However ES cells have so far only been derived up to the late blastocyst stage. ES lines have not been derived from early embryonic layers once the embryonic ectoderm has been formed. Although cell lines known as EpiSCs which share some features of ES cells have been isolated from pre-gastrulation stage embryonic ectoderm (Brons *et al.*, 2007; Tesar *et al.*, 2007) these cells, in common with human ES cells, do not depend on LIF to maintain their undifferentiated state and furthermore, like EPL cells do not contribute to chimaeras. The unique gonadal tumours, teratocarcinomas, may hold a clue to the origin and *in vivo* homology of ES cells. Testicular teratocarcinomas spontaneously arise in the gonads of strain 129 mice by proliferation of germ cells in the genital ridge of male mice (Stevens, 1967) and can be induced experimentally by the transplantation of Day11.25-12.5 genital ridge to an ectopic site (Stevens, 1964). However, an embryonic origin for these cells has also been postulated. This was tested by transplantation of embryos to ectopic sites (Stevens, 1970) and it was shown that teratocarcinomas were formed from pre-implantation embryos. Furthermore, from post-implantation embryos

up to gastrulation stage, only fragments which included embryonic ectoderm would form teratocarcinomas (Diwan and Stevens, 1976).

A transplantable teratocarcinoma generated from a preimplantation embryo by Stevens, was used to derive clonally isolated cell cultures which were shown to be pluripotent both *in vivo* by forming teratocarcinomas on injection into mice (Evans, 1972), and *in vitro* via embryoid bodies (Martin and Evans, 1975a; Martin and Evans, 1975b). These cells were shown to be able to contribute to chimaeras. Subsequently, ES cells were derived from delayed blastocysts without the *in vivo* tumour step (Evans, 1972; Evans and Kaufman, 1981). These cells share with EC cells the ability to grow indefinitely in culture, to differentiate *in vivo* as teratocarcinomas and *in vitro* as embryoid bodies. Furthermore, when injected into a blastocyst, they not only contribute to the somatic tissues of the animal but also to the germ line.

The central question to this thesis is whether ES cells have any homology to the pluripotent compartment of the mouse and to what degree they are an artefact of culture. Molecular phenotyping using techniques for profiling the expression of individual genes such as Northern blotting, PCR and *in situ* hybridisation have provided much information but fail to give a wider picture. Expression of individual genes associated with the pluripotent state in EC and ES cells such as Oct3/4 have been investigated in the embryo and their function *in vivo* studied via knockouts of the genes. A functional genomics approach using microarrays would be expected to show this broader picture as well as yield information on individual genes. Although the proteome can be argued to be more representative of the state of a cell than the transcriptome, not all mRNA is translated into protein; the ability to carry out comparisons is restricted by the amount of material available. While the amount of material available for comparisons between

Chapter 6: Embryonic Stem Cells and Transcriptomic Analysis

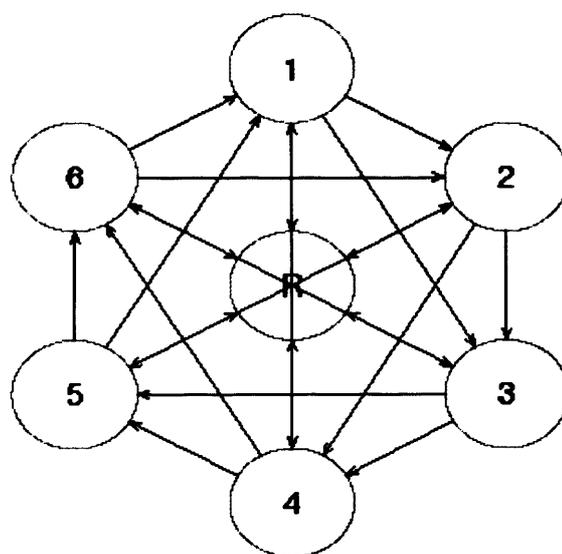
ES and EC cells is limited only by the number of vessels they can be grown in, the amount of material from the pluripotent compartment of the embryo is very small. It is possible to amplify the transcriptome while maintaining complexity and without loss of the original transcript levels.

Microarrays are uniquely able to provide expression data for thousands of genes in parallel and thus provide a snapshot of gene activity. This, in combination with the ability to amplify RNA, can be utilised to compare ES cells to the pluripotent compartment of the embryo at various stages. The data thus obtained can be used to identify which developmental stage ES cells most resemble and perhaps reveal pathways to specific fates. The NIA 15k cDNA gene set is derived from libraries of ESTs from embryonic tissues making it a valuable resource for studies such as these. This clone set consists of 15,247 clones which represent approximately 12,000 unique transcripts. It was assembled from cDNA libraries of pre-implantation embryo stages, day 7.5 embryonic and extra embryonic portions, ovary/mesonephros from Day 12.5 embryos and ovary from newborn pups (Carter *et al.*, 2003). Furthermore, the clones are ESTs and thus the 3' bias and reduced transcript length of the amplified target is less problematic than for oligonucleotide arrays.

In this study ES cells were compared to various pluripotent cells of the peri-implantation embryo namely the Day 3.5 ICM, Dy4.5 ICM and the early post-implantation embryonic epiblast and ectoderm on Day5.5 and Day6.5 of gestation respectively. In addition, ES cells were compared with ICMs from delayed blastocysts (Section1.1.2) on the equivalent day of gestation (EDG) 6 and 8. Frequently ES cell lines are isolated from such blastocysts. Furthermore, at these developmental stages, unmanipulated embryos would normally have implanted.

6.1.1. Design of Experiment

The design of a microarray experiment is not always obvious and many features have to be considered if the primary aim of the experiment is to be achieved. Balanced dye loop designs have the advantage of economy; they usually require less array slides and have lower variances, as the number of conditions being compared increase, the paths between the comparisons increase leading to increased variances. This can be overcome by using an interwoven loop design. However, as the number of conditions being compared increases, the experiments become more complex with greater capacity for operator error.



A-Optimality Score for Contrasts: $\text{Tr}[\text{Inv}(CX)] = 3.7145$.
Nr of arrays = 24 . Nr of conditions = 7

Figure 6.1: Balanced dye loop design with reference sample. The reference sample in this case would be ES cells. This design was produced using the online tool http://exgen.ma.umist.ac.uk/R_public/ExpDes_v2.htm with the help of Ben Routley. This design uses both an interwoven loop without dye reversal and a common reference with the reference sample being of biological significance. The common reference slides require dye reversals to be carried out. This design would require 24 slides for 7 samples.

The addition of the common reference design to the interwoven loop design minimises the variances but for example shown in Figure 6.1 with seven samples twenty-four slides are required. Furthermore, the failure of any individual array becomes more critical.

A reference design does require more array slides and, as most or all of the comparisons are indirect (on separate slides); the variances are higher than for a direct or a loop design. However, the reference design is more flexible (Figure 6.2). If desired, more conditions can be added to the experiment. Furthermore, loss of a slide due to breakage or hybridisation failure is recoverable as long as there is enough RNA sample. Analysis of the comparisons of interest is also simpler.

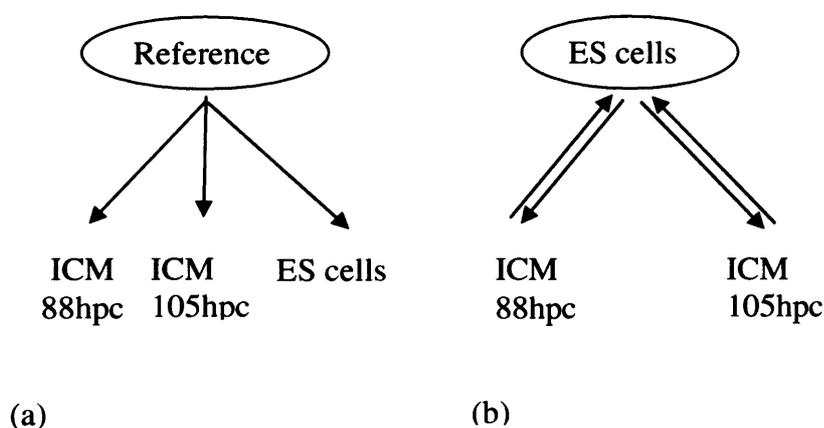


Figure 6.2: Two examples of experiment designs using a common pool. In (a), a common reference pool of no intrinsic interest is used requiring three array slides to compare the three samples of interest with indirect comparisons. In (b), four slides are required as dye swaps have to be performed. Although the comparison between the ICM samples and ES cells are direct, the difference between the two ICM samples can be estimated indirectly.

Although an important factor in the choice of reference is abundance, whether the choice of reference could also be a sample of interest, should also be considered. On

one hand, if the reference sample is of no interest, then half the possible information on the slide is of no interest. On the other hand, when the common reference is also a sample, gene specific dye effects must be taken into consideration so more slides have to be used. Another concern is the number of spots detected by the reference sample. Ideally all the spots on the array should be detected by the reference sample otherwise there will be loss of data. Consequently, for small arrays, the library represented by the spots is often used as a reference sample.

The ease with which ES cells can be genetically manipulated has led to the belief that the DNA is open and that transcriptional control is leaky (Evans *et al.*, 1997). Recently, the DNA from ES cells has been shown to be in a more open configuration (Meshorer *et al.*, 2006). If this does indeed lead to expression of all transcripts albeit at a very low level, ES cells could be unique as a universal reference sample in two colour arrays. Alternatively, a pool of all the tissues which were used to compile the gene libraries represented on the arrays can be used as a common reference. The pool strategy has the advantage that no dye swaps are required since the comparisons of interest are all labelled with the same dye. The disadvantage is that all the comparisons of interest are indirect, leading to increases in the variances.

A pilot experiment was undertaken to test the usefulness of ES cells as a universal reference before making a reference pool.

The reference pool consisted of the embryos and tissues which were used to construct the libraries from which the NIA 15k gene set was derived (Carter *et al.*, 2003). (Section 6.2.2, Table 6.1)

6.1.2. Image Analysis

Image analysis was carried out using the commercial packages ImaGene™ (BioDiscovery) and Spot (CSIRO). ImaGene™ (BioDiscovery) uses histograms to segment the image pixels into foreground and background. Although this appears to function well, it suffers from the drawback that the version of the program held by this institution does not process batches of slides thus every slide must have the grid applied and adjusted individually. Furthermore, the local background measurement in most cases can not simply be subtracted from the foreground as this often produces too many zero or negative values which not only compromises “between array” normalisation but leads to loss of data. The variance in the data can also be increased leading to loss of data especially with low intensity spots (Yang *et al.*, 2001). Spot software is implemented as a package within R (<http://www.r-project.org/>) allowing the powerful suite of R tools, particularly the package limma (BioConductor) to be used directly on the results produced by Spot software. Spot is an adaptive shape segmentation method which uses seed points from which regions grow outwards preferentially according to the difference between a pixel’s value and the running mean of values in an adjoining region. This software has the big advantage of being able to use a template to automatically process a batch of slides thus reducing the amount of user intervention required. It also accommodates irregular shaped spots. The background measurement method ‘morph’ has the effect of smoothing and damping the variance of the background and usually returns a lower background than ImaGene™. It allows simple background subtraction without generation of many zero or negative values.

6.1.3. Analysis

The image files generated by either ImaGene™ (BioDiscovery) or Spot (CISRO) were input to either limma using limmaGUI version where appropriate and the data was normalised. The Bayesian statistics were generated in limma and tables of the highest differentially expressed genes were generated.

The BioConductor package limma is an excellent tool for both quality control and normalisation of data. A drawback, however, is the difficulty in retrieving the manipulated data for analysis in other software packages. The GUI (Graphical User Interface) version of limma handles data and performs in the same way as the command driven version though with more restricted options. However, it is very easy to extract the normalised M and A data to Excel files from which it is easy to calculate normalised, \log_2 red and green intensity values which can be used in other tools such as the online NIA Array Analysis tool (<http://stemcell.biosi.cf.ac.uk/>). This suite of tools can be used to carry out ANOVA and produce lists of differentially regulated genes as well as hierarchical clustering.

6.1.4. Characteristics of the Embryonic Stages.

The collection of the embryonic material for the experimental comparisons was carried out such that the groups were not only collected by the estimated time of development but by morphological landmarks (Figure 6.3, 6.4, 6.5 and 6.6). The blastocysts harvested between 86 and 86.5h.p.c. were selected on the basis that they had a blastocoel and retained the zona pellucida. The initiation of fluid accumulation begins towards the end of the fifth cleavage division when the embryo has 27-29 cells (Smith and McLaren, 1977). All morulae, defined as having no fluid accumulating cells, all

very early blastocysts, defined as having a blastocoel cavity of less than half the volume of the blastocyst and all hatched blastocysts, defined by having no zona pellucida were discarded. Thus the majority of ICMs in this group would not have primitive endoderm on the blastocoelic surface. Furthermore, if explanted into culture they would regenerate TE or if injected into a host blastocyst would be expected to be capable of contributing to all tissues of the conceptus, including the trophectoderm (Handyside, 1978; Hogan and Tilly, 1978).

The blastocysts harvested between 103 and 107h.p.c. were only selected when hatched. Some of the embryos in this group, particularly from the later harvest times had evidence of endometrial cells adhering to the TE indicating the initiation of implantation. Isolated ICMs from this group would be expected to have lost the ability to form trophectoderm both *in vitro* and *in vivo*. They would also be expected to have primitive endoderm and perhaps some parietal endoderm starting to migrate around the inner surface of the mural trophectoderm (Enders, 1971; Nadijcka and Hillman, 1974; Snell and Stevens, 1966). No attempt was made to remove the endoderm from these ICMs after immunosurgery. Immunosurgery was used to isolate the ICM from blastocysts.

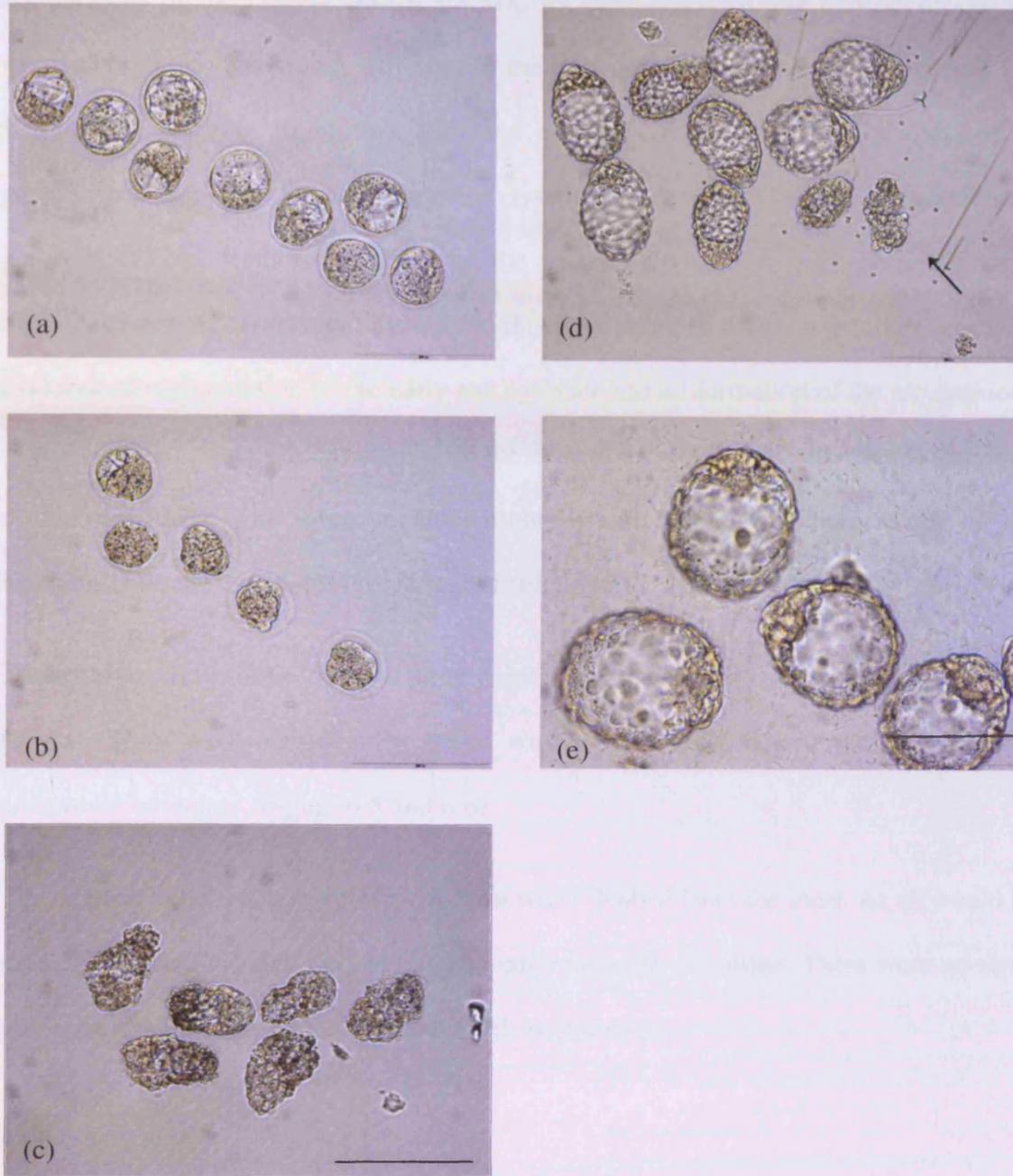


Figure 6.3: Examples of embryos from which ICMs collected (a) early blastocysts flushed 88hpc (b) morulae and early blastocysts from the same group as (a) but rejected from immunosurgery. (c) late blastocysts flushed 105hpc (d) delayed blastocysts 136hpc, arrow indicates collapsed blastocyst this would re-expand in culture (e) delayed blastocysts 180hpc Bar is 200mm.

As the cells of the trophectoderm are held by tight junctions, the blastocysts can be exposed to antibodies which will bind to the trophectoderm but be unable to reach the ICM. After washing to remove unbound antibody, the blastocysts are exposed to complement and the trophectoderm is lysed leaving the ICM intact (Solter and Knowles, 1975). Embryos isolated at the 132-135h.p.c. time point were the most difficult to detect, harvest and dissect. At this time point, in some cases, there was little evidence of segmentation of the early egg cylinder and no formation of the pro-amniotic cavity (pac) so some contamination of the embryonic ectoderm with extra-embryonic ectoderm is likely. This group included embryos with an early pac but without the pac extending into the extra-embryonic region (Figure 6.4).

The embryos at the 150-154h.p.c. time point were both easy to see in the uterus and to harvest. They were significantly larger with a well defined embryonic and extra-embryonic boundary (Figure 6.5 and 6.6).

The delayed blastocysts were very uniform when flushed from the uteri. As all would be expected to have hatched, any unhatched embryos were discarded. There were no signs of implantation having initiated in any of these embryos.

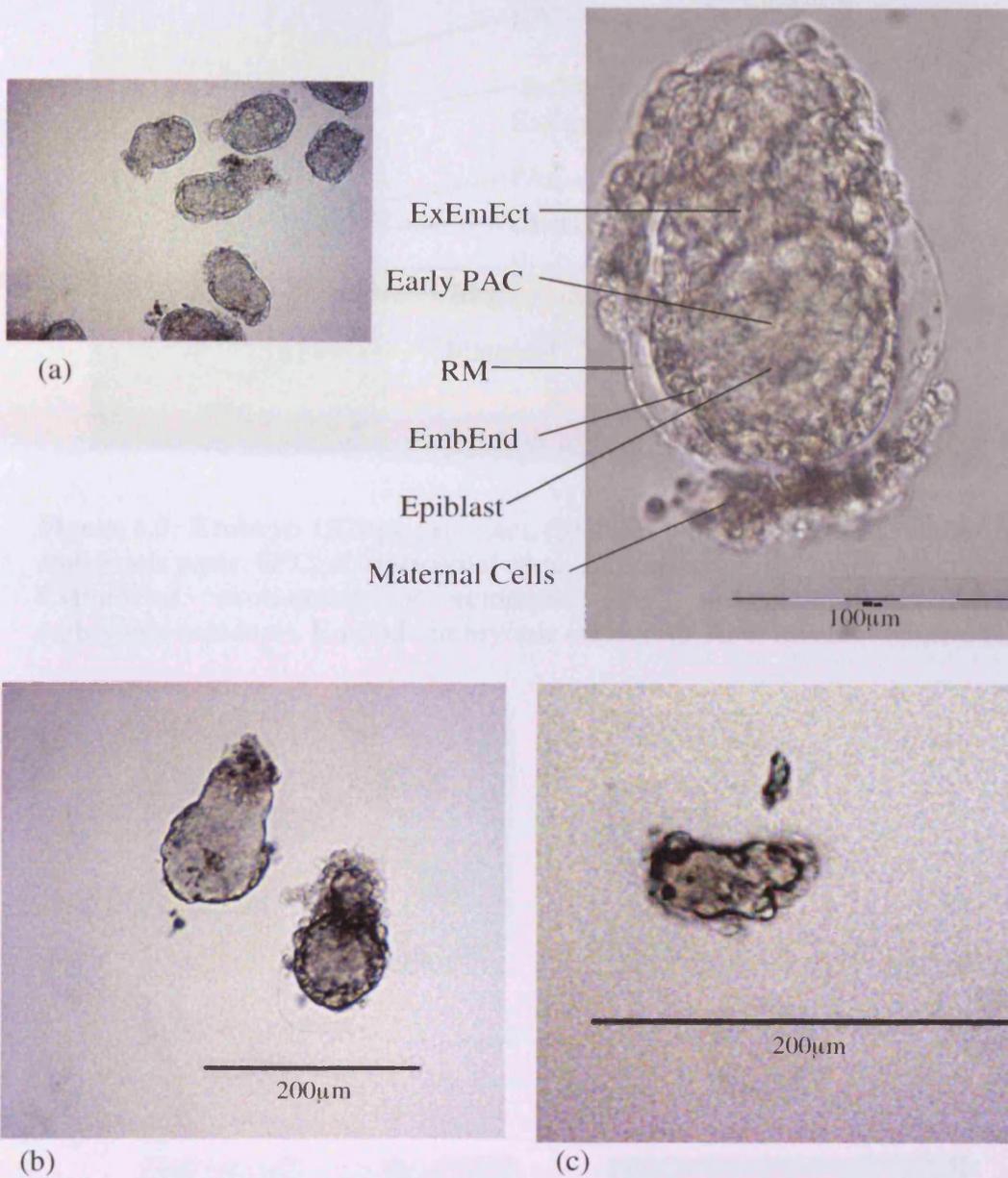


Figure 6.4: Examples of the isolation of the embryonic ectoderm from 130h.p.c. embryos. The top row (a) shows a sample of intact embryos. Reichert's membrane was removed with tungsten needles where necessary. The embryo was then incubated in pancreatin-trypsin and the endoderm peeled off (b) left hand embryo. (c) The embryonic portion was then isolated with tungsten needles. ExEmEct:extra-embryonic ectoderm, PAC: pro-amniotic cavity, EmEnd: embryonic endoderm, RM: Reichert's membrane.

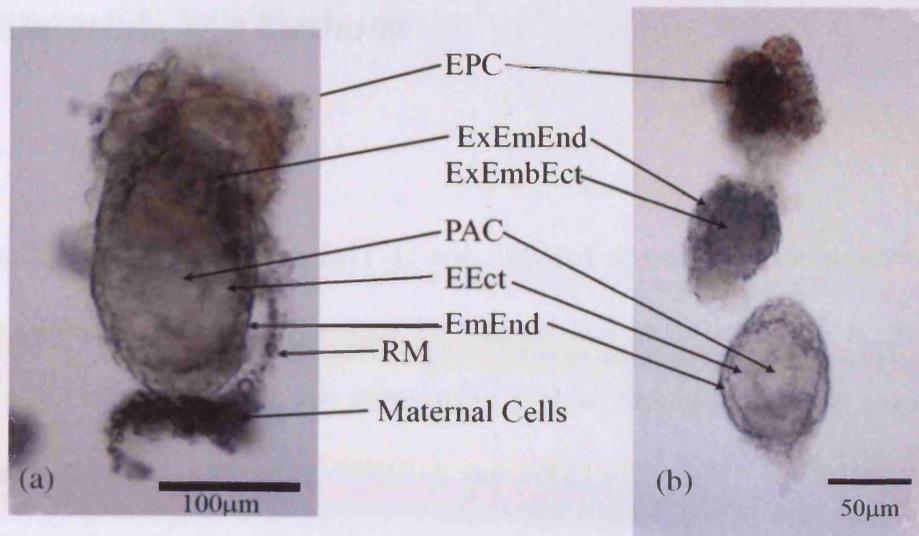


Figure 6.5: Embryo 152hpc (a) intact, (b) dissected into embryonic and extra-embryonic parts. EPC: ectoplacental cone, ExEmEnd: extra-embryonic endoderm, ExEmbEct: extra-embryonic ectoderm, PAC: pro-amniotic cavity, EEct: embryonic ectoderm, EmEnd: embryonic endoderm, RM: Reichert's membrane.

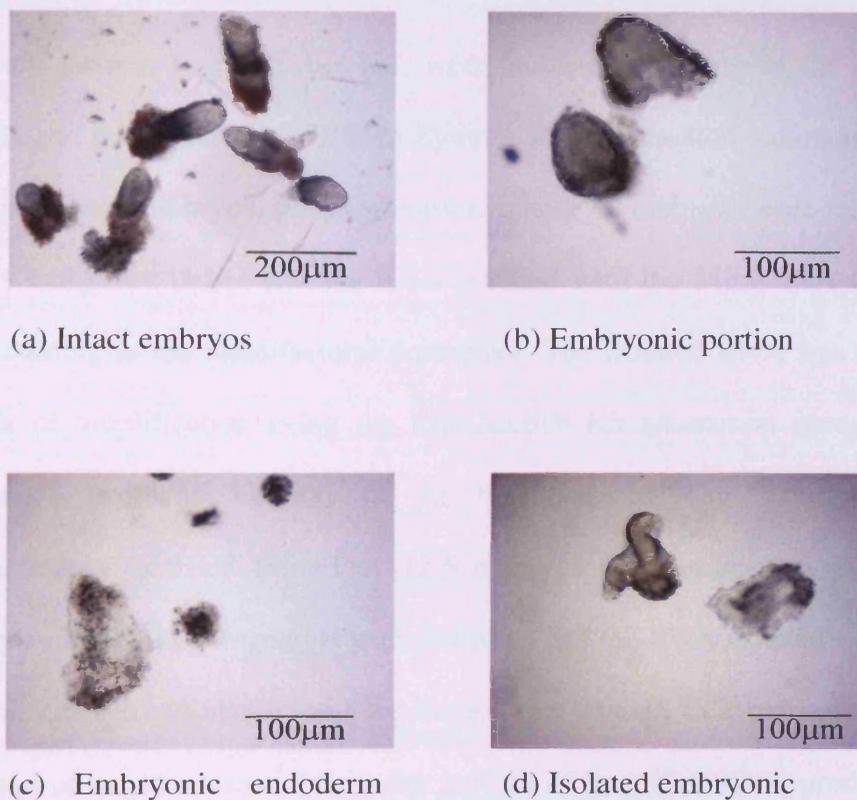


Figure 6.6: Example of the isolation of the embryonic ectoderm from 152hpc embryos. First Reichert's membrane is removed with tungsten needles and the embryonic portion (b) is incubated in pancreatin-trypsin. The embryonic endoderm, (c) is then peeled off leaving the embryonic ectoderm (d).

6.2. Materials and Methods

6.2.1. Cell Culture.

The embryonic stem cell line IMT-11 was cultured according to the protocol in Section 2.1. A plate of IMT-11 at passage 16 was washed twice in DPBS-A, groups of ten colonies were picked from the plate using a glass capillary, put into microcentrifuge tubes in a minimal volume of DPBS-A and 200µl of Zymo™ RNA Extraction Buffer was added. Samples were stored at -70°C until extraction. RNA was purified according to the manufacturer's protocol (2.3.1b).

6.2.2. Reference Pool

Embryos and tissues, as per Table 6.1, were isolated according to the protocols in Section 2.2a and b and stored at -70°C in Zymo™ RNA extraction buffer. In the case of the pre-implantation embryos, pools of approximately 40 embryos were made in 200µl of Zymo™ extraction buffer and the RNA purified with the Mini RNA Isolation Kit (Zymo) according to the manufacturer's protocol. The isolated RNA was subjected to two rounds of amplification using the RiboAmp™ Kit (Arcturus) according to the manufacturer's protocol Version C. In the case of day 7.5 embryos, the ovary/mesonephros isolated from Day 12.5 embryos and newborn ovary, groups of eight embryos and tissues fragments were collected and the RNA isolated as for the pre-implantation embryos. When amplifying these larger tissues, to avoid saturation of the second round of *in vitro* transcription, the amount of amplified RNA produced by the first round of amplification was measured by spectrophotometry. The second round of amplification was then initiated with 0.5µg of amplified RNA from the first round of amplification.

Tissue	Day of Gestation	Harvested hpc
oocytes	0.5	7-7.5
1 cell	0.5	7.5-8.5
2 cell	1.5	32-35
3-4 cell	2.0	45-46.5
4-8 cell	2.5	55.5-58
8-16 cell	3	67-68
Late morulae-early blastocysts	3.5	80-84
Whole embryo, mid-streak stage	7.5	
Ovary+mesonephros	12.5	
Ovary	Newborn	

Table 6.1: Composition of the common pool.

Three separate pools, each composed of 5µg of amplified RNA from each tissue sample, were made and a 5µg sample from each pool was labelled with Cy5™. A 1µl sample of each labelled pool sample was run on a slide gel (Section 2.3.5). A grand pool made up of 50µl each of the labelled pools Pool 1, 2 and 3 was made. A 15pmol sample of each Pool 1, 2, 3 and the grand pool was combined with 15pmol of Cy3™ labelled IMT-11 ES cell amplified RNA. The volume adjusted, combined samples were heated with human COT1 DNA (Gibco BRL) and Poly-dA³⁰ (MWG Biotech) oligoA before adding an equal volume of two times concentrated hybridisation buffer (2.4.3). The samples were then applied to blocked microarray slides (2.4.2) and covered with Hybrislips (Sigma) (2.4.3). Hybridisation was performed at 42°C overnight followed by washing as in Section 2.4.4. Slides were then scanned with the ScanArray® Express HT (Perkin Elmer) with 65pmt, 5µm resolution and half speed. After ensuring the individual pools labelled well and hybridised to the slides, the amplified RNA was pooled, the volume adjusted and random hexamers were added to a final concentration of 100µM. This was then stored in aliquots ready for labelling. Labelled (Cy5™) aliquots were pooled and redistributed in 15pmol aliquots (a single hybridisation) and

stored frozen at -20°C ready for combination with the labelled experimental samples (Cy3TM). All samples were labelled using the method developed in Chapter 3.

6.2.3 Collection of Tissues

Embryos were harvested from animals and tissues isolated according to the protocols in Section 2.2 (Table 6.2.).

Tissue	Embryo Stage	Time hpc
ICM	Day 3.5	86-86.5
ICM	Day4.5	103-107
Delayed ICM	EDG6	134-138
Delayed ICM	EDG8	179-180.75
Embryonic epiblast	Day5.25	132-135
Embryonic ectoderm	Day6.5	150-154

Table 6.2: Tissue samples being compared by microarray

6.3. Results and Discussion

6.3.1 ES Cells as Reference

As part of a preliminary experiment, the amplified RNA from colonies of the ES cell line IMT-11 at passage 16 was hybridised according to the plan in Table 6.3. For this experiment the slides were scanned with the LSIV scanner (Genomic Solutions) and analysed with ImaGeneTM (BioDiscovery). The mean foreground and background intensities were exported to Excel (Microsoft) and the number of spots with foreground intensity greater than the mean local background intensity plus two standard deviations (SD) was counted. Since the Cy3^{TMTM} and Cy5^{TMTM} labelled samples of IMT-11 were pooled after labelling 5 μg aliquots of amplified RNA, and the developmental stages are six different tissues, the SD of the number of spots for IMT-11 should be lower than for the developmental stage tissues and this is in fact the case.

Slide Number	Cy3™	Cy5™
1	IMT-11	ICM ~86hpc
2	IMT-11	ICM ~86hpc
3	IMT-11	ICM ~105hpc
4	IMT-11	ICM ~105hpc
5	IMT-11	Embryonic epiblast ~133hpc
6	IMT-11	Embryonic epiblast ~133hpc
7	IMT-11	Embryonic ectoderm ~152hpc
8	IMT-11	Embryonic ectoderm ~152hpc
9	IMT-11	Delayed ICM ~136hpc
10	IMT-11	Delayed ICM ~136hpc
11	IMT-11	Delayed ICM ~180hpc
12	IMT-11	Delayed ICM ~180hpc
13	ICM ~86hpc	IMT-11
14	ICM ~86hpc	IMT-11
15	ICM ~105hpc	IMT-11
16	ICM ~105hpc	IMT-11
17	Embryonic epiblast ~133hpc	IMT-11
18	Embryonic epiblast ~133hpc	IMT-11
19	Embryonic ectoderm ~152hpc	IMT-11
20	Embryonic ectoderm ~152hpc	IMT-11
21	Delayed ICM ~136hpc	IMT-11
22	Delayed ICM ~136hpc	IMT-11

Table 6.3: Hybridisation plan of ES cells compared with ectoderm from various developmental stages.

Although the number of spots detected as two standard deviations above the mean background fluorescence intensity on these slides is very low, there is no evidence that ES cells hybridise with more spots than any of the tissues tested. Student's two tailed t-tests for heteroscedastic samples (samples of unequal variance) were carried out. The only slide set with a significant difference ($p < 0.05$), more spots were detected in the developmental stages.

	ES cells Cy3™	Developmental Stages Cy5™	ES cells Cy5™	Developmental Stages Cy3™	ES cells combined	Developmental Stages combined
No of Spots	1392.25	2859.92	1489.80	973.40	1436.59	2002.41
sd	976.84	1659.01	841.43	536.26	919.05	1585.68
n, p	n=12, p=0.02		n=10, p=0.14		n=22, p=0.17	

Table 6.4: Mean number of spots per slide above background fluorescence intensity+2sd.

An obvious problem with this experiment is the very low number of spots detected. There are 15,240 data spots on this particular slide batch and at best only 18.76% of the spots were detected as the mean+ 2SD above the global background. This was probably due to the batch of slides being of sub-optimal quality.

6.3.2 Pooled Embryonic Tissues as Reference

A further experiment was initiated with a new array slide set. The slides used in this section, which compare the ES cells to the common pool, are part of the larger experiment described in Section 6.3.4. They have been used in this section used to ask the question as to whether the ES cells are suitable as a reference sample and yet again in Section 6.6.3 to compare ImaGene™ with Spot.

Table 6.5 shows the number of spots detected in the two channels of the ES cells against the array. This slide set has 15248 data spots on each slide. In these calculations the poor quality spots were removed from the data. This is also a conservative estimate of spots above background as local background was used as the estimate of background. Thus 87.56% and 88.81% of spots on the slides were hybridised to the target from ES cells and the pooled reference respectively. This suggests that ES cells may indeed

make a good universal reference at least for amplified material where very low expression levels are more reliably detected (Polacek *et al.*, 2003) The embryonic stem cells may only make a good reference with embryonic genes.

	ES cells Cy3™	Pool Cy5™
Mean Number of Spots	13351.17	13541.33
SD(n=6)	387.83	369.22

Table 6.5: Mean number of good spots per slide detected above background fluorescence+2sd.

6.6.3 Comparison of *ImaGene*™ and *Spot*

In this study, much of the image analysis was carried out using *ImaGene*™. However, it became clear as more data was generated; an image analysis program which required less user intervention was necessary. The image analysis program *Spot*, not only uses the seeded growth method of spot detection and segmentation, which has the advantage of being able to segment irregular spots, but also can be used to process a whole batch of slides which for our printer is fifty slides (Figure 6.7)

Since each slide is analysed individually more “between slide” variation in data may be acquired from *ImaGene*™ analysis, and consequently noisier data might be expected. Furthermore, *ImaGene*™ files are handled differently from *Spot* files in *limma* in particular with regard to background correction. A set of slides hybridised with IMT-11, labelled with Cy3™ and the reference pool labelled with Cy5™, were analysed using both *ImaGene*™ and *Spot*. The data were then scrutinized for discrepancies.

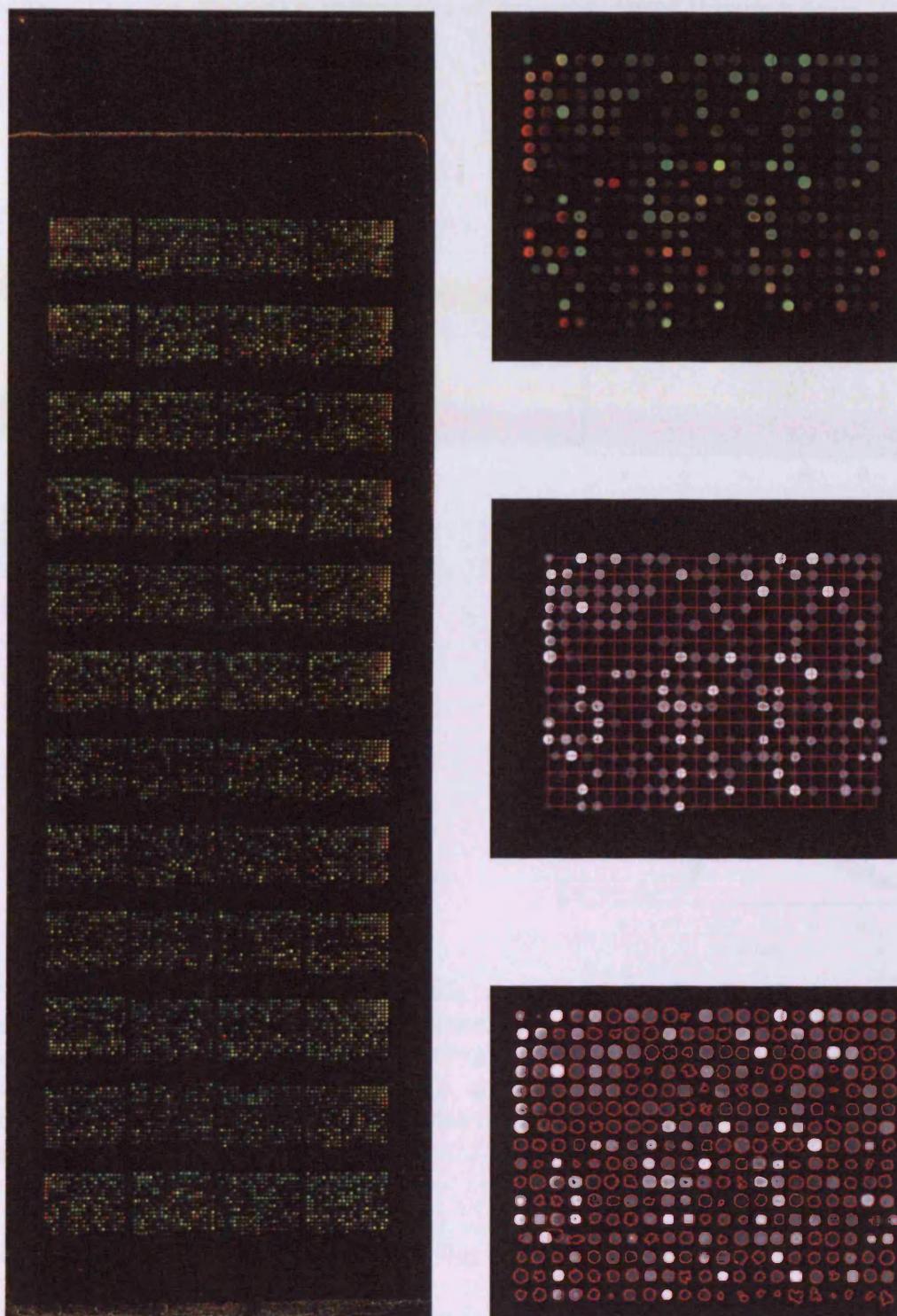


Figure 6.7: Example of a microarray analysed with Spot (CISRO) Left is whole array and right column is one example grid. Top unprocessed, middle with addressing grid applied and bottom segmented. In the pseudo colour images red is higher expression in the Cy5™, green is higher in Cy3™ and yellow is equal in both channels.

The first difference detected is in the range of intensities found (Figure 6.8).

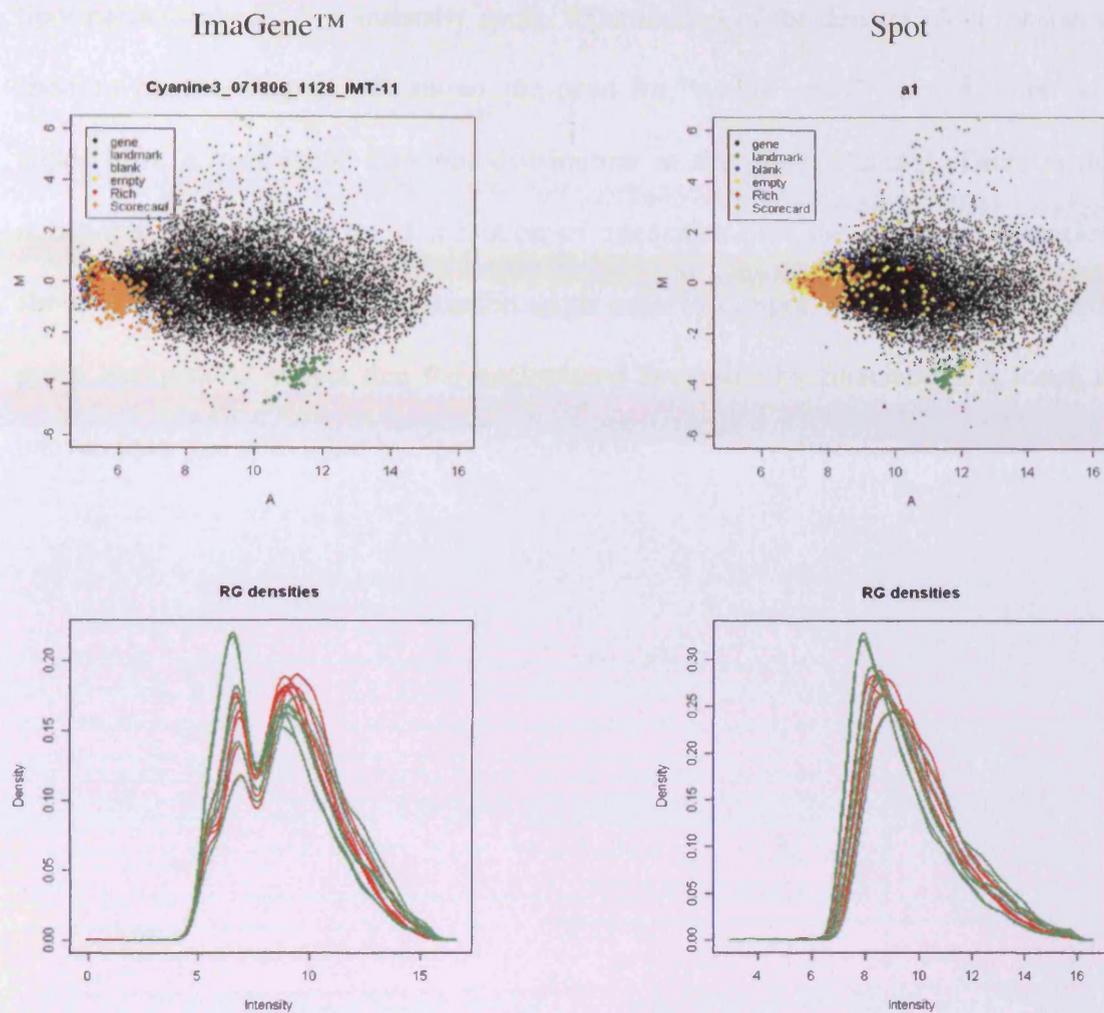


Figure 6.8: MA plot of sample slide and density plots for the raw \log_2 transformed red and green intensities. Left column ImaGene™ and right column Spot. M is $\log_2\text{Cy5}^{\text{TM}} - \log_2\text{Cy3}^{\text{TM}}$ i.e. the fold change for each spot and A is the $0.5(\log_2\text{Cy5}^{\text{TM}} + \log_2\text{Cy3}^{\text{TM}})$ i.e. the total intensity for each spot. The density plots show that two of the slides have a slightly different distribution of intensity in the green (Cy3^{TM}) channel

Examination of the MA plot of the raw \log_2 red and green intensity values shows that the lowest intensity spot was greater than 6 when analysed with Spot and less than 5 when analysed with ImaGene™ suggesting greater discrimination at lower spot intensity for the program Spot. This is not surprising as the histogram method of segmentation uses the mean pixel values between two high percentiles, usually the 80th and 95th. The MA plots of the ImaGene™ and Spot red and green intensities (top

graphs, Figure 6.8) also show the ImaGene™ data does have greater variance than for Spot particularly for low intensity spots. Examination of the density plots for this data, (bottom graphs, Figure 6.8) shows the need for “within array” normalisation as two slides have a noticeably different distribution in the green channel. There is also a noticeable difference in the distribution of intensities over the slides with ImaGene™ showing a double peak rather than the single peak from Spot. Image plots of the red and green background shows that the background generated by ImaGene™ is much more uneven than that generated by Spot (Figure 6.9).

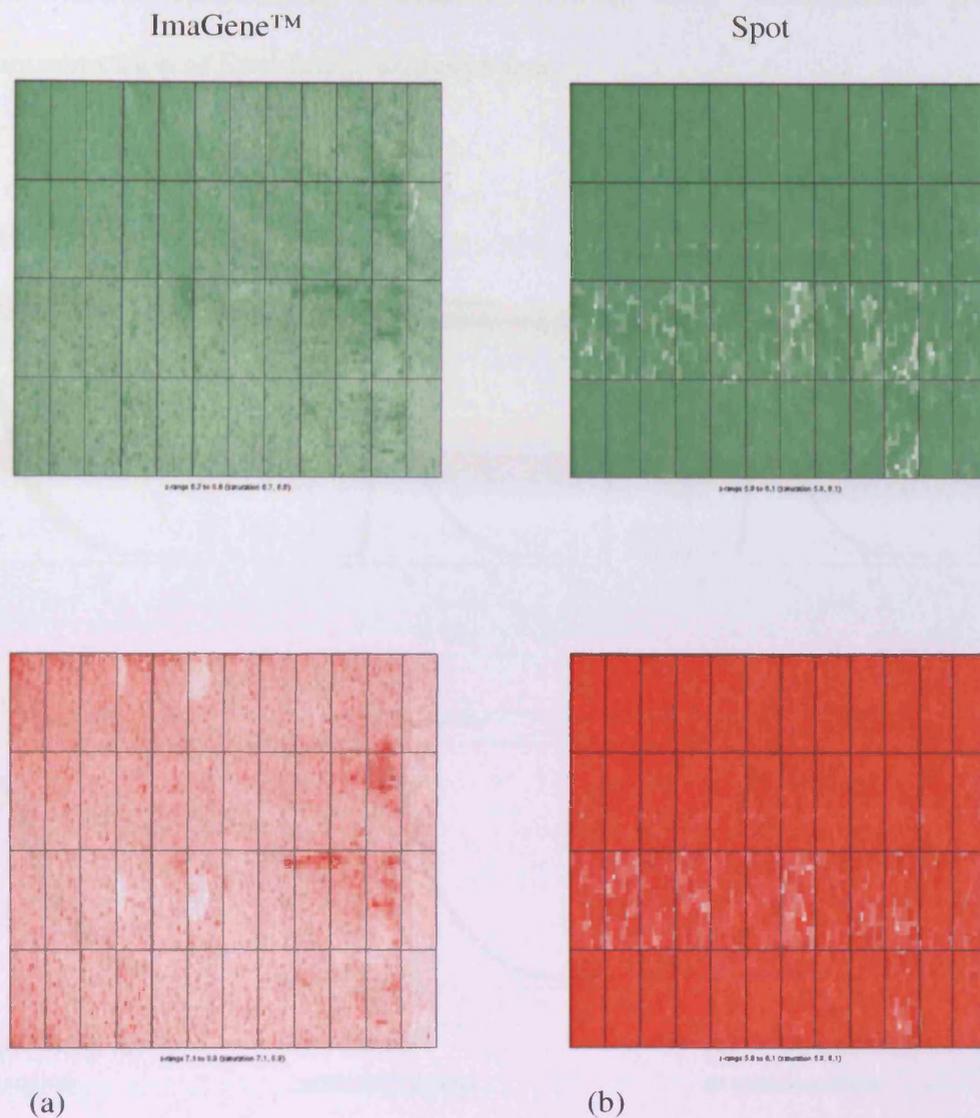


Figure 6.9: Image plots of red and green background for a slide analysed with (a) ImaGene™ and (b) Spot. These are regenerated images of the background signal.

Examination of the density plots of the within slide normalised data for Spot, which have not been background subtracted (Figure 6.10) show that these slides require very little “between array” normalisation. Unfortunately, it was not possible to put these plots on the same scale making comparisons more difficult. Thus it was not possible to say if background subtraction increased the variance between slides, a phenomenon often seen (Yang *et al.*, 2002a). However it appears that the simple subtraction of morph

background function followed by A-quantile “between array” normalisation gives excellent normalisation of Spot image analysed data.

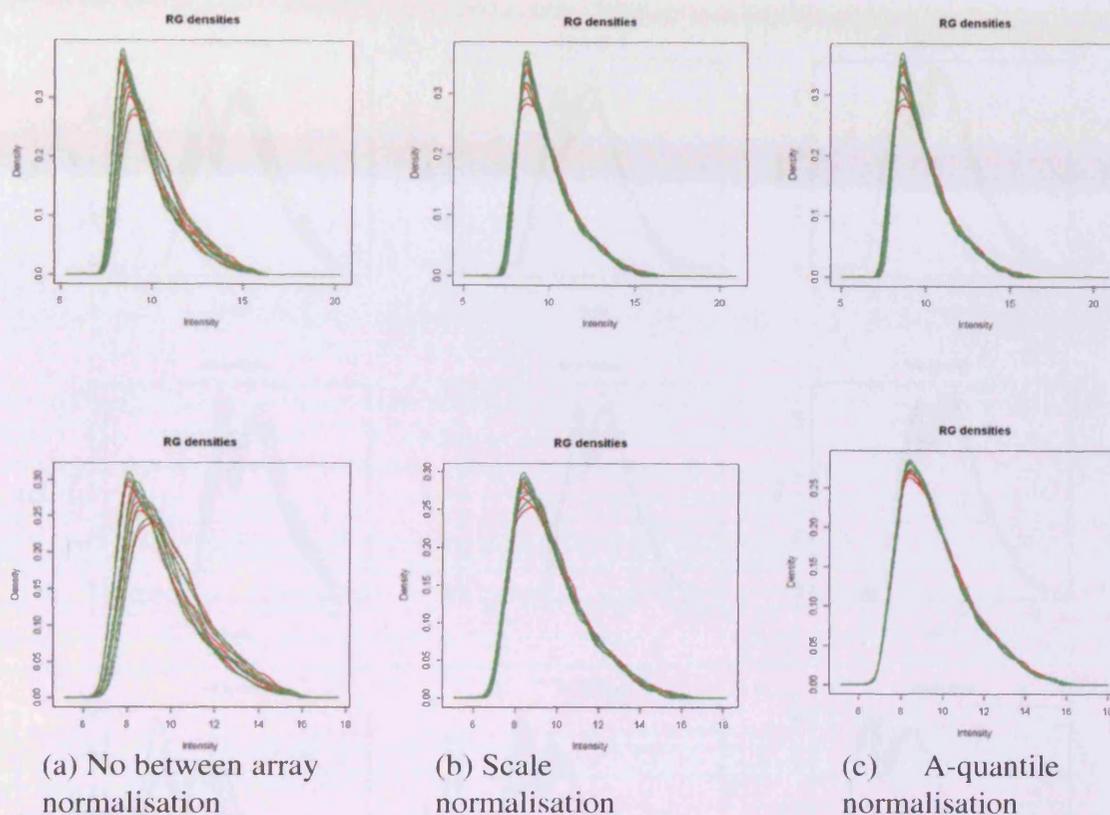


Figure 6.10: Density plots of data with image analysis using Spot. The top row shows the effect of (a) no between array normalisation, (b) scale between array normalisation and (c) A-quantile between array normalisation on data which has not had any background removed before print tip loess “within array” normalisation. The bottom row has had the Spot generated value “morph” removed before print tip loess normalisation within the arrays followed by between array normalisation as for the top row.

Examination of the ImaGene “within array” normalised data (Figure 6.11) suggests that background subtraction in this case does increase the between slide variances and magnifies anomalies in intensity distribution.

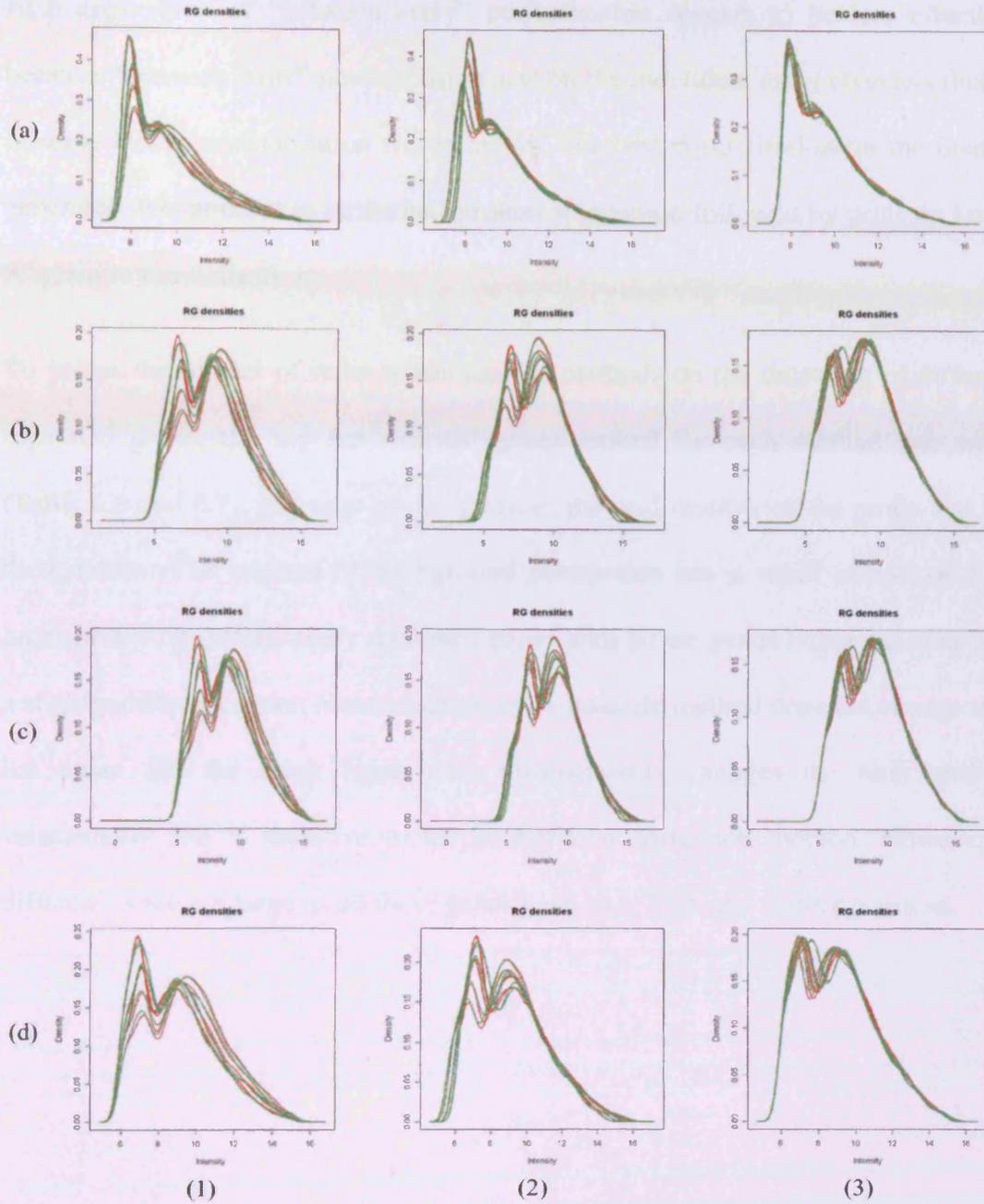


Figure 6.11: Density plots of the data obtained from the image analysis program ImaGene™ Column1: no between array normalisation. Column2: scale between array normalisation. Column3: A-quantile between array normalisation. Row (a): no background subtraction, Row (b): “subtract”. Row(c): “half” background subtraction. Row (d): “normexp, offset=50” background subtraction.

Also application of “between array” normalisation appears to be less effective and because “between array” normalisation acts on the individual array channels this results in some loss of normalisation within arrays. The best normalisation for the ImaGene™ generated data appears to be no background subtraction followed by print-tip Loess and A-quantile normalisation.

To assess the impact of these normalisation methods on the detection of differentially regulated genes, the top ten returned genes ranked for each method was examined (Table 6.6 and 6.7). For ease of comparison, the grid number of the probe was used in these tables. The method of background subtraction has a small impact on the rank order of the top differentially regulated genes with all ten genes being the same albeit in a slightly different order. Normalisation by A-quantile method does not change the gene list order. On the other hand scale normalisation changes the rank order more substantially and is sensitive to the background correction method. However, these differences are not large as all these genes have very high and similar p values.

Rank	Probe Number											
1	8162	6997	8162	8162	2601	8162	8162	2601	8162	8162	8889	8162
2	13651	4293	13651	1485	8889	1485	1485	8889	1485	5639	2601	5639
3	5639	5088	5639	8278	8162	8278	8278	8162	8278	1485	6997	1485
4	6866	15761	6866	15296	3946	15296	15296	3946	15296	8278	7985	8278
5	15457	7985	15457	5639	1485	5639	5639	1485	5639	3946	8162	3946
6	3946	8889	3946	3946	6997	3946	3946	6997	3946	15263	1485	15263
7	91	6897	91	9914	10567	9914	9914	10567	9914	9914	4293	9914
8	2547	6993	2547	15263	12959	15263	15263	12959	15263	15296	15263	15296
9	15263	2783	15263	14937	7985	14937	14937	7985	14937	13651	9337	13651
10	13441	2927	13441	2601	15296	2601	2601	15296	2601	2601	3946	2601
Background correction	none			subtract			half			normexp		
Between Array correction	none	scale	Aquantile	none	scale	Aqantile	none	scale	Aquantile	none	scale	Aquantile

Table 6.6: Top ten differentially expressed genes from slides analysed with ImaGene™.

Rank	Probe Number					
1	8162	8889	8162	8162	8889	8162
2	5639	6534	5639	5639	6534	5639
3	15263	2601	15263	13651	8162	13651
4	15687	15761	15687	15263	2601	15263
5	13651	13455	13651	15687	2547	15687
6	1257	8162	1257	6866	6972	6866
7	2547	1485	2547	1257	1485	1257
8	5643	2547	5643	4211	12959	4211
9	6866	2819	6866	5643	8684	5643
10	4211	6972	4211	2547	9521	2547
Background Correction	none			morph		
“between array” Normalisation	none	scale	A-quantile	none	scale	A-quantile

Table 6.7: Top ten differentially expressed genes from Spot analysed slides. “Morph” is the method of smoothing the background variances before subtraction used by Spot.

The rank order by p-value for the slides analysed with ImaGene™ is significantly different from the order found after Spot analysis. When the returned genes for Spot and ImaGene™ which have been analysed by the same method are compared (morph in Spot is the same as subtract in ImaGene as both simply subtract the background as is measured by the image analysis program) the number of genes in common (Table 6.8) are at best four out of ten.

Background correction	None			Subtract/Morph		
Between array correction	None	Scale	A quantile	None	Scale	A quantile
Common genes	4	2	4	1	4	1

Table 6.8: Number of common genes in top ten differentially expressed genes comparing Spot and ImaGene™ image analysis.

Both data sets, irrespective of the image analysis program used, show that background correction has an impact on the top ten genes returned and that this impact is greater in slides analysed using ImaGene™. Furthermore, scale “between array” normalisation changes the top ten returned genes irrespective of the method of background correction while the lists returned when either no “between array” or A-quantile “between array” normalisation was applied are identical.

When the genes lists for the ImaGene™ slides analysed using no background subtraction and A-quantile “between array” normalisation are compared to Spot analysed data using morph background correction and A-quantile “between array” normalisation, the methods which give the best normalisation by examination of the density distribution, the number of common genes rises to six.

The Spot and ImaGene™ data were reanalysed using limmaGUI. In both cases they were analysed in the sequence: background subtract by the method “subtract”, “within array” normalisation by the method print-tip loess, and “between array” normalisation by the method A-quantile. The M and A values were then exported to Excel(Microsoft). The green and red intensities were calculated using the formulae below (Box1). The derivation of the formulae is presented in Appendix 4.

$$\text{Log}_2 \text{Cy3}^{\text{TM}} \text{ Intensity} = A - M/2$$

$$\text{Log}_2 \text{Cy5}^{\text{TM}} \text{ Intensity} = A + M/2$$

These intensities were put into a format suitable for the NIA Array Analysis tool and loaded. The data was analysed using the maximum of averaged and actual error variance model which is the most conservative reducing the number of false positives.

The standard deviation should vary with the log intensity and for the Spot data does but shows a peak with the ImaGene™ data (Figure 6.12). The correlation matrix of intensities (Figure 6.13) shows a slightly better correlation between repeats with ImaGene™ than Spot although the correlation in both cases is high. The data presented here suggests that Spot performs well at segmenting the images in these slides. It does suffer from the disadvantage that the individual spots cannot be adjusted so some losses do occur (Figure 6.6 bottom right panel). However, the measurement of background for Spot is much less noisy and results in less loss of data.

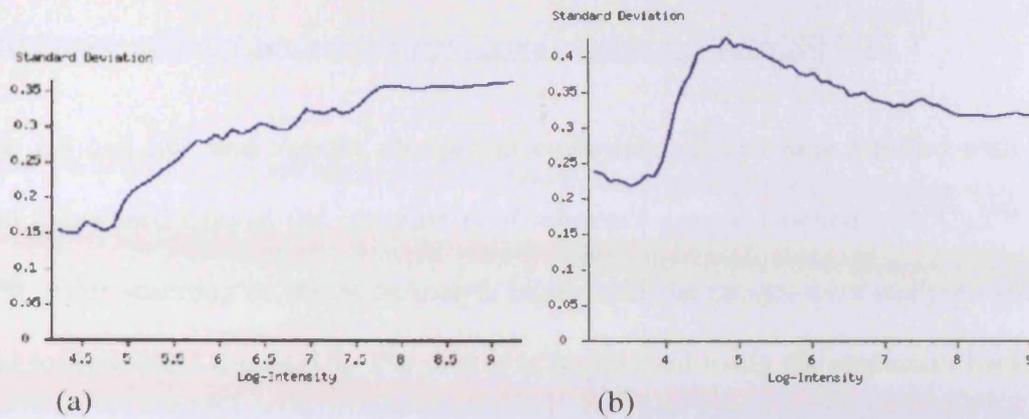


Figure 6.12: Error plot for (a) Spot and (b) ImaGene™ analysed slides. Log-intensity (x-axis) is plotted against the Standard Deviation (y-axis)

	1	2	3	4	5	6
1. IMT_11:Rep1	1	0.941	0.896	0.901	0.884	0.887
2. IMT_11:Rep2	0.941	1	0.929	0.935	0.911	0.911
3. IMT_11:Rep3	0.896	0.929	1	0.963	0.899	0.898
4. IMT_11:Rep4	0.901	0.935	0.963	1	0.907	0.906
5. IMT_11:Rep5	0.884	0.911	0.899	0.907	1	0.972
6. IMT_11:Rep6	0.887	0.911	0.898	0.906	0.972	1

(a)

	1	2	3	4	5	6
1. IMT-11:Rep1	1	0.938	0.909	0.913	0.899	0.896
2. IMT-11:Rep2	0.938	1	0.926	0.933	0.911	0.907
3. IMT-11:Rep3	0.909	0.926	1	0.954	0.907	0.904
4. IMT-11:Rep4	0.913	0.933	0.954	1	0.912	0.909
5. IMT-11:Rep5	0.899	0.911	0.907	0.912	1	0.961
6. IMT-11:Rep6	0.896	0.907	0.904	0.909	0.961	1

(b)

Figure 6.13: Correlation matrix of intensities for (a) Spot and (b) ImaGene™. The darkest red indicates the highest correlation, 1 being the maximum, shading orange for correlation coefficients between 0.900 and 0.999 and yellow for between 0.800 and 0.899.

6.3.4 Comparison of ES Cells with Pluripotent Embryonic Tissue

The data produced by this experiment has been lodged with GEO.

GEO Accession: GSE8881:

<http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE8881>

The ES cell line and various pluripotent embryonic tissues were labelled with Cy3™ and hybridised against the common pool reference sample labelled with Cy5™ (Table 6.9). After scanning on the ScanArray® ExpressHT the images were analysed with Spot and loaded into LimmaGUI. The data was normalised using the sequence: background subtract by method subtract, “within array” normalise with print-tip loess and “between array” normalise with the method A-quantile. The normalised M and A values were exported to Excel (Microsoft) and converted to normalised log₂ red and green intensities as before (6.3.3). This manipulated data was then input to the NIA Microarray Analysis tool and analysed. Examination of the correlation matrix suggested that some of the slides were not good enough and visual inspection showed them to be smeared. These were eliminated from the analysis namely slides 11, 12, 30 and 42. However, the correlation coefficients for each tissue were very high. The most difficult tissue to isolate and dissect was the Day 5.5 epiblast and this might be expected to be the most variable, however the correlation coefficients for the biological repeats lay between 0.935 and 0.975 showing the degree of reproducibility in sample collection and processing was very high. Furthermore, this suggests a high degree of similarity in gene expression within the groups of embryos staged by morphological criteria.

A wealth of data has been generated by this experiment and is beyond the scope of this thesis to examine it all. The analysis examined here is aimed at investigating the homology of embryonic stem cells with the pluripotent embryonic portion of the conceptus.

Slide Number	Cy3™ labelled Tissue	Number in sample	Yield amplified RNA(µg)
1	IMT-11 colonies Sample1	10	
2	IMT-11 Sample1		
3	IMT-11 Sample2	10	
4	IMT-11 Sample2		
5	IMT-11 Sample3	10	
6	IMT-11 Sample3		
7	ICM 86hpc Sample1	25	43.9
8	ICM 86hpc Sample1		
9	ICM 86hpc Sample2	34	94.08
10	ICM 86hpc Sample2		
11	ICM 86hpc Sample3	27	60.24
12	ICM 86hpc Sample3		
13	ICM 105hpc Sample1	28	109.4
14	ICM 105hpc Sample1		
15	ICM 105hpc Sample2	35	107.16
16	ICM 105hpc Sample2		
17	ICM 105hpc Sample3	34	73.2
18	ICM 105hpc Sample3		
19	Delayed ICM 136hpc Sample 1	16	63.5
20	Delayed ICM 136hpc Sample 1		
21	Delayed ICM 136hpc Sample 2	33	84.06
22	Delayed ICM 136hpc Sample 2		
23	Delayed ICM 136hpc Sample 3	31	46.56
24	Delayed ICM 136hpc Sample 3		
25	Delayed ICM 180hpc Sample1	30	20.82
26	Delayed ICM 180hpc Sample1		
27	Delayed ICM 180hpc Sample2	26	21.36
28	Delayed ICM 180hpc Sample2		
29	Delayed ICM 180hpc Sample3	37	31.26
30	Delayed ICM 180hpc Sample3		
31	Embryonic epiblast Day5.5 Sample1	3	36.24
32	Embryonic epiblast Day5.5 Sample1		
33	Embryonic epiblast Day5.5 Sample2	5	27.72
34	Embryonic epiblast Day5.5 Sample2		
35	Embryonic epiblast Day5.5 Sample3	2	27.7
36	Embryonic epiblast Day5.5 Sample3		
37	Embryonic ectoderm Day6.5 Sample1	2	33.9
38	Embryonic ectoderm Day6.5 Sample1		
39	Embryonic ectoderm Day6.5 Sample2	3	107.28
40	Embryonic ectoderm Day6.5 Sample2		
41	Embryonic ectoderm Day6.5 Sample3	2	53.7
42	Embryonic ectoderm Day6.5 Sample3		

Table 6.9 continued: Plan of hybridisation. Highlighted slides were removed from the analysis.

Pairwise comparisons of the means were carried out and plots of log ratio against log intensity were made (Figure 6.14). With respect to the number of genes up and down regulated, the ES line IMT-11 is most similar to the epiblast portion of the day 5.5 embryo. Only 97 genes are detected as differentially regulated between the ES sample and this embryonic tissue. Also very interesting but not pertinent to this thesis is the similarity between the early (88h.p.c.) and late (105h.p.c.) ICM and the late ICM 105h.p.c. with the delayed ICM, 180h.p.c. The number of regulated genes in each pair wise combination is shown (Table 6.10)

Comparison	Number Up Regulated	Number Down Regulated	Total
IMT-11 v ICM 88hpc	1344	984	2382
IMT_11 v ICM 105hpc	2022	1807	3829
IMT-11 v delayed ICM 136hpc	530	307	837
IMT-11 v delayed ICM 180hpc	1708	1879	3587
IMT-11 v epiblast Day5.5	31	66	97
IMT-11 v embryonic ectoderm Day6.5	959	1124	2083

Table 6.10: Number of differentially regulated genes for the comparisons between the ES cells, IMT-11 and embryonic tissues.

Figures 6.15 to 6.20 present the data for each comparison of the ES cells against with the top ten up-regulated (in red) and down-regulated (in green) genes.

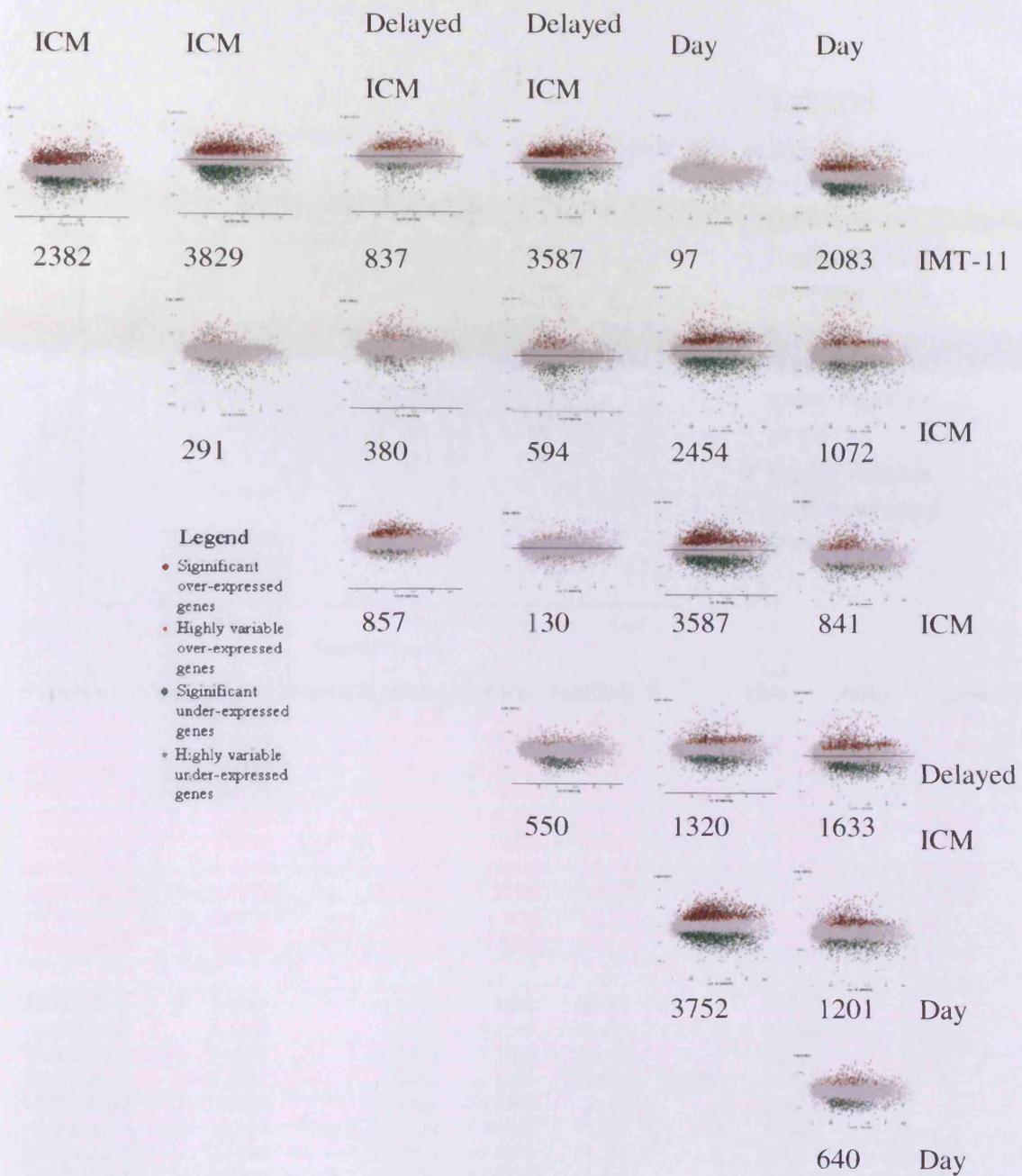


Figure 6.14: Plots of all pairwise comparisons extracted from ANOVA. The number under each graph is the numbers of regulated genes in each comparison.

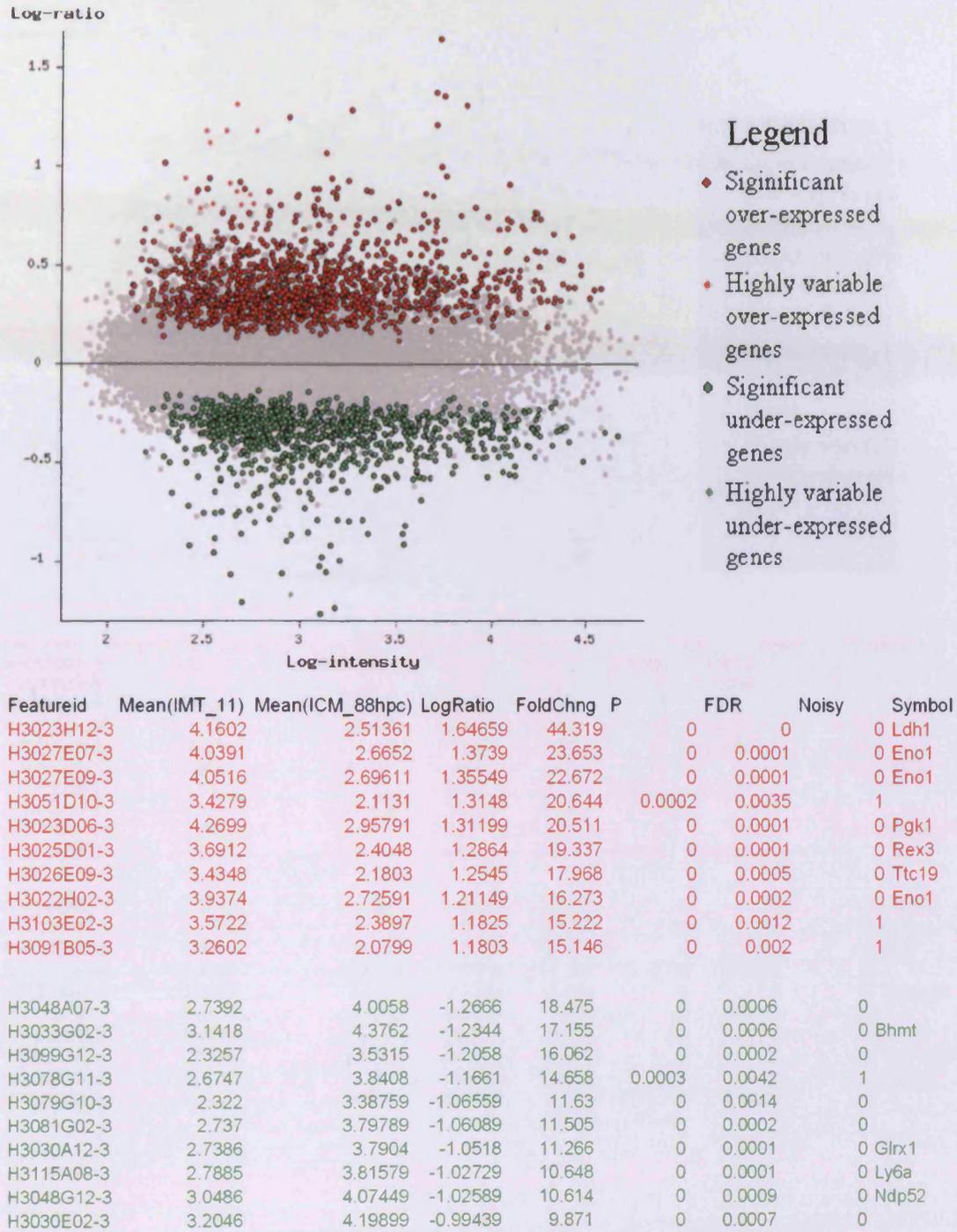


Figure 6.15: IMT-11 vs. ICM 88hpc. Top ten up regulated genes in red and down regulated genes in green below scatter plot.

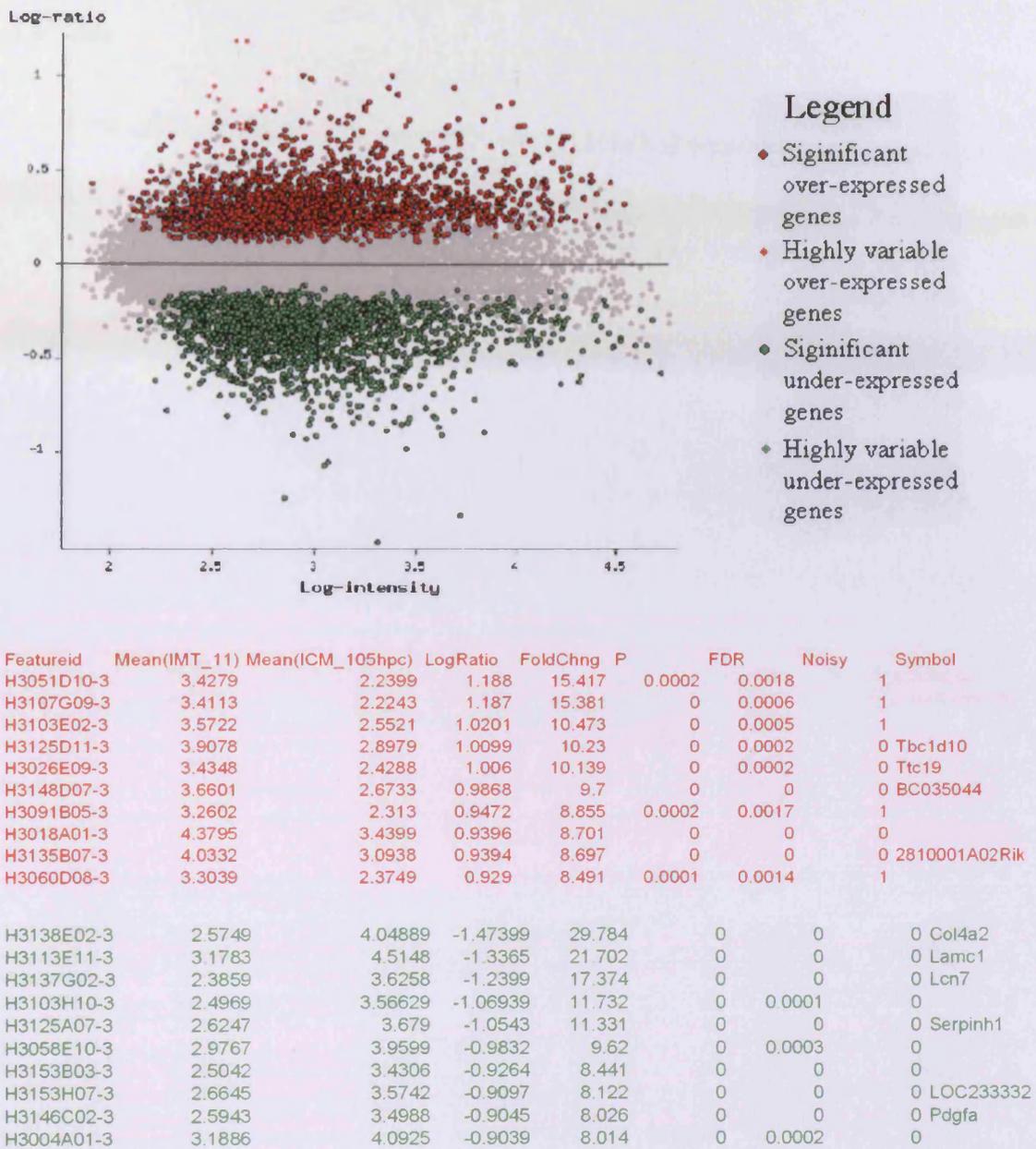


Figure 6.16: IMT-11 vs. ICM 105hpc. Top ten up regulated genes in red and down regulated genes in green below scatter plot.

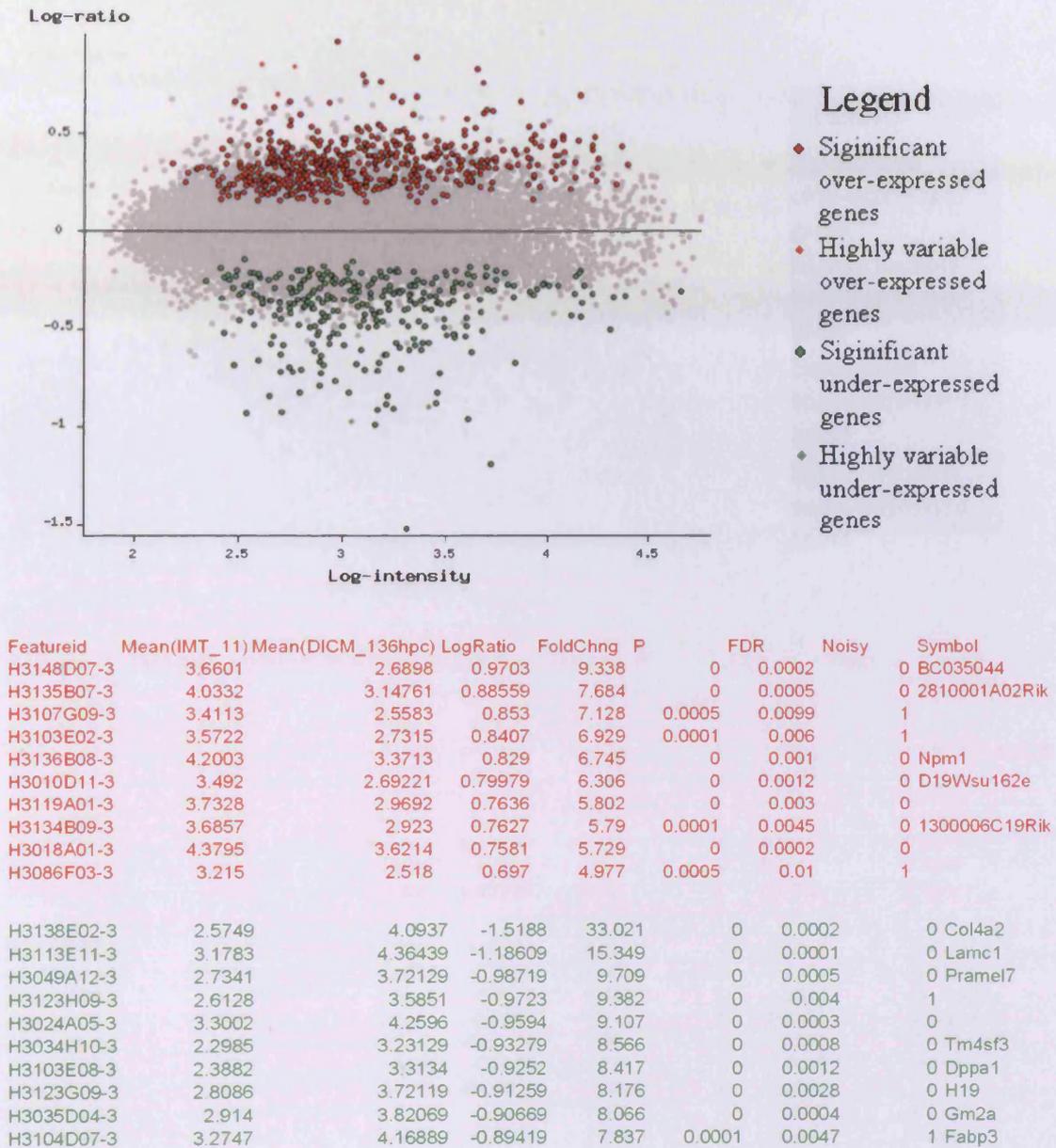
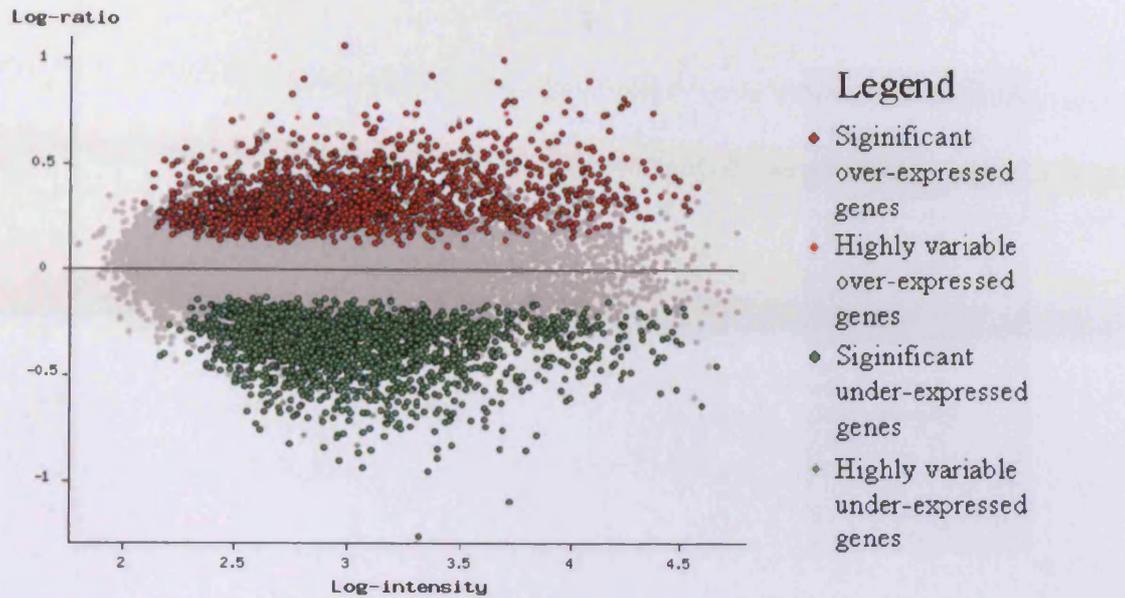


Figure 6.17: IMT-11 vs. Delayed ICM 136hpc. Top ten up regulated genes in red and down regulated genes in green below scatter plot.



Featureid	Mean(IMT_11)	Mean(DICM180hpc)	LogRatio	FoldChng	P	FDR	Noisy	Symbol
H3148D07-3	3.6601	2.59601	1.06409	11.59		0	0.0001	0 BC035044
H3051D10-3	3.4279	2.4197	1.0082	10.19	0.001	0	0.0058	1
H3018A01-3	4.3795	3.3834	0.9961	9.91	0	0	0.0001	0
H3135B07-3	4.0332	3.11201	0.92119	8.34	0	0	0.0001	0 2810001A02Rik
H3048E11-3	3.2952	2.39171	0.90349	8.007	0	0	0.0003	0
H3021C01-3	4.6413	3.7465	0.8948	7.848	0	0	0.0002	0 Prkch
H3136B08-3	4.2003	3.3115	0.8888	7.741	0	0	0.0002	0 Npm1
H3023D06-3	4.2699	3.42411	0.84579	7.011	0	0	0.0002	0 Ppk1
H3125D09-3	3.2383	2.4055	0.8328	6.804	0	0	0.0001	0 St13
H3028A04-3	4.7448	3.92881	0.81599	6.546	0	0	0.0002	0 Ppia
H3138E02-3	2.5749	3.83789	-1.26299	18.322	0	0	0.0002	0 Col4a2
H3113E11-3	3.1783	4.2774	-1.0991	12.563	0	0	0.0001	0 Lamc1
H3153C07-3	2.8466	3.8045	-0.9579	9.076	0	0	0.0003	0 Lrpap1
H3125A07-3	2.6247	3.5641	-0.9394	8.697	0	0	0.0002	0 Serpinh1
H3147B02-3	2.5727	3.47479	-0.90209	7.981	0	0	0.0001	0 Rab11fip5
H3129B08-3	2.9778	3.8666	-0.8888	7.741	0	0	0.0005	0 H19
H3103H10-3	2.4969	3.36569	-0.86879	7.392	0	0	0.0007	0
H3097A02-3	3.3725	4.2297	-0.8572	7.197	0	0	0.0002	0 Lap3
H3153C04-3	2.5819	3.4371	-0.8552	7.164	0	0	0.0002	0 Ddt
H3153A11-3	2.9373	3.7863	-0.849	7.063	0	0	0.0001	0

Figure 6.18: IMT-11 vs. delayed ICM 180hpc. Top ten up regulated genes in red and down regulated genes in green below scatter plot.

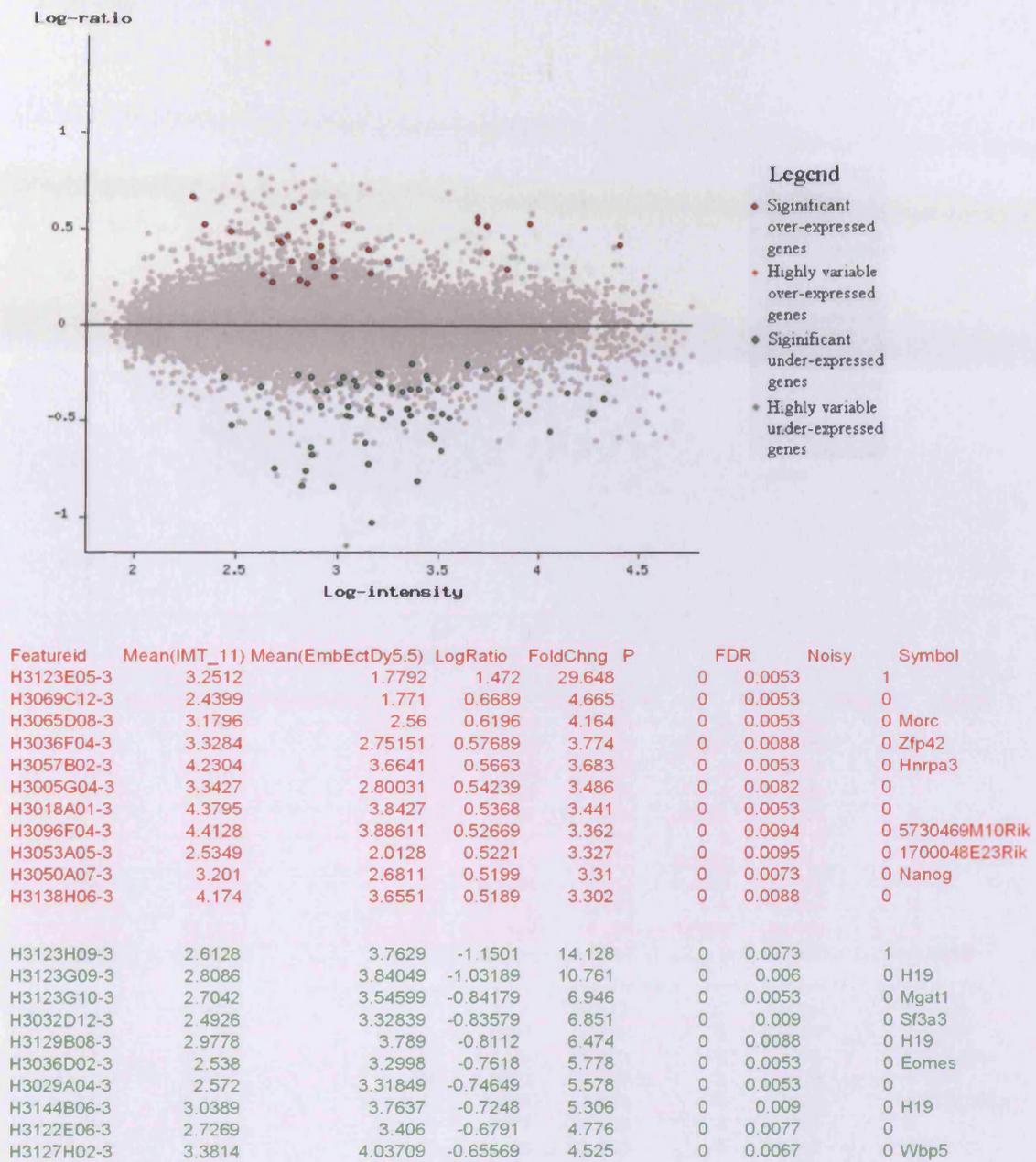
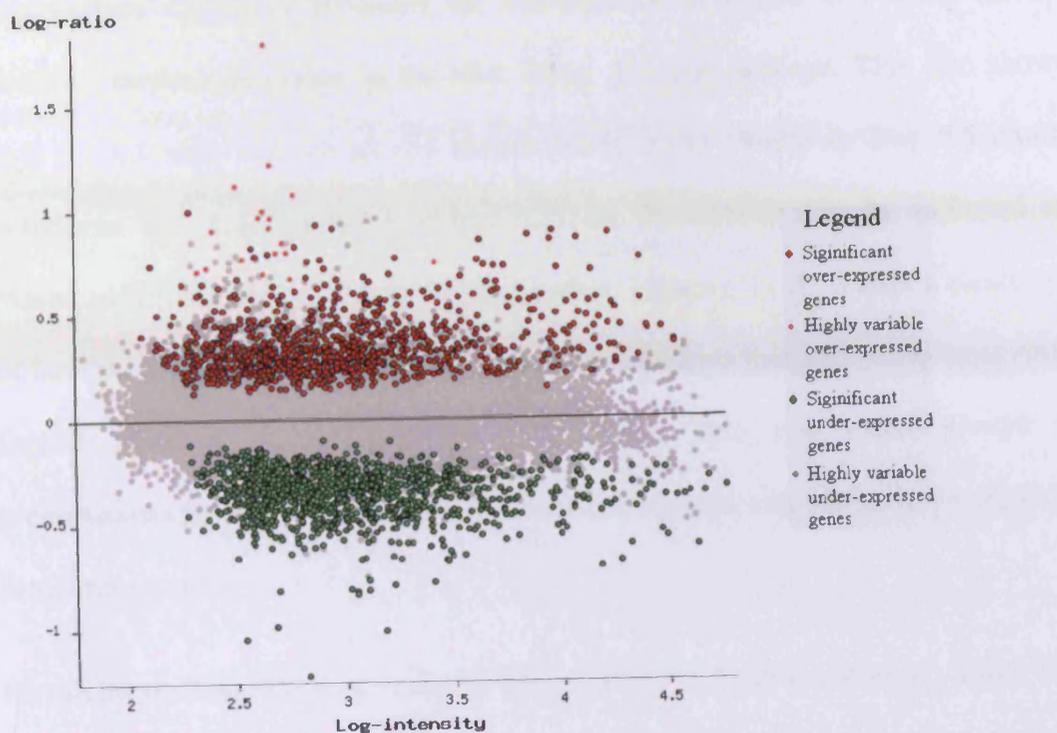


Figure 6.19: IMT-11 vs. epiblast day 5.5. Top ten up regulated genes in red and down regulated genes in green below scatter plot.



Featureid	Mean(IMT_11)	Mean(EmEct_Dy6.5)	LogRatio	FoldChng	P	FDR	Noisy	Symbol
H3123E05-3	3.2512	1.4542	1.797	62.661	0	0.0007	1	
H3051D10-3	3.4279	2.2091	1.2188	16.55	0.0002	0.0039	1	
H3091B05-3	3.2602	2.1378	1.1224	13.255	0	0.0021	1	
H3103E02-3	3.5722	2.49651	1.07569	11.903	0	0.0015	1	
H3085A10-3	3.2775	2.2779	0.9996	9.99	0.0001	0.0024	1	
H3148F08-3	2.8796	1.8811	0.9985	9.965	0	0.0013	0	
H3107G09-3	3.4113	2.4439	0.9674	9.276	0.0002	0.0041	1	
H3103A04-3	3.2926	2.3272	0.9654	9.234	0.0001	0.0028	1	
H3005G04-3	3.3427	2.41661	0.92609	8.435	0	0.0005	0	
H3056H08-3	3.1424	2.22671	0.91569	8.235	0.0002	0.0033	1	
H3036F04-3	3.3284	2.417	0.9114	8.154	0	0.0006	0	Zfp42
H3032D12-3	2.4926	3.7172	-1.2246	16.772	0	0.0008	0	Sf3a3
H3135E08-3	2.298	3.3487	-1.0507	11.238	0	0.0008	0	Gm47
H3003C03-3	2.7797	3.7945	-1.0148	10.346	0	0.0014	0	Dnmt3b
H3095H12-3	2.4135	3.4038	-0.9903	9.779	0	0.0009	0	Car4
H3149D10-3	2.8034	3.64279	-0.83939	6.908	0	0.0005	0	BC021523
H3103H10-3	2.4969	3.3246	-0.8277	6.725	0	0.0018	0	
H3153C04-3	2.5819	3.40379	-0.82189	6.635	0	0.0008	0	Ddt
H3029A04-3	2.572	3.3891	-0.8171	6.562	0	0.0009	0	
H3159E01-3	2.6609	3.474	-0.8131	6.502	0	0.0005	0	Sall2
H3155F02-3	2.7167	3.5108	-0.7941	6.224	0	0.0013	0	Atox1

Figure 6.20: IMT-11 vs. embryonic ectoderm day6.5. Top ten up regulated genes in red and down regulated genes in green below scatter plot.

Hierarchical clustering produced the dendrograms in Figure 6.21 using the average distance method described in the NIA Array Analysis package. This also shows that IMT-11 ES cells are, of all the tissues tested, most similar to Day 5.5 embryonic ectoderm. The dendrogram in Figure 6.21 (a) shows both that the technical repeats cluster together and the tissues cluster together. Figure 6.21 (b) shows a clearer picture of how the tissues are related to each other and confirms that IMT-11 is most similar to Day5.5 epiblast. Furthermore the data clusters into two major groups: firstly preimplantation which consists of all the ICM samples and secondly the epiblast and ectoderm samples.

Hierarchical clustering was repeated using Cluster 3.0 (de Hoon *et al.*, 2004; Eisen *et al.*, 1998) and the Spearman rank correlation as the similarity metric, the data was loaded into Java TreeView (Saldanha, 2004) First only the significantly changed genes (from the NIA annotation) for the whole experiment were loaded (Figure 6.22). As for the NIA analysis, the tissues group together and in most cases the technical repeats group together. Furthermore, the same pattern (that IMT-11 most resembles day 5.5 epiblast) is seen. Second, when the whole gene set was examined this gave the same result. Furthermore, where more than one probe for a particular gene has been printed on the array, they have clustered together (Figure 6.23). This clustering method uses non-parametric methods and, in combination with the NIA analysis, suggests the data is robust.

Principal component analysis (PCA) was also used to associate genes with tissues and to produce gene expression patterns (Figure 6.24) and again show tissue repeats clustering together. When the biplots are examined with the 3D viewer it is also clear that the ES line IMT-11 is most similar to Day5.5 epiblast. This could not only be

compared to the pattern produced by Cluster 3.0 but could be used for comparison to published data.

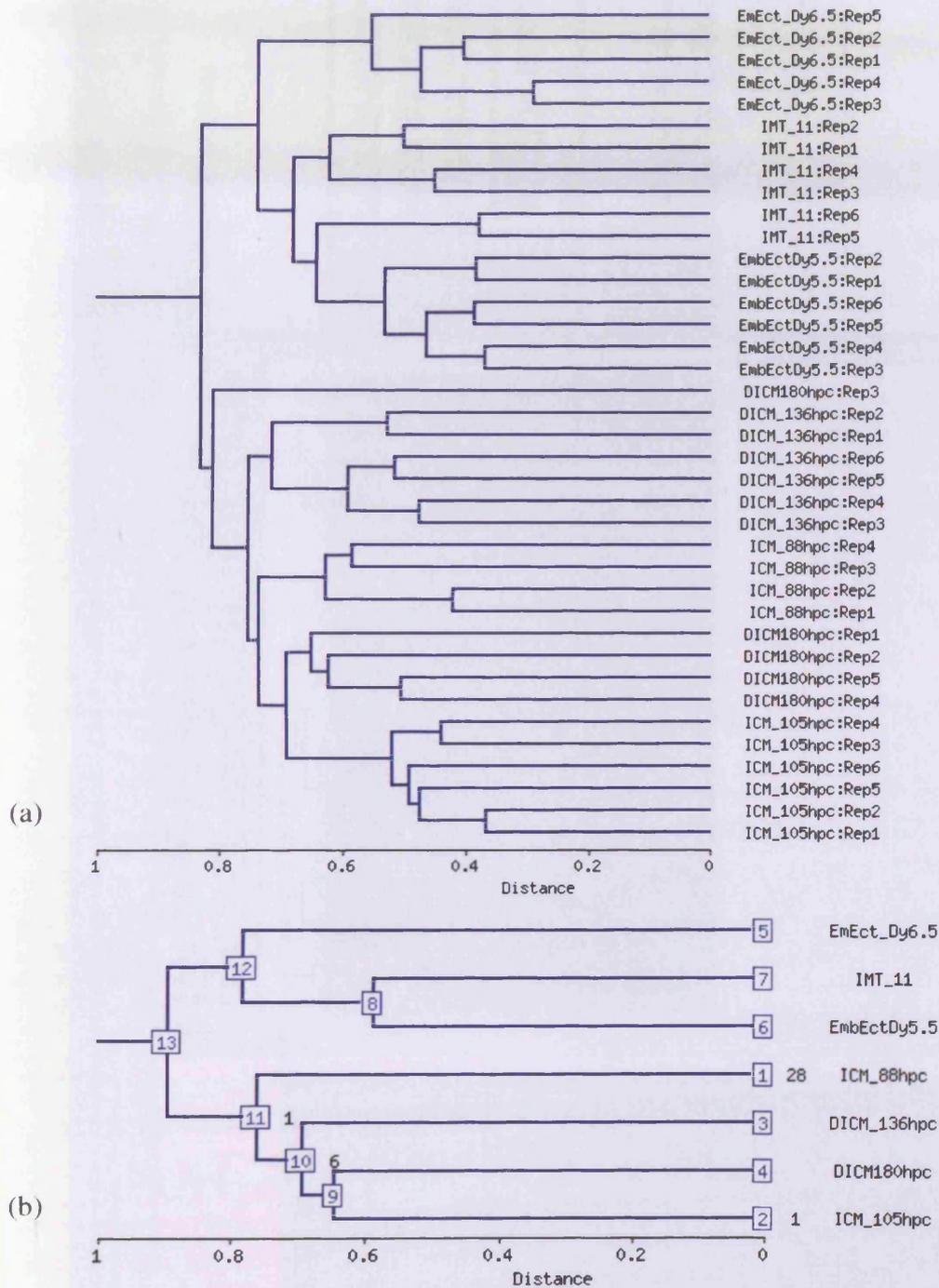


Figure 6.21 Dendrogram of hierarchical clustering of ES cell line IMT-11 and pluripotent embryonic tissues. The top dendrogram (a) shows the clustering of the repeated samples and (b) of the tissues.



Figure 6.22 : View of Java TreeView for regulated genes. In (b) red indicates high expression, green low.

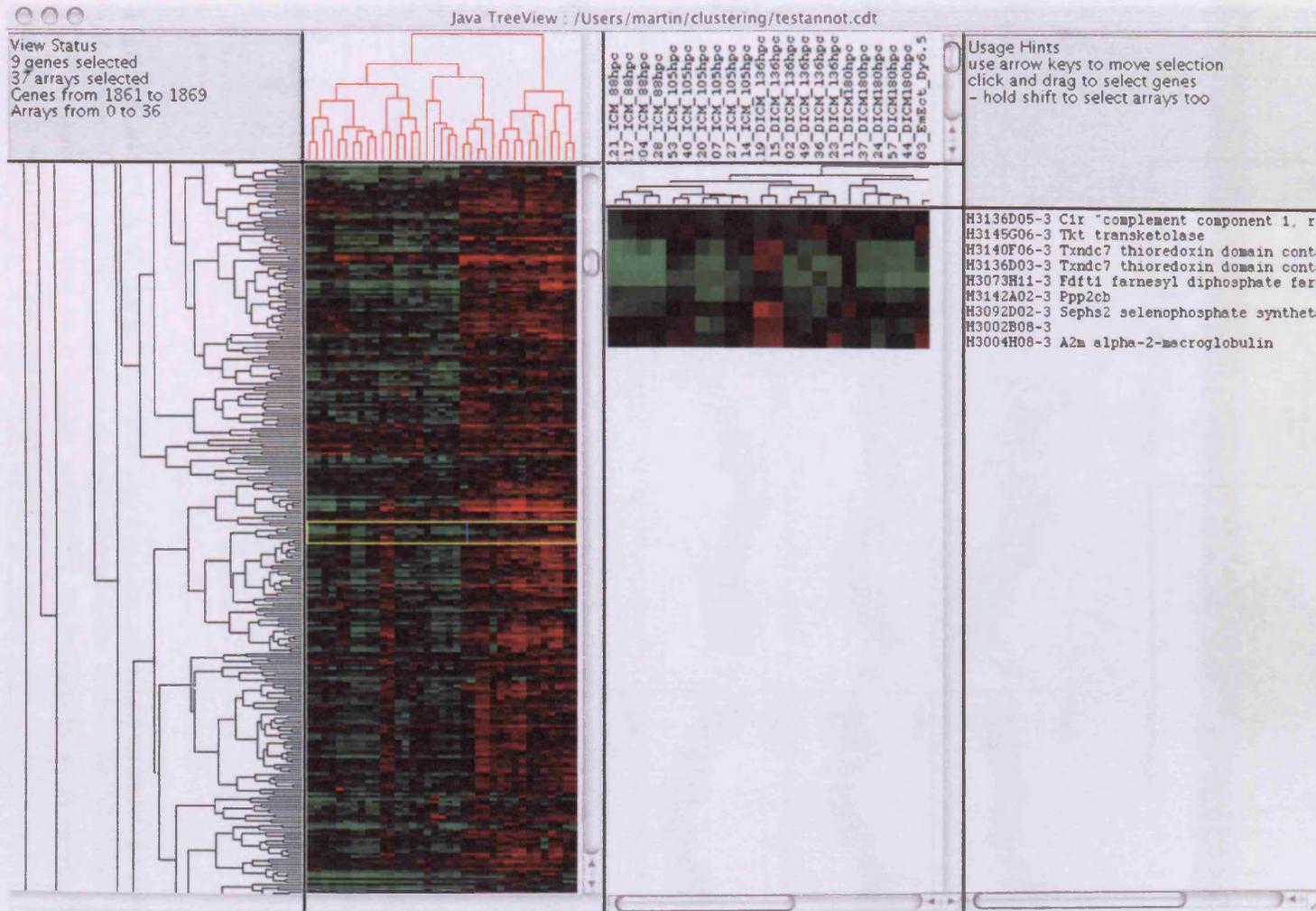


Figure 6.23: Image of Java TreeView analysis for all genes. In (b) red indicates high expression, green low.

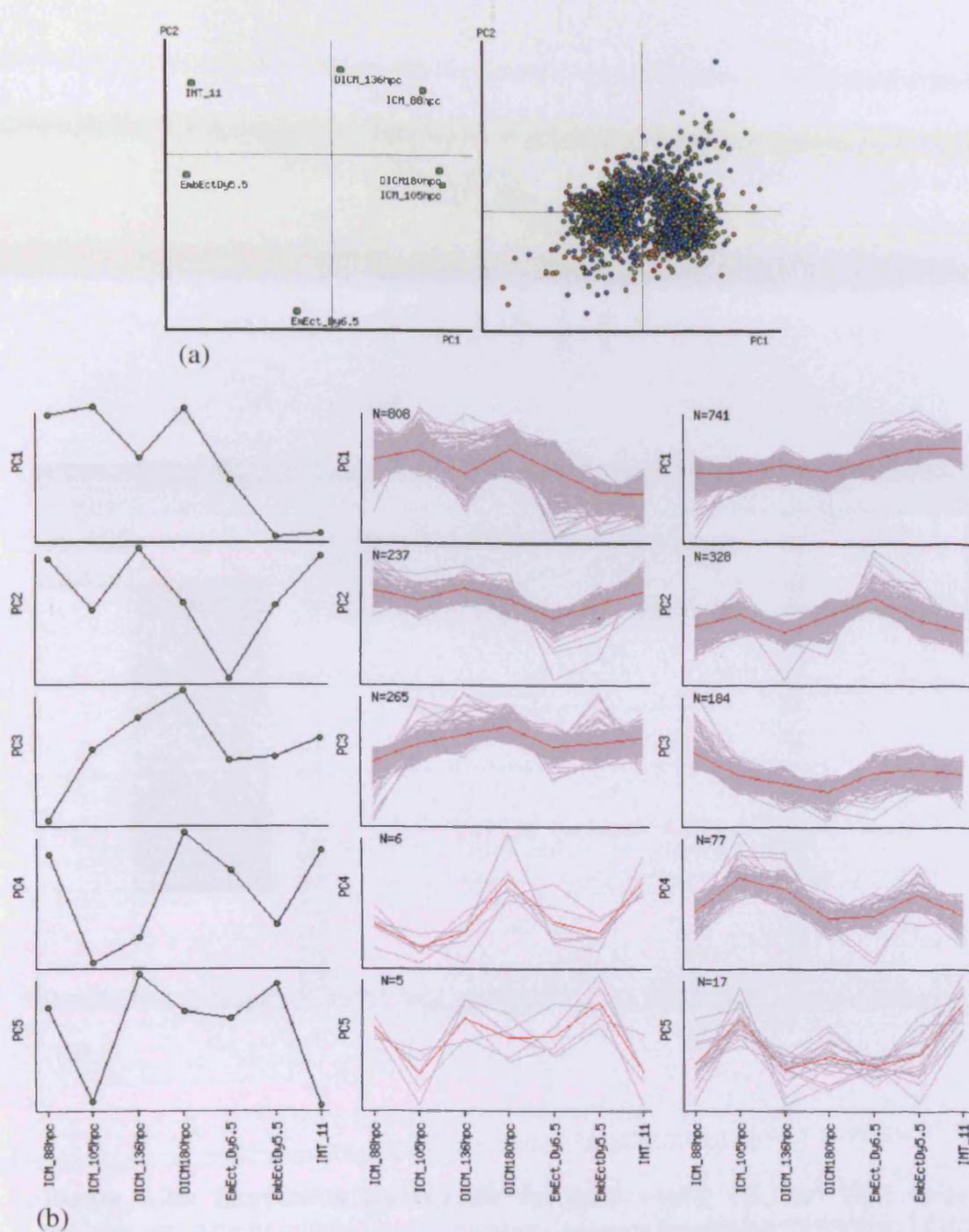
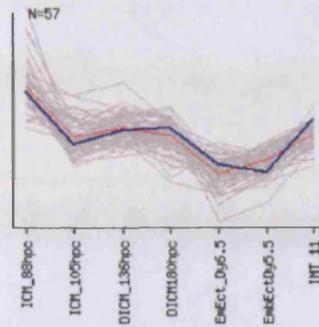
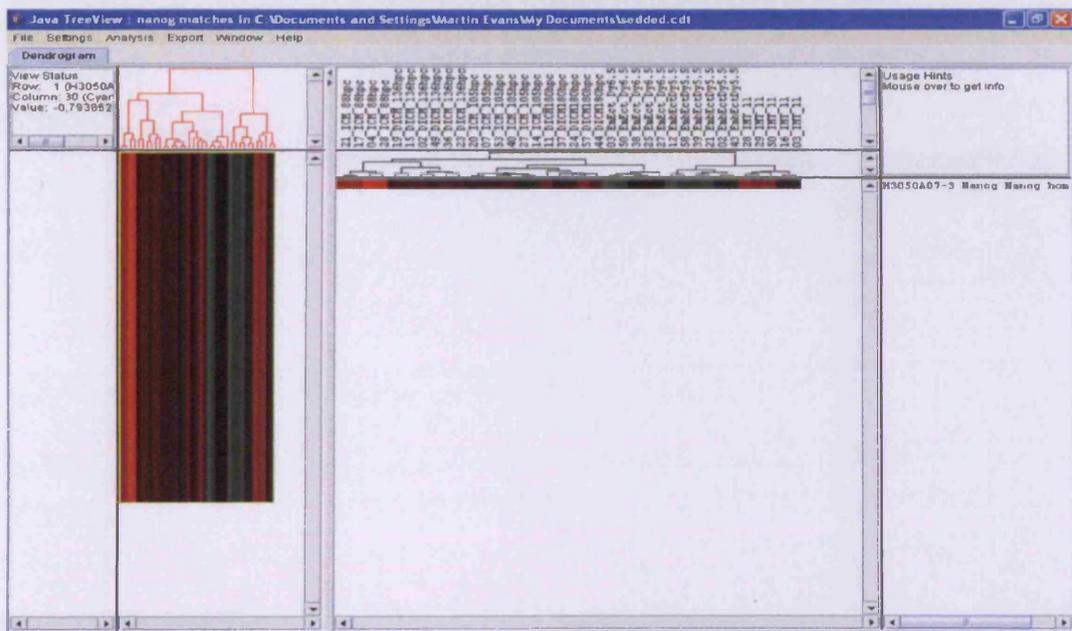


Figure 6.24: PCA of data. (a) biplot showing the relationship between the tissues and genes. This can be visualised in three dimensions within the NIA tool. (b) The PCA assigns each gene to a cluster according to a negative or positive direction correlation with the 5 principal components.

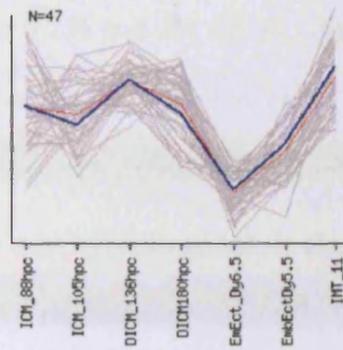


(a)

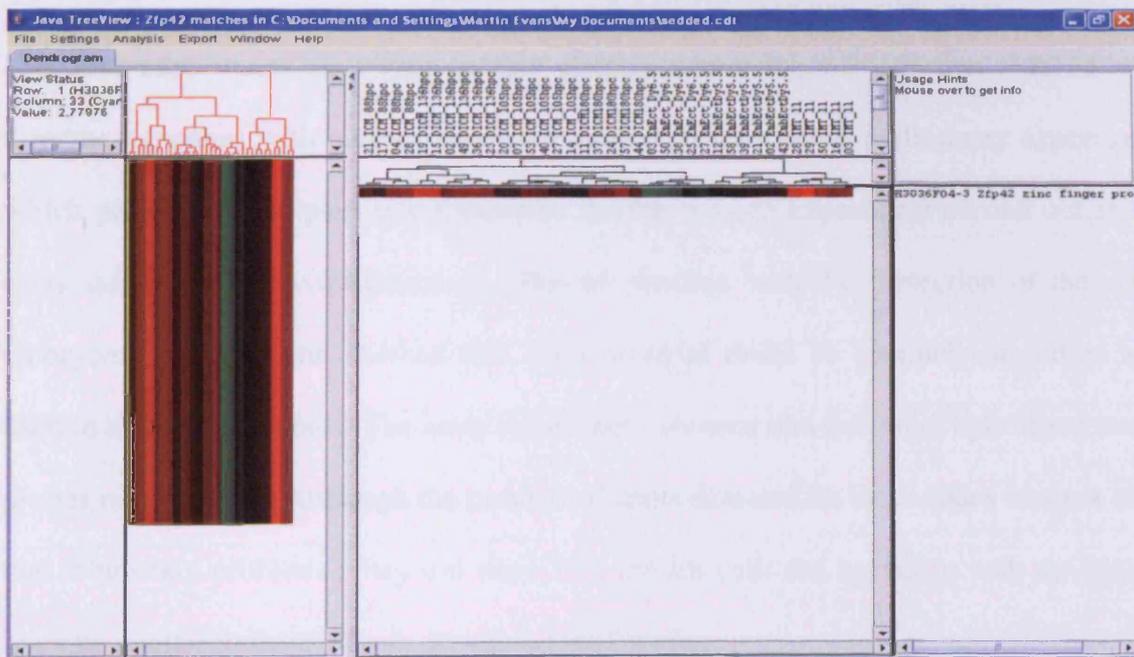


(b)

Figure 6.25: Expression pattern for the gene *nanog* (a) from NIA Array analysis and pattern matching and (b) from Cluster 3.0 and Java TreeView. In (b) red indicates high expression, green low.



(a)



(b)

Figure 6.26: Expression pattern for Zfp42. (a) from NIA Array analysis pattern matching and (b) from Cluster 3.0 and Java TreeView. In (b) red indicates high expression, green low.

The patterns for *nanog* (Figure 6.25) and for *Zfp42* (Figure 6.26) are as published for the embryo tissues (Chambers *et al.*, 2003; Pelton *et al.*, 2002). This demonstrates both that the expression profiles produced by these studies reflect well the profiles produced by other means (in this case *in situ* hybridisation). It also suggests that the study of the expression of single genes can be misleading.

6.4. Conclusions

The first experiment described in this chapter where the ES cell line IMT-11 was directly compared with various dissected embryo tissues was a preliminary experiment which performed two principal functions. As this was an experiment carried out at the early stages of the investigations it allowed practice with the dissection of the early embryonic material and showed that such material could be routinely amplified and used in array experiments. The array experiment showed that the target hybridised to the probes on the slides. Although the number of spots detected on these slides was not high due to printing problems, they did show that the ES cells did hybridise with the probes at a rate similar to the different developmental stages.

Notwithstanding the poor hybridisation on these slides, it was clear that a common reference design where the reference sample would not be part of the final analysis would use fewer array slides and simplify the analysis in this study. Therefore a common reference sample was assembled from amplified RNA from the tissues used to make the libraries represented in the NIA 15kprobe set. This also produced data about the potential usefulness of amplified RNA from ES cells as a universal reference sample. This showed that the number of probe spots hybridised at an intensity more than two standard deviations above the mean background intensity is high, greater than

87%, and very similar for ES cells and the common reference pool. This suggests that amplified RNA from ES cells would probably make a good universal reference sample at least for the NIA 15K gene set.

Image analysis of large numbers of array slides is very time consuming and potentially a source of variation. Batch processing allows each slide in a set from the same print run to be analysed using the same conditions and minimising operator intervention. This should both reduce operator variation and the amount of time required to analyse the slide images. The image analysis program ImaGene™ was used in most of the early experiments but this institute does not have the module which allows for batch processing thus the grid must be applied for each slide. While this produces very good results due to the flexibility of the placement and spot finding procedure, only four slides can be analysed in a day and the slides should only be analysed by one person. Spot, which can analyse a batch of slides so the grid is only set for one slide, is slightly less flexible in the placement of the grids so some spots will be lost and cannot be retrieved by manual adjustment. As the two programs use different methods of addressing and spot finding the results obtained are slightly different. As Spot uses all the information in a spot and does not discard the highest and lowest pixel values as is done by ImaGene™ it is more sensitive at lower spot intensities. Furthermore, the background variance is lower due to the method of background measurement and is more stable in Spot. Thus the results from Spot may reflect better the actual gene expression ratios; particularly for low expression levels.

The measurement of background has implications for the normalisation of the data. Limma with its wide range of normalisation options and visualisation methods was ideal for looking at the interaction between the image analysis methods and the

normalisation procedure. It was clear from the density distribution that the Spot analysed slides needed less adjustment than those analysed using ImaGene™ i.e. the data produced had less noise. Furthermore, background subtraction with ImaGene™ analysed data introduced noise, a phenomena noted by other workers (personal communication, Ben Routley).

The normalised data was extracted from limma and put into the NIA array analysis tool (<http://stemcell.biosi.cf.ac.uk/>) for ANOVA (analysis of variance), PCA (principal component analysis) and clustering. The use of Cluster 3.0 (de Hoon *et al.*, 2004; Eisen *et al.*, 1998) Java TreeView (Saldanha, 2004) allowed the clustering result to be confirmed using the non-parametric Spearman Rank correlation as the similarity measurement. The clustering pattern of the technical and biological repeats suggested high quality data as did the clustering of tissues. The number of regulated genes between each tissue pair was used as a measure of similarity and showed that the ES cell line IMT-11 was most similar to the epiblast on day5.5 of gestation despite the expression of individual genes suggesting ES cells are homologous to the ICM of the early blastocyst stage.

Chapter 7

Discussion

7. Discussion

The aim of this project was to develop techniques which would allow transcriptome phenotyping of the pluripotent compartment of the mouse embryo and embryonic stem cells. This would also enable gene expression comparisons to be carried out at a global level and has involved development of techniques for labelling of the target, amplification of RNA and handling of the data generated.

7.1 Labelling of Target

The development of labelling techniques was carried out using cytoplasmic RNA. Not only is this easily obtained in large quantities but amplified RNA takes several days to generate and is therefore a valuable resource. Nevertheless, the methods of labelling amplified RNA would be expected to be similar to those for cytoplasmic RNA.

Direct labelling of the mRNA was worth exploring as it is less likely to suffer from sequence specific dye biases or 3' biases which are inherent in methods which use enzymes. This method produced excellent array data (not presented in this thesis). The instability, however, of the reagents reduced reliability and increased cost.

It is generally recommended that at least 10pmols of dye labelled target is applied to each array slide. The direct integration of labelled nucleotides (Cy-dUTP) into cDNA by means of reverse transcription, which was already in use in this institute required 10µg of total RNA to produce barely enough Cy5™ labelled cDNA for a single microarray slide. This is irrelevant where the amount of RNA is not limiting but the amount of RNA available from the early embryo tissue samples was inevitably small. Furthermore, the unequal incorporation in the two dyes is also a potential problem. If

the dyes are not adequately balanced, the normalisation process can be detrimentally affected.

Integration of aminoallyl dUTP into the cDNA followed by conjugation of the dyes appeared to have a number of advantages over the direct incorporation of dye modified nucleotides. Firstly, the aminoallyl modified nucleotides are much less bulky than the dye modified molecules and thus are more efficiently incorporated into the cDNA. Secondly, the monoreactive dyes, Cy3™ and Cy5™, bind with similar avidity to the modified cDNA. Thirdly the dyes bound via a linker cause less interference with hybridisation.

The use of the reverse transcription reaction is a basic tool in this laboratory and it might seem that with optimised levels of reagents, in particular the relative level of aminoallyl dUTP to unmodified dUTP, there would be little benefit in using a commercial kit. However, the results obtained without a kit were highly variable. The relative levels of modified to unmodified dUTP can be expected to affect both the yield of cDNA and the level of incorporation. Much of the variation in yield can be attributed to the purification steps between reverse transcription and dye coupling and in the final removal of unbound free dye from the finished probe. With a kit, both yield of cDNA and the amount of dye incorporated are still variable but on the whole are more reliable than the in-house methods.

Destruction of the RNA template can be achieved by either the enzyme RNase H or sodium hydroxide. The kits which were being trialled, the Atlas™ Glass Fluorescent Labelling kit and the Atlas™ PowerScript™ Fluorescent Labelling kit (BD Biosciences) both used RNase H to destroy the template. As RNase H only destroys the RNA which is hybridised to DNA, spectrophotometric analysis could not be used to

assess the amount of cDNA synthesised against the background of cellular total RNA. Furthermore, inefficient removal of the RNA template could potentially lead to a drop in hybridisation of the probe due to these RNA fragments hybridising back to the labelled cDNA. Therefore RNase H and NaOH degradation of the RNA was tested. Degradation with NaOH allowed the amount of cDNA and the frequency of incorporation (FOI) of the labelled nucleotides to be estimated. This is important because cDNA with too high an FOI is inhibited from hybridisation. Further analysis of microarrays demonstrated that neither the number of spots detected nor the total intensity was affected by the method of RNA degradation when the same amount of dye labelled probe (15pmol) is hybridised with the probes on the slide. Thus the evidence is that the RNase H removes the template as efficiently as NaOH however to give a clearer picture of probe quality before use, the degradation of the template was in subsequent experiments carried out with NaOH.

The amount of cDNA synthesised in the reverse transcription reaction is affected by the molecule being incorporated. Aminoallyl dUTP does not inhibit the synthesis of cDNA to the same extent as Cy3TM or Cy5TM dUTP. Thus amount of cDNA synthesised per 10µg of RNA is higher for aminoallyl incorporation than for CyTM conjugated dyes as is the amount of dye incorporated into target. Although Cy3TM dUTP incorporates into cDNA at a similar FOI as aminoallyl dUTP Cy5TM only incorporates about half as much. Overall 5µg of total RNA will provide enough labelled target for 1 array slide when incorporating Cy3TM dUTP or only half the required amount when incorporating Cy5TM dUTP. When using the incorporation of aminoallyl dUTP 5µg of total RNA provides enough labelled target for 2 and often 3 array slides irrespective of which CyTM dye is being used. Interestingly, although both probes appeared to be equally labelled, the slope of the correlation line between Cy3TM and Cy5TM probes (Section 3. 3. 3c,

Figure 3.7 and Table 3.22) shows a difference in hybridisation between the two dyes, namely Cy5™ labelled target appears to hybridise with a higher affinity than Cy3™.

Overall, the labelling by the indirect method of incorporating aminoallyl dUTP into cDNA by reverse transcription followed by conjugation of monoreactive Cy™ dye has proved to be the best in terms of yield of labelled target. There was no detectable difference in the performance of the CyScribe™ Post-Labeling Kit (Amersham Pharmacia Biotech) and the Atlas™ PowerScript™ Labelling Kit (BD Biosciences) with the modification of NaOH degradation. The procedure with the later is slightly shorter so this was used in all subsequent experiments. The method was further modified by adding PelletPaintNF™ (Novagen, Merk Biosciences) (Section 3. 4) to the ethanol precipitation of the aminoallyl modified cDNA. This reduces losses of the aminoallyl cDNA by improving visibility of the pellet and making a more compact pellet which is less easily dislodged.

As a result of these tests it was decided to use the Atlas™ PowerScript™ Labelling kit (BD Biosciences) modified by degrading the RNA template with NaOH and the addition of PelletPaintNF™ (Novagen, Merk Biosciences) to promote pellet recovery during ethanol precipitation

7.1.2 Amplification

The primary aim of this thesis was to apply global transcriptome analysis to investigate the homology of ES cells to the pluripotent compartment of the embryo. Paucity of experimental material has restricted these studies in the past. (Tanaka and Ko, 2004) used 2510 morulae and 1149 blastocysts (+ other stages) from superovulated females to carry out a microarray experiment. The work of (Hamatani *et al.*, 2006) involved the

collection of pools of 500 embryos for each sample to generate the gene expression profile from superovulated embryos. Although it is generally adequate to produce early pre-implantation stages by superovulation, superovulated mothers rarely yield satisfactory embryos at later pre-implantation stages such as morula and blastocysts. Furthermore, not only would post-implantation development not be expected to be satisfactory after superovulation, but also the amount of “hands on” manipulation of each post-implantation embryo to isolate the embryonic ectoderm would hamper such studies. Thus development of an amplification method for the RNA was central to this study.

In the same way as the labelling procedure the enzymatic reactions carried out are routine (reverse transcription, double stranding and in vitro transcription) but are being carried out at the limits of their sensitivity. This not only means the optimisation of each reaction step is critical, but also the cleaning and buffer exchange steps are critical to ensure no carryover of components from one reaction to the next or loss of scant sample.

Although each reaction appeared to function and even some amplified RNA was made, the whole procedure did not lead to amplification. To achieve adequate amplification in a reproducible manner required the RiboAmp™ kit (Arcturus).

Labelled amplified RNA has the advantage of producing much brighter arrays than labelled total or cytoplasmic RNA. This is because the 3-5 μ g of amplified RNA being put into the labelling reaction is the equivalent of 3 μ g of highly purified mRNA. The disadvantage of amplified RNA is that it is 3' biased due to the lack of processivity of reverse transcriptase. Although this would have an impact on oligonucleotide arrays it has little impact on the cDNA arrays being used in this study because the probes are, in

the majority of cases derived from the 3' end of the gene. The disadvantage of cDNA arrays is that they do not give absolute transcript levels, but only relative. Furthermore, the NIA 15k gene set has a great many unidentified genes in it and the annotation is poor.

7.1.2 Image Analysis

Printed microarray slides are rarely perfect. Bent or worn pins introduce distortions in the grids of spots. More problematical is variation between slides.

In this study two different image analysis programs were used; namely ImaGene™ and Spot. ImaGene™ is more labour intensive in that the grid is placed for each slide image. However, this hand placed grid can be distorted and the metagrids moved to allow for variation from a true square grid. As this institution does not have the batch processor for this program, between slides variation is not a problem. On the other hand, with Spot, after initial calibration for a batch of slides, variation is handled by the program limited by preset parameters.

Segmentation into foreground and background with ImaGene™ is carried out by histograms which unlike fixed circle or adaptive circle segmentation does not depend on a precise location for the spot if it is within the mask area, nor is a perfectly circular spot required. Pixels are allocated to foreground or background on the basis of their intensity. If a spot is too far from its expected location, it can be adjusted manually. Spot, an adaptive shape segmentation method, also does not require perfectly circular spots or very precise information about the location of the spots as there are multiple seed points for the growth of the features. However, if the spot is too far from its

expected location, it can be missed. As it is not possible to adjust this manually, the data is lost (Figure 6.6 bottom left picture of segmented metagrid).

Another variable in image analysis is background acquisition. ImaGene™ with its histogram method uses the low intensity pixels between two low percentiles as background. This has the effect of filtering out high pixels due to dust, scratches or neighbouring spots resulting in lower backgrounds than for fixed or adaptive circle segmentation. Spot on the other hand takes seed points round the intersection point of the valleys between spots and calculates a background based on a mask more than twice the size of the spot distance. This has the effect of “peeling off” the spots leaving the background behind. The background measurement is taken from directly under the spot. This method has the effect of returning a lower background than fixed or adaptive circle methods and to a lesser extent also to histogram methods. However, it also has the effect of smoothing out the background and assumes that spatial changes in background are gradual. This is often not the case, background can change more abruptly. Thus ImaGene™ probably more closely reflects the reality on the slide. However, the aim is to reduce noise introduced by the background measurement so the Spot method and morph reduces the dependence of variability on intensity.

In addition to the background measurements for ImaGene™ and Spot, blank spots printed only with buffer are available for background subtraction. These are distributed across the whole slide to avoid spatial errors and can be used in a global manner or by metagrid. At present we have no algorithms available and can only use these with spreadsheet programs such as Excel.

7.1.3 Background Subtraction and Normalisation

The purpose of normalisation is to account for loading differences thus allowing comparison between slides, and to remove systematic biases inherent in the experiments which permit the extraction of data from microarrays. The most common forms of bias are dye bias and spatial bias. In these experiments, since a common reference design using a reference sample is used, sequence specific dye bias is not a problem. However, the dye balance between the two channels, which is not only related to differences in the efficiency of incorporation of the two labels but also differences in excitation and detection and loading, must be corrected. Most microarrays require some form of spatial normalisation as they often have darker areas especially round the edges. In this study, we used print-tip loess as our method of normalisation. Loess normalisation estimates the dependence of the log ratio M on the log intensity A . Applied globally, it does not correct for spatial variation and a separate algorithm would have to be applied. It is only appropriate where the bulk of the genes are not differentially regulated. Most spatial variation is associated with individual print tips due to for example wear or a bent pin, or edge effects due to drying of the hybridisation mix at the edge of the coverslips. When applied to each print-tip group individually, loess also performs as a spatial normalisation method; providing the print tip groups are not too large and there is not excessive variation across the metagrid area due to other spatial factors i.e. the metagrid must not act as an individual array. If the print tip groups are too small, then there will not be enough spots to carry out a reliable regression. Although other algorithms are available for spatial normalisation, as the number of separate manipulations to the data increase, the noise in the data increases. Thus print-tip loess carries out the jobs both of global normalisation; balancing the channels and smoothing the regression line, and of spatial normalisation.

Background subtraction is a more contentious issue. It is generally believed that spot intensity consists of signal due to hybridisation of labelled target to the probe, plus other fluorescent contamination from slide coatings, scanner filters etc, and furthermore is assumed to be additive. This contamination is represented by background. The measurement of background has its own noise associated with it, and since this can be high, subtraction can lead to zero and negative red or green intensity values and thus loss of data. This is strongly affected by the method of background measurement. Moreover additional noise is being added to the data which impacts on detection of differential gene expression particularly at lower intensities. Although it has been argued that background removal should be omitted from the analysis because Spot provides the morph measurement of background it was subtracted in these analyses. For ImaGene™ analysis, the limma function `normexp` was used as the background removal algorithm which gives a similar result to the subtraction of morph in Spot.

7.2 Embryonic Stem Cells

The only other study to compare the global transcriptome of ES cells and embryos was the study by Sharov *et al* (2003). This study compared the expression profile of ES cells, lineage restricted stem cells, whole embryos and newborn tissues. They showed that ES cells are closest in expression profile to the day 6.5 embryo rather than the day 4.5 blastocyst and EG cells are more similar to post-gastrulation day 8.5 embryos. Embryos on day 5.5 of gestation were not included in this study. Other studies have looked only at comparisons of different categories of stem cells in attempts to identify “stemness” genes and genes associated with the pluripotent state (Ivanova *et al.*, 2002; Kelly, 1977; Ramalho-Santos *et al.*, 2002; Sene *et al.*, 2007; Tanaka *et al.*, 2002) or pre-

implantation development and processes such as the activation of the zygotic genome (Tanaka *et al.*, 2000; Wang *et al.*, 2004; Zeng and Schultz, 2005) blastulation, (Tanaka and Ko, 2004) two-cell block (Jeong *et al.*, 2006) and activation of the diapause embryo (Hamatani *et al.*, 2004b) or have compared tissues from the same developmental stage (Frankenberg *et al.*, 2007). The study described here differs from the Sharov *et al* (2003) in many significant ways. First the Sharov *et al* (2003) study did not use microarray as its experimental platform but used the frequency of ESTs in many different libraries to describe their samples. This had the advantage of using the large number of clones in libraries already identified. This work led to the production of the NIA cDNA gene set used in our study. This study had the advantage of breadth and showed that many of the genes identified in microarray studies as unique to ES cells in other studies were in fact expressed in a differentiated tissue.

A feature of many array studies is the lack of repeatability between different research groups. This has been particularly noted in studies of stem cells. Although in many cases this variability can be at least partly assigned to different array platforms some can be put down to heterogeneity of cells in culture. This may be a particular problem in ES cells as they show heterogeneity which cannot simply be ascribed to the cell cycle. The gene *nanog* is an example of an ES marker which appears to be expressed at different levels in the population of cells growing in a dish (Singh *et al.*, 2007). This may partially account for the differences between the studies of Ivanova *et al* (2002) and Ramalho-Santos *et al* (2002) (Fortunel *et al.*, 2003) both of which used the Affimetrix gene chip platform.

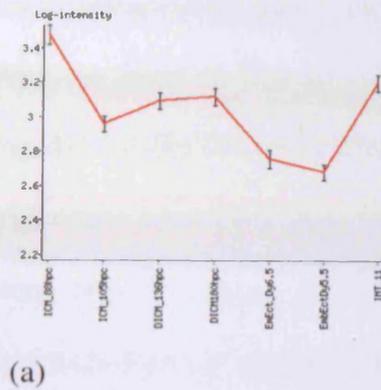
Examination of the under-expressed gene list from the ES cells against Day5.5 epiblast reveals a number of genes normally associated with the extra-embryonic portion of the

embryo namely *eomes* and H19. This would suggest contamination of the epiblast samples with extra-embryonic ectoderm which might be expected since the embryos are both very small and the egg cylinder only partially segmented. *Eomes* is expressed in the polar trophoctoderm and giant cells of the embryo on day 5.0 of gestation but this extends into the primitive streak by day 5.75 of gestation (Russ *et al.*, 2000). Thus *eomes* expression may not indicate contamination rather a spread in developmental stages. On the other hand H19 is expressed in the extra-embryonic ectoderm and ectoplacental cone on day 5.4 of gestation and is not expressed in the embryonic portion on day 6.5 of gestation (Poirier *et al.*, 1991) and most probably indicates contamination. This is confirmed on examination of the over-expressed gene list for day 5.5 epiblast against the ectoderm on day 6.5 (data not shown) where although *eomes* is not detected as differentially expressed, H19 and Plet1 (H3011D11) which is expressed in the distal extra-embryonic ectoderm (Frankenberg *et al.*, 2007) are. *Hand1* appears to be more proximally expressed in the extra-embryonic ectoderm on day 5.5 of gestation (Auman *et al.*, 2002) and is not listed in the under-expressed gene list for ES cells against day 5.5 epiblast although it is found in the list of over-expressed genes day 5.5 against day 6.5. However the expression data published for *Hand1* is not clear and *Hand1* may be expressed at low levels in ES cells assayed by PCR (Takada *et al.*, 2005).

The vast majority of ES lines have been isolated from the ICM of blastocyst stage embryos. They have also been isolated from morulae (Eistetter, 1989) and from early post-implantation, egg cylinder stages on day 5.5 pc (Wells *et al.*, 1991). However their homology to these *in vivo* stages is not clear. Although ES cell lines have been derived from up to day 5.5 of development, the expression of the known pluripotency markers reflects that of the earlier ICM (Section 1.1.7, Figure 1.1). Furthermore, their ability to contribute to chimaeras reflects this earlier homology. EC cells, which may be an ES

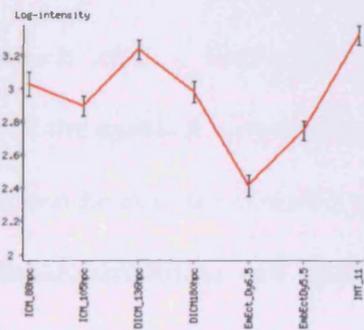
homologue, were compared to the ICM in an earlier proteomics study (Lovell-Badge and Evans, 1980). In this study 600 polypeptide spots could be compared with confidence. On the basis of the 32 spots which varied over the time course of the EC cells differentiating in vitro, the ICM from early blastocysts (shortly after blastulation) most resembled EC cells 6-12h after initiation of differentiation by embryoid body formation. At this time the embryoid bodies would be expected to be differentiating a layer of endoderm as would the ICMs and these common spots were postulated to be associated with this process. An advantage of this approach is the role of post-translational modification can be examined. However, although the technology has improved since this study it is still difficult to identify the proteins.

In this study global gene expression was analysed by transcriptome analysis for ES cells and the pluripotent component of the embryo between the blastocyst stage on day 3.5 and the embryonic ectoderm stage before gastrulation on day 6.5. Using the 15kNIA gene set allowed the expression profile of approximately 12,000 unique genes, for many of which the identity is known. A clear answer was obtained that the ES cell samples most resemble the day 5.5 epiblast samples, the two tissues only differing by 97 spots out of 15,248.



Tissue	Log-mean	SD
ICM_88hpc	3.4777	0.0584
ICM_105hpc	2.959	0.0477
DICM_136hpc	3.0967	0.0477
DICM180hpc	3.12	0.0522
EmEct_Dy6.5	2.7576	0.0522
EmbEctDy5.5	2.6811	0.0477
IMT_11	3.201	0.0477

(a)



Tissue	Log-mean	SD
ICM_88hpc	3.03	0.0716
ICM_105hpc	2.8958	0.0584
DICM_136hpc	3.2388	0.0584
DICM180hpc	2.9835	0.064
EmEct_Dy6.5	2.417	0.064
EmbEctDy5.5	2.7514	0.0584
IMT_11	3.3283	0.0584

(b)

Figure 7.1: Expression profile for (a) *nanog* and (b) *Zfp42*. Generated by the NIA array analysis tool. These graphs do not imply a time course.

Prominent in the list of over-expressed genes are *Zfp42* and *nanog*. The expression level of *Nanog* and *Zfp42* in the ES cells is significantly different from those of the day 5.5 epiblast. *Nanog* expression in the ES cells is most similar to the ICM of delayed blastocysts isolated at 180h.p.c. and *Zfp42* to the ICM of delayed embryos 136h.p.c. For homology to Day 5.5 of development, expression of these two genes would be expected to be lower in ES cells. This is unlikely to be an artefact of the microarray methodology as the expression profiles for *Zfp42* reflect the published expression profile (Pelton *et al.*, 2002). The expression profile for *Nanog* suggests slightly earlier down-regulation

than the published literature but this may be due to differences in embryo staging. Thus the known pluripotency genes may not be useful for characterisation of homology. Conversely, the level of Gsk-3 β is present in the list of down-regulated genes perhaps reflecting the activity of the canonical Wnt pathway. The expression of higher levels of the pluripotency associated genes in ES cells may reflect a metastable state adopted by pluripotent cells in culture under the serendipitous influence of the Gp130-STAT3 pathway which does not appear to act to maintain pluripotency in the embryo.

Over this period of development the embryo is initiating a wave of morphological change which leads to formation of the pro-amniotic cavity and rearrangement of the structure of the epiblast to form the ectoderm. It has been shown that a distinct cell type, EPL cells can be derived from ES cells by means of culture medium conditioned by the human hepatocarcinoma cell line HepG2 (Lake *et al.*, 2000). These cells have an expression profile of pluripotency markers like day 5.5 embryonic ectoderm (Pelton *et al.*, 2002). These cells are distinguished by being more like human ES cells in morphology; flatter, epithelial colonies rather than the heaped colonies of mouse ES. These cells also do not integrate into the blastocyst to form chimaeras. However, this state can be reversed. By modifying the growth conditions, notably by the addition of Lif, they can revert to an ES phenotype and regain their ability to contribute to chimaeras. It has also been demonstrated that under the influence of these culture conditions, the day 5.5 embryonic ectoderm can be maintained and furthermore these can be induced to give ES cell like lines which survive for up to five weeks (Rathjen *et al.*, 2003).

Recently, two groups have isolated stable cell lines, post-ES cells, from day 5.5 epiblast which had initiated pro-amniotic cavity formation (Brons *et al.*, 2007; Tesar *et al.*, 2007). These cells like EPL cells and human ES cells express Fgf5 and Oct3/4 but not Rex1. They also share with murine EPL and human ES an epithelial morphology and the inability to contribute to chimaeras (obviously this later property has not been tested in humans). These cells were derived using the same culture conditions as for human ES cells. This suggests that post ES cells, EPL cells and human ES cells, which grow in the absence of Lif, may reflect the *in vivo* epiblast. ES cells, being for the most part homologous to Day5.5 epiblast under the influence of Lif, the pluripotency markers of an earlier stage are enhanced.

It may be possible to test this by attempting to grow post-ES cells in the presence of Lif and observing if they can like the EPL cells described by Lake *et al* (Lake *et al.*, 2000) revert and form ES like lines which can contribute to chimaeras. The expression of individual pluripotency markers reflects the adaptation of these cells to culture and in particular the use of the gp130, Jak/STAT pathway which *in vivo* does not appear to have a role in the maintenance of the pluripotent state.

In combination with the ability to grow ES cells in the reversible EPL state it appears that ES cells in culture most resemble the embryonic epiblast on day 5.5 at the early pro-amniotic cavity stage. At this stage the pro-amniotic cavity is confined to the embryonic region of the early egg cylinder and has not extended into the extra-embryonic ectoderm and is equivalent to between day 5.25 and 5.5 as assessed by Pelton *et al*. The embryos used by Tesar *et al* (personal communication) to derive post-ES cells were also at this stage. Unfortunately, the paper by Wells *et al* (Wells *et al.*, 1991) was not precise in staging the embryos used

7.3 Future Work

As well as suggesting further work this data set contains a great deal of unanalysed information. Experiments to confirm gene expression profiles by other methods such as PCR are a priority although the errors from careful microarray studies such as those reported here are in many cases no larger than those from quantitative PCR. Further to the homology of ES cells to other pluripotent cell types, profiles of EC cells both of germ cell origin and ectoderm origin should be examined as should EG cells and primordial germ cells. Hopefully in the future it will be possible to profile EPL cells and post-ES cells

.

Chapter 8

Bibliography

8. Bibliography

- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F. et al.** (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-6.
- Alwine, J. C., Kemp, D. J. and Stark, G. R.** (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* **74**, 5350-4.
- Ambrosetti, D.-C., Scholer, H. R., Dailey, L. and Basilico, C.** (2000). Modulation of the Activity of Multiple Transcriptional Activation Domains by the DNA Binding Domains Mediates the Synergistic Action of Sox2 and Oct-3 on the Fibroblast Growth Factor-4 Enhancer. *Journal of Biological Chemistry* **275**, 23387-23397.
- Arnaud-Barbe, N., Cheynet-Sauvion, V., Oriol, G., Mandrand, B. and Mallet, F.** (1998). Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Res* **26**, 3550-4.
- Artzt, K., Dubois, P., Bennett, D., Condamine, H., Babinet, C. and Jacob, F.** (1973). Surface antigens common to mouse cleavage embryos and primitive teratocarcinoma cells in culture. *Proc Natl Acad Sci U S A* **70**, 2988-92.
- Aubert, J., Dunstan, H., Chambers, I. and Smith, A.** (2002). Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nat Biotechnol* **20**, 1240-5.
- Auman, H. J., Nottoli, T., Lakiza, O., Winger, Q., Donaldson, S. and Williams, T.** (2002). Transcription factor AP-2 $\{\gamma\}$ is essential in the extra-embryonic lineages for early postimplantation development. *Development* **129**, 2733-2747.
- Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R.** (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126-140.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merckenschlager, M. et al.** (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532.
- Baugh, L. R., Hill, A. A., Brown, E. L. and Hunter, C. P.** (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* **29**, E29.

Bazan, J. F. (1990). Structural Design and Molecular Evolution of a Cytokine Receptor Superfamily. *Proceedings of the National Academy of Sciences* **87**, 6934-6938.

Beddington, R. S. P. and Robertson, E. J. (1998). Anterior patterning in mouse. *Trends in Genetics* **14**, 277.

Beddington, R. S. P. and Robertson, E. J. (1999). Axis Development and Early Asymmetry in Mammals. *Cell* **96**, 195.

Belyavsky, A., Vinogradova, T. and Rajewsky, K. (1989). PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res* **17**, 2919-32.

Benjamini, Y. and Hochberg, Y. (1995). Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *Journal Of The Royal Statistical Society Series B-Methodological* **57**, 289-300.

Bennett, D. (1956). Developmental analysis of a mutation with pleiotropic effects in the mouse. *J. Morph* **98**, 199-234.

Berger, C. N. and Sturm, K. S. (1997). Self renewal of embryonic stem cells in the absence of feeder cells and exogenous leukaemia inhibitory factor. *Growth Factors* **14**, 145-59.

Berk, A. J. and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**, 721-32.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315.

Biebricher, C. K. and Luce, R. (1996). <http://www.nugeninc.com/technology/index.shtml>

Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase. *Embo J* **15**, 3458-65.

Bierbaum, P., MacLean-Hunter, S., Ehlert, F., Moroy, T. and Muller, R. (1994). Cloning of embryonal stem cell-specific genes: characterization of the transcriptionally controlled gene *esg-1*. *Cell Growth Differ* **5**, 37-46.

Bowles, J., Teasdale, R. P., James, K. and Koopman, P. (2003). Dppa3 is a marker of pluripotency and has a human homologue that is expressed in germ cell tumours. *Cytogenet Genome Res* **101**, 261-5.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-6.

Brennan, J., Lu, C. C., Norris, D. P., Rodriguez, T. A., Beddington, R. S. P. and Robertson, E. J. (2001). Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965.

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191.

Brook, F. A. and Gardner, R. L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5709-5712.

Burdon, T., Stracey, C., Chambers, I., Nichols, J. and Smith, A. (1999). Suppression of SHP-2 and ERK Signalling Promotes Self-Renewal of Mouse Embryonic Stem Cells. *Developmental Biology* **210**, 30.

Carter, M. G., Piao, Y., Dudekula, D. B., Qian, Y., VanBuren, V., Sharov, A. A., Tanaka, T. S., Martin, P. R., Bassey, U. C., Stagg, C. A. et al. (2003). The NIA cDNA project in mouse stem cells and early embryos. *C R Biol* **326**, 931-40.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K. and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132**, 885-896.

Chamberlin, M. J. and Ryan, T. (1982). In *The Enzymes*, (ed. P. Boyer). New York: Academic.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-55.

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

Chenchik, A. Z., Y.Y.; Diatchenko, L.; Li, R.; Hill, J. and Siebert, P.D. (1998). In *Gene Cloning and Analysis by RT-PCR*, (ed. P. a. L. J. Siebert), pp. 305-319. Natick, MA: Biotecniques Books.

Chomczynski, P. and Sacchi, N. (1987). Single-Step Method Of Rna Isolation By Acid Guanidinium Thiocyanate Phenol Chloroform Extraction. *Analytical Biochemistry* **162**, 156-159.

Conlon, F. L., Lyons, K. M., Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B. and Robertson, E. J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919-1928.

Conover, J. C., Ip, N. Y., Poueymirou, W. T., Bates, B., Goldfarb, M. P., Dechiara, T. M. and Yancopoulos, G. D. (1993). Ciliary Neurotrophic Factor Maintains the Pluripotentiality of Embryonic Stem-Cells. *Development* **119**, 559-565.

Coucouvanis, E. and Martin, G. R. (1995). Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. *Cell* **83**, 279-87.

Dafforn, A., Chen, P., Deng, G., Herrler, M., Iglehart, D., Koritala, S., Lato, S., Pillarisetty, S., Purohit, R., Wang, M. et al. (2004). Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. *Biotechniques* **37**, 854-7.

Damjanov, I., O., B. and Solter, D. (1983). Genetic and epigenetic factors regulate the evolving malignancy of embryo-derived teratomas. In *Cold Spring Harb. Conf. Cell Proliferation*, vol. 10 (ed., pp. 501-517. Cold Spring Harbor.

Dani, C., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B., Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J. et al. (1998). Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev Biol* **203**, 149-62.

Darnell, J. E., Jr., Kerr, I. M. and Stark, G. R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415-21.

de Hoon, M. J., Imoto, S., Nolan, J. and Miyano, S. (2004). Open source clustering software. *Bioinformatics* **20**, 1453-4.

Diwan, S. B. and Stevens, L. C. (1976). Development of teratomas from the ectoderm of mouse egg cylinders. *J. Natl. Cancer Inst.* **57**, 937-942.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45.

Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. and Trent, J. M. (1999). Expression profiling using cDNA microarrays. *Nat Genet* **21**, 10-4.

Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *PNAS* **95**, 14863-14868.

Eistetter, H. R. (1989). Pluripotent Embryonal Stem Cell Lines Can Be Established from Disaggregated Mouse Morulae. (mouse embryonal stem (ES) cells/disaggregated morulae/in vitro development/teratoma). *Development, Growth & Differentiation* **31**, 275-282.

Endege, W. O., Steinmann, K. E., Boardman, L. A., Thibodeau, S. N. and Schlegel, R. (1999). Representative cDNA libraries and their utility in gene expression profiling. *Biotechniques* **26**, 542-8, 550.

Enders, A. C. (1971). The fine structure of the blastocyst. In *Biology of the Blastocyst*, (ed. R. J. Blandau), pp. 71-94. Chicago: University of Chicago Press.

Ernst, M., Oates, A. and Dunn, A. R. (1996). gp130-mediated Signal Transduction in Embryonic Stem Cells Involves Activation of Jak and Ras/Mitogen-activated Protein Kinase Pathways. *Journal of Biological Chemistry* **271**, 30136-30143.

Evans, M. (1981). Origin of mouse embryonal carcinoma cells and the possibility of their direct isolation into tissue culture. *J Reprod Fertil* **62**, 625-31.

Evans, M. J. (1972). The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J. Embryol. exp. Morph.* **28**, 163-176.

Evans, M. J., Carlton, M. B. L. and Russ, A. P. (1997). Gene trapping and functional genomics. *Trends in Genetics* **13**, 370.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156.

Fortunel, N. O., Otu, H. H., Ng, H.-H., Chen, J., Mu, X., Chevassut, T., Li, X., Joseph, M., Bailey, C., Hatzfeld, J. A. et al. (2003). Comment on " 'Stemness': Transcriptional Profiling of Embryonic and Adult Stem Cells" and "A Stem Cell Molecular Signature" (I). *Science* **302**, 393b-.

Frankenberg, S., Smith, L., Greenfield, A. and Zernicka-Goetz, M. (2007). Novel gene expression patterns along the proximo-distal axis of the mouse embryo before gastrulation. *BMC Developmental Biology* **7**, 8.

Frohman, M. A., Dush, M. K. and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A* **85**, 8998-9002.

Gardner, R. L., Lyon, M. F., Evans, E. P. and Burtenshaw, M. D. (1985). Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. *J Embryol Exp Morphol* **88**, 349-63.

Gupta, V., Cherkassky, A., Chatis, P., Joseph, R., Johnson, A. L., Broadbent, J., Erickson, T. and DiMeo, J. (2003). Directly labeled mRNA produces highly precise and unbiased differential gene expression data. *Nucleic Acids Res* **31**, e13.

Haegeler, L., Ingold, B., Naumann, H., Tabatabai, G., Ledermann, B. and Brandner, S. (2003). Wnt signalling inhibits neural differentiation of embryonic stem cells by controlling bone morphogenetic protein expression. *Molecular and Cellular Neuroscience* **24**, 696.

Hamatani, T., Carter, M. G., Sharov, A. A. and Ko, M. S. H. (2004a). Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. *Developmental Cell* **6**, 117.

Hamatani, T., Daikoku, T., Wang, H., Matsumoto, H., Carter, M. G., Ko, M. S. H. and Dey, S. K. (2004b). Global gene expression analysis identifies molecular pathways distinguishing blastocyst dormancy and activation. *Proceedings of the National Academy of Sciences* **101**, 10326-10331.

Hamatani, T., Sh Ko, M., Yamada, M., Kuji, N., Mizusawa, Y., Shoji, M., Hada, T., Asada, H., Maruyama, T. and Yoshimura, Y. (2006). Global gene expression profiling of preimplantation embryos. *Human Cell* **19**, 98-117.

Hamazaki, T., Oka, M., Yamanaka, S. and Terada, N. (2004). Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. *J Cell Sci* **117**, 5681-5686.

Handyside, A. H. (1978). Time of commitment of inside cells isolated from preimplantation mouse embryos. *J Embryol Exp Morphol* **45**, 37-53.

Handyside, A. H. and Hunter, S. (1986). Cell division and death in the mouse blastocyst before implantation. *Roux Arch. Dev. Biol.* **195**, 519-526.

Handyside, A. H., O'Neill, G. T., Jones, M. and Hooper, M. L. (1989). Use of BRL-conditioned medium in combination with feeder cells to isolate a diploid embryonal stem cell line. *Roux Arch. Dev. Biol.* **198**, 48-55.

Hao, J., Li, T.-G., Qi, X., Zhao, D.-F. and Zhao, G.-Q. (2006). WNT/[beta]-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Developmental Biology* **290**, 81.

Hatano, S.-y., Tada, M., Kimura, H., Yamaguchi, S., Kono, T., Nakano, T., Suemori, H., Nakatsuji, N. and Tada, T. (2005). Pluripotential competence of cells associated with Nanog activity. *Mechanisms of Development* **122**, 67.

He, S., Pant, D., Schiffmacher, A., Meece, A. and Keefer, C. L. (2008). Lef1 Mediated Wnt Signaling Promotes the Initiation of Trophoblast Lineage Differentiation in Mouse Embryonic Stem Cells. *Stem Cells*, 2007-0356.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N. and Quackenbush, J. (2000). A concise guide to cDNA microarray analysis. *Biotechniques* **29**, 548-+.

Hibi, M., Murakami, M., Saito, M., Hirano, T., Taga, T. and Kishimoto, T. (1990). Molecular-Cloning And Expression Of An Il-6 Signal Transducer, Gp130. *Cell* **63**, 1149-1157.

Hogan, B. and Tilly, R. (1978). In vitro development of inner cell masses isolated immunosurgically from mouse blastocysts. II. Inner cell masses from 3.5- to 4.0-day p.c. blastocysts. *J Embryol Exp Morphol* **45**, 107-21.

Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C. and Birchmeier, W. (2000). Requirement for {beta}-Catenin in Anterior-Posterior Axis Formation in Mice. *J. Cell Biol.* **148**, 567-578.

Hunter, S. M. and Evans, M. (1999). Non-surgical method for the induction of delayed implantation and recovery of viable blastocysts in rats and mice by the use of tamoxifen and Depo-Provera. *Molecular Reproduction And Development* **52**, 29-32.

Hunter, S. M., Mansergh, F. C. and Evans, M. J. Optimization of minuscule samples for use with cDNA microarrays. *Journal of Biochemical and Biophysical Methods In Press, Corrected Proof.*

Iscove, N. N., Barbara, M., Gu, M., Gibson, M., Modi, C. and Winegarden, N. (2002). Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat Biotechnol* **20**, 940-3.

- Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A. and Lemischka, I. R.** (2002). A Stem Cell Molecular Signature. *Science* **298**, 601-604.
- Jacob, F.** (1978). Mouse teratocarcinoma and mouse embryo. *Proc. R. Soc. Lond. B Biol. Sci.* **201**, 249-270.
- James, D., Levine, A. J., Besser, D. and Hemmati-Brivanlou, A.** (2005). TGF beta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem. *Development* **132**, 1273-1282.
- Jeong-Heon Lee, S. R. L. H. D. G. S.** (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. *genesis* **38**, 32-38.
- Jeong, H.-J., Kim, H. J., Lee, S.-H., Kwack, K., Ahn, S.-Y., Choi, Y.-J., Kim, H.-G., Lee, K.-W., Lee, C.-N. and Cha, K.-Y.** (2006). Gene expression profiling of the pre-implantation mouse embryo by microarray analysis: Comparison of the two-cell stage and two-cell block. *Theriogenology* **66**, 785.
- Johnson, M. H. and Ziomek, C. A.** (1981). The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71.
- Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R. and Kondoh, H.** (1995). Involvement of SOX proteins in lens-specific activation of crystallin genes. *Embo J* **14**, 3510-9.
- Kanehisa, M. and Goto, S.** (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* **28**, 27-30.
- Kelly, S. J.** (1977). Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J Exp Zool* **200**, 365-76.
- Ko, M. S., Threat, T. A., Wang, X., Horton, J. H., Cui, Y., Pryor, E., Paris, J., Wells-Smith, J., Kitchen, J. R., Rowe, L. B. et al.** (1998). Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome. *Hum Mol Genet* **7**, 1967-78.
- Labosky, P. A., Barlow, D. P. and Hogan, B. L. M.** (1994). Mouse Embryonic Germ (Eg) Cell-Lines - Transmission Through The Germline And Differences In The Methylation Imprint Of Insulin-Like Growth-Factor 2 Receptor (Igf2r) Gene Compared With Embryonic Stem (Es) Cell-Lines. *Development* **120**, 3197-3204.

- Lake, J., Rathjen, J., Remiszewski, J. and Rathjen, P. D.** (2000). Reversible programming of pluripotent cell differentiation. *J Cell Sci* **113**, 555-566.
- Lawson, K. A., Meneses, J. J. and Pedersen, R. A.** (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891-911.
- Liang, P. and Pardee, A. B.** (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-71.
- Lloyd, S., Fleming, T. P. and Collins, J. E.** (2003). Expression of Wnt genes during mouse preimplantation development. *Gene Expr Patterns* **3**, 309-12.
- Lovell-Badge, R. H. and Evans, M. J.** (1980). Changes in protein synthesis during differentiation of embryonal carcinoma cells, and a comparison with embryo cells. *J Embryol Exp Morphol* **59**, 187-206.
- Luo, L., Salunga, R. C., Guo, H., Bittner, A., Joy, K. C., Galindo, J. E., Xiao, H., Rogers, K. E., Wan, J. S., Jackson, M. R. et al.** (1999). Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* **5**, 117-22.
- Martin, G. R.** (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7634-7638.
- Martin, G. R. and Evans, M. J.** (1974). The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**, 163-172.
- Martin, G. R. and Evans, M. J.** (1975a). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1441-1445.
- Martin, G. R. and Evans, M. J.** (1975b). Multiple differentiation of clonal teratocarcinoma stem cells following embryoid body formation in vitro. *Cell* **6**, 467-474.
- Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T. and Yokota, T.** (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *Embo J* **18**, 4261-9.
- Matsui, Y., Toksoz, D., Nishikawa, S., Nishikawa, S.-I., Williams, D., Zsebo, K. and Hogan, B. L. M.** (1991). Effect of Steel factor and leukaemia inhibitory factor on murine primordial germ cells in culture. *Nature* **353**, 750.

- Matsui, Y., Zsebo, K. and Hogan, B. L. M.** (1992). Derivation Of Pluripotential Embryonic Stem-Cells From Murine Primordial Germ-Cells In Culture. *Cell* **70**, 841-847.
- Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. and Chenchik, A.** (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res* **27**, 1558-60.
- Meijer, L., Skaltsounis, A. L., Magiatis, P., Polychronopoulos, P., Knockaert, M., Leost, M., Ryan, X. Z. P., Vonica, C. A., Brivanlou, A., Dajani, R. et al.** (2003). GSK-3-selective inhibitors derived from Tyrian purple indirubins. *Chemistry & Biology* **10**, 1255-1266.
- Meno, C., Shimon, A., Saijoh, Y., Yashiro, K., Mochida, K., Ohishi, S., Noji, S., Kondoh, H. and Hamada, H.** (1998). *lefty-1* Is Required for Left-Right Determination as a Regulator of *lefty-2* and *nodal*. *Cell* **94**, 287.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T. and Misteli, T.** (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-16.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S.** (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-42.
- Munoz-Sanjuan, I. and Brivanlou, A. H.** (2002). Neural induction, the default model and embryonic stem cells. *Nat Rev Neurosci* **3**, 271-80.
- Murakami, M., Hibi, M., Nakagawa, N., Nakagawa, T., Yasukawa, K., Yamanishi, K., Taga, T. and Kishimoto, T.** (1993). IL-6-induced homodimerization of gp130 and associated activation of a tyrosine kinase. *Science* **260**, 1808-10.
- Nadijcka, M. and Hillman, N.** (1974). Ultrastructural studies of the mouse blastocyst substages. *J Embryol Exp Morphol* **32**, 675-95.
- Nagy, A. G., M; Vintersten, K and Behringer, R.** (2003). *Manipulating the Mouse Embryo: A Laboratory Manual.*, (ed. Cold Spring Harbour, New York: Cold Spring Harbour Press.
- Nagy, Z. B., Kelemen, J. Z., Feher, L. Z., Zvara, A., Juhasz, K. and Puskas, L. G.** (2005). Real-time polymerase chain reaction-based exponential sample amplification for microarray gene expression profiling. *Anal Biochem* **337**, 76-83.

- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S.** (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotech* **26**, 101.
- Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J.-i., Matoba, R., Ko, M. S. H. et al.** (2006). Klf4 Cooperates with Oct3/4 and Sox2 To Activate the Lefty1 Core Promoter in Embryonic Stem Cells. *Mol. Cell. Biol.* **26**, 7772-7782.
- Nichols, J., Chambers, I. and Smith, A.** (1994). Derivation of Germline Competent Embryonic Stem-Cells with a Combination of Interleukin-6 and Soluble Interleukin-6 Receptor. *Experimental Cell Research* **215**, 237-239.
- Nichols, J., Davidson, D., Taga, T., Yoshida, K., Chambers, I. and Smith, A.** (1996). Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis. *Mech Dev* **57**, 123-31.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. and Smith, A.** (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-91.
- Nicolson, G. L., Yanagimachi, R. and Yanagimachi, H.** (1975). Ultrastructural localization of lectin-binding sites on the zonae pellucidae and plasma membranes of mammalian eggs. *J Cell Biol* **66**, 263-74.
- Niwa, H.** (2001). Molecular Mechanism to Maintain Stem Cell Renewal of ES Cells. *Cell Structure and Function* **26**, 137.
- Niwa, H.** (2007). How is pluripotency determined and maintained? *Development* **134**, 635-646.
- Niwa, H., Burdon, T., Chambers, I. and Smith, A.** (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-60.
- Niwa, H., Miyazaki, J. and Smith, A. G.** (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics* **24**, 372-376.
- Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H.** (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**, 461-72.

Okita, K., Ichisaka, T. and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. and Matsubara, K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat Genet* **2**, 173-9.

Okumura-Nakanishi, S., Saito, M., Niwa, H. and Ishikawa, F. (2005). Oct-3/4 and Sox2 Regulate Oct-3/4 Gene in Embryonic Stem Cells. *Journal of Biological Chemistry* **280**, 5307-5317.

Ouhibi, N., Sullivan, N. F., English, J., Colledge, W. H., Evans, M. J. and Clarke, N. J. (1995). Initial culture behaviour of rat blastocysts on selected feeder cell lines. *Mol Reprod Dev* **40**, 311-24.

Paling, N. R. D., Wheadon, H., Bone, H. K. and Welham, M. J. (2004). Regulation of Embryonic Stem Cell Self-renewal by Phosphoinositide 3-Kinase-dependent Signaling. *Journal of Biological Chemistry* **279**, 48063-48070.

Palmieri, S. L., Peter, W., Hess, H. and Scholer, H. R. (1994). Oct-4 Transcription Factor Is Differentially Expressed In The Mouse Embryo During Establishment Of The First 2 Extraembryonic Cell Lineages Involved In Implantation. *Developmental Biology* **166**, 259-267.

Papaioannou, V. E., McBurney, M. W. and Gardner, R. L. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258**, 70-73.

Pelton, T. A., Bettess, M. D., Lake, J., Rathjen, J. and Rathjen, P. D. (1998). Developmental complexity of early mammalian pluripotent cell populations in vivo and in vitro. *Reproduction Fertility And Development* **10**, 535-549.

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J. and Rathjen, P. D. (2002). Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J Cell Sci* **115**, 329-39.

Pennica, D., Shaw, K. J., Swanson, T. A., Moore, M. W., Shelton, D. L., Zioncheck, K. A., Rosenthal, A., Taga, T., Paoni, N. F. and Wood, W. I. (1995). Cardiotrophin-1 - Biological-Activities and Binding to the Leukemia Inhibitory Factor-Receptor Gp130 Signaling Complex. *Journal of Biological Chemistry* **270**, 10915-10922.

Pera, M. F., Andrade, J., Houssami, S., Reubinoff, B., Trounson, A., Stanley, E. G., Ward-van Oostwaard, D. and Mummery, C. (2004). Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. *J Cell Sci* **117**, 1269-80.

Pevny, L. H. and Lovell-Badge, R. (1997). Sox genes find their feet. *Current Opinion in Genetics & Development* **7**, 338.

Pfister, S., Steiner, K. A. and Tam, P. P. L. (2007). Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. *Gene Expression Patterns* **7**, 558.

Poirier, F., Chan, C. T., Timmons, P. M., Robertson, E. J., Evans, M. J. and Rigby, P. W. (1991). The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in the developing embryo. *Development* **113**, 1105-1114.

Polacek, D. C., Passerini, A. G., Shi, C., Francesco, N. M., Manduchi, E., Grant, G. R., Powell, S., Bischof, H., Winkler, H., Stoeckert, C. J., Jr. et al. (2003). Fidelity and enhanced sensitivity of differential transcription profiles following linear amplification of nanogram amounts of endothelial mRNA. *Physiol Genomics* **13**, 147-56.

Pratt, H. P., Ziomek, C. A., Reeve, W. J. and Johnson, M. H. (1982). Compaction of the mouse embryo: an analysis of its components. *J Embryol Exp Morphol* **70**, 113-32.

Qi, X., Li, T. G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y. and Zhao, G. Q. (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A* **101**, 6027-32.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. and Melton, D. A. (2002). "Stemness": Transcriptional Profiling of Embryonic and Adult Stem Cells. *Science* **298**, 597-600.

Rathjen, J., Lake, J. A., Bettess, M. D., Washington, J. M., Chapman, G. and Rathjen, P. D. (1999). Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *Journal Of Cell Science* **112**, 601-612.

Rathjen, J., Washington, J. M., Bettess, M. D. and Rathjen, P. D. (2003). Identification of a Biological Activity That Supports Maintenance and Proliferation of Pluripotent Cells from the Primitive Ectoderm of the Mouse. *Biol Reprod* **69**, 1863-1871.

Resnick, J. L., Bixler, L. S., Cheng, L. Z. and Donovan, P. J. (1992). Long-Term Proliferation Of Mouse Primordial Germ-Cells In Culture. *Nature* **359**, 550-551.

- Robertson, E. J.** (1987). Embryo-derived stem cell lines. In *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, (ed. E. J. Robertson), pp. 71-112. Oxford: IRL Press.
- Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W. J. and Staudt, L. M.** (1990). A Pou-Domain Transcription Factor In Early Stem-Cells And Germ-Cells Of The Mammalian Embryo. *Nature* **345**, 686-692.
- Rossant, J. and Cross, J. C.** (2001). Placental development: Lessons from mouse mutants. *Nat Rev Genet* **2**, 538.
- Russ, A. P., Wattler, S., Colledge, W. H., Aparicio, S. A. J. R., Carlton, M. B. L., Pearce, J. J., Barton, S. C., Surani, M. A., Ryan, K., Nehls, M. C. et al.** (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95.
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N.** (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-4.
- Saldanha, A. J.** (2004). Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248.
- Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1998). *Molecular Cloning - a Laboratory Manual*. New York, USA: Cold Spring Harbour.
- Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. and Brivanlou, A. H.** (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55-63.
- Sato, N., Sanjuan, I. M., Heke, M., Uchida, M., Naef, F. and Brivanlou, A. H.** (2003). Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol* **260**, 404-13.
- Schena, M., Shalon, D., Davis, R. W. and Brown, P. O.** (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-70.
- Scholer, H. R., Hatzopoulos, A. K., Balling, R., Suzuki, N. and Gruss, P.** (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J* **8**, 2543-50.

- Sene, K., Porter, C., Palidwor, G., Perez-Iratxeta, C., Muro, E., Campbell, P., Rudnicki, M. and Andrade-Navarro, M.** (2007). Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* **8**, 85.
- Seth, D., Gorrell, M. D., McGuinness, P. H., Leo, M. A., Lieber, C. S., McCaughan, G. W. and Haber, P. S.** (2003). SMART amplification maintains representation of relative gene expression: quantitative validation by real time PCR and application to studies of alcoholic liver disease in primates. *J Biochem Biophys Methods* **55**, 53-66.
- Singh, A. M., Hamazaki, T., Hankowski, K. E. and Terada, N.** (2007). A Heterogeneous Expression Pattern for Nanog in Embryonic Stem Cells. *Stem Cells* **25**, 2534-2542.
- Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D.** (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90.
- Smith, A. G. and Hooper, M. L.** (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol* **121**, 1-9.
- Smith, L., Underhill, P., Pritchard, C., Tymowska-Lalanne, Z., Abdul-Hussein, S., Hilton, H., Winchester, L., Williams, D., Freeman, T., Webb, S. et al.** (2003). Single primer amplification (SPA) of cDNA for microarray expression analysis. *Nucleic Acids Res* **31**, e9.
- Smith, R. and McLaren, A.** (1977). Factors affecting the time of formation of the mouse blastocoele. *J Embryol Exp Morphol* **41**, 79-92.
- Smith, T. A. and Hooper, M. L.** (1983). Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Exp Cell Res* **145**, 458-62.
- Snell, G. D. and Stevens, L. C.** (1966). Early embryology. In *Biology of the Laboratory Mouse*, (ed. E. L. Green), pp. 205-245. New York: McGraw-Hill.
- Snow, M. H. L.** (1976). Embryo growth during the immediate postimplantation period. In *Ciba Foundation symposium*, (ed.: Elsevier; Excerpta Medica; North Holland, Amsterdam).
- Solter, D.** (2006). From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat Rev Genet* **7**, 319.

Solter, D. and Knowles, B. B. (1975). Immunosurgery of mouse blastocyst. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 5099-5102.

Solter, D. and Knowles, B. B. (1978). Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5565-5569.

Solter, D., Skreb, N. and Damjanov, I. (1970). Extrauterine Growth of Mouse Egg-cylinders results in Malignant Teratoma. *Nature* **227**, 503.

Sousa, R., Patra, D. and Lafer, E. M. (1992). Model for the mechanism of bacteriophage T7 RNAP transcription initiation and termination. *J Mol Biol* **224**, 319-34.

Stahl, N. and Yancopoulos, G. D. (1994). The tripartite CNTF receptor complex: activation and signaling involves components shared with other cytokines. *J Neurobiol* **25**, 1454-66.

Stears, R. L., Getts, R. C. and Gullans, S. R. (2000). A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol Genomics* **3**, 93-9.

Stevens, L. C. (1964). Experimental production of testicular teratomas in mice. *Proc. Natl. Acad. Sci. U.S.A.* **52**, 654-661.

Stevens, L. C. (1967). Origin of testicular tumours from primordial germ cells in mice. *J. Natl. Cancer Inst.* **38**, 549-552.

Stevens, L. C. (1968). The development of teratomas from intratesticular grafts of tubal mouse eggs. *J. Embryol. exp. Morph.* **20**, 329-341.

Stevens, L. C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Developmental Biology* **21**, 364-382.

Stewart, C. L., Gadi, I. and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. *Dev Biol* **161**, 626-8.

Stewart, C. L., Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Kontgen, F. and Abbondanzo, S. J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-9.

Stewart, T. A. and Mintz, B. (1981). Successive generations of mice produced from an established culture line of euploid teratocarcinoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6314-6318.

Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophoctoderm in the mouse blastocyst. *Development* **132**, 2093-2102.

Sutherland, A. E., Speed, T. P. and Calarco, P. G. (1990). Inner cell allocation in the mouse morula: The role of oriented division during fourth cleavage. *Developmental Biology* **137**, 13.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodriguez-Esteban, C. and Belmonte, J. C. (2006a). Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. *Nat Clin Pract Cardiovasc Med* **3 Suppl 1**, S114-22.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodriguez-Esteban, C. and Izpisua Belmonte, J. C. (2006b). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. *Proceedings of the National Academy of Sciences* **103**, 10294-10299.

Tada, M., Tada, T., Lefebvre, L., Barton, S. C. and Surani, M. A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *Embo Journal* **16**, 6510-6520.

Tada, T., Tada, M., Hilton, K., Barton, S. C., Sado, T., Takagi, N. and Surani, M. A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. *Development Genes And Evolution* **207**, 551-561.

Taga, T., Hibi, M., Hirata, Y., Yamasaki, K., Yasukawa, K., Matsuda, T., Hirano, T. and Kishimoto, T. (1989). Interleukin-6 Triggers The Association Of Its Receptor With A Possible Signal Transducer, Gp130. *Cell* **58**, 573-581.

Takada, T., Nemoto, K.-i., Yamashita, A., Kato, M., Kondo, Y. and Torii, R. (2005). Efficient gene silencing and cell differentiation using siRNA in mouse and monkey ES cells. *Biochemical and Biophysical Research Communications* **331**, 1039.

Takahashi, K. and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663.

Takaoka, K., Yamamoto, M., Shiratori, H., Meno, C., Rossant, J., Saijoh, Y. and Hamada, H. (2006). The Mouse Embryo Autonomously Acquires Anterior-Posterior Polarity at Implantation. *Developmental Cell* **10**, 451.

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T. and Akira, S. (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc Natl Acad Sci U S A* **94**, 3801-4.

Tam, A. W., Smith, M. M., Fry, K. E. and Larrick, J. W. (1989). Construction of cDNA libraries from small numbers of cells using sequence independent primers. *Nucleic Acids Res* **17**, 1269.

Tanaka, T. S., Jaradat, S. A., Lim, M. K., Kargul, G. J., Wang, X., Grahovac, M. J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R. et al. (2000). Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *PNAS* **97**, 9127-9132.

Tanaka, T. S. and Ko, M. S. (2004). A global view of gene expression in the preimplantation mouse embryo: morula versus blastocyst. *Eur J Obstet Gynecol Reprod Biol* **115 Suppl 1**, S85-91.

Tanaka, T. S., Kunath, T., Kimber, W. L., Jaradat, S. A., Stagg, C. A., Usuda, M., Yokota, T., Niwa, H., Rossant, J. and Ko, M. S. H. (2002). Gene Expression Profiling of Embryo-Derived Stem Cells Reveals Candidate Genes Associated With Pluripotency and Lineage Specificity. *Genome Res.* **12**, 1921-1928.

Tarkowski, A. K., Ozdzinski, W. and Czolowska, R. (2005). Identical triplets and twins developed from isolated blastomeres of 8- and 16-cell mouse embryos supported with tetraploid blastomeres. *Int J Dev Biol* **49**, 825-32.

Tesar, P. J. (2005). Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proceedings of the National Academy of Sciences* **102**, 8239-8244.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196.

Thomas, K. R. and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503.

- Thomas, P. Q., Brown, A. and Beddington, R. S.** (1998). Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* **125**, 85-94.
- Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M. and Okuda, A.** (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucl. Acids Res.* **30**, 3202-3213.
- Vallier, L., Alexander, M. and Pedersen, R. A.** (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *Journal Of Cell Science* **118**, 4495-4509.
- Vallier, L., Reynolds, D. and Pederson, R. A.** (2004). Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Developmental Biology* **275**, 403-421.
- van Belkum, A., Linkels, E., Jelsma, T., van den Berg, F. M. and Quint, W.** (1994). Non-isotopic labeling of DNA by newly developed hapten-containing platinum compounds. *Biotechniques* **16**, 148-53.
- Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D. and Eberwine, J. H.** (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A* **87**, 1663-7.
- Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W.** (1995). Serial analysis of gene expression. *Science* **270**, 484-7.
- Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. and Marincola, F. M.** (2000). High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* **18**, 457-9.
- Wang, Q. T., Piotrowska, K., Ciemerych, M. A., Milenkovic, L., Scott, M. P., Davis, R. W. and Zernicka-Goetz, M.** (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev Cell* **6**, 133-44.
- Wang, S. H., Tsai, M. S., Chiang, M. F. and Li, H.** (2003). A novel NK-type homeobox gene, ENK (early embryo specific NK), preferentially expressed in embryonic stem cells. *Gene Expression Patterns* **3**, 99-103.
- Ware, C. B., Horowitz, M. C., Renshaw, B. R., Hunt, J. S., Liggitt, D., Koblar, S. A., Gliniak, B. C., McKenna, H. J., Papayannopoulou, T., Thoma, B. et al.** (1995). Targeted disruption of the low-affinity leukemia inhibitory factor receptor gene causes

placental, skeletal, neural and metabolic defects and results in perinatal death. *Development* **121**, 1283-99.

Wells, D. N., McWhir, J., Hooper, M. L. and Wilmut, I. (1991). Factors influencing the isolation of murine embryonic stem cells. *Theriogenology* **35**, 293.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E. and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-7.

Willison, K. R. and Stern, P. L. (1978). Expression of a Forssman antigenic specificity in the preimplantation mouse embryo. *Cell* **14**, 785.

Yang, Y. H., Buckley, M. J., Dudoit, S. and Speed, T. P. (2002a). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics* **11**, 108-136.

Yang, Y. H., Buckley, M. J. and Speed, T. P. (2001). Analysis of cDNA microarray images. *Brief Bioinform* **2**, 341-9.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. and Speed, T. P. (2002b). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**, e15.

Yeakley, J. M., Fan, J. B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M. S. and Fu, X. D. (2002). Profiling alternative splicing on fiber-optic arrays. *Nat Biotechnol* **20**, 353-8.

Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Scholer, H. R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-894.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. and Smith, A. (2003a). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotech* **21**, 183.

Ying, Q. L., Nichols, J., Chambers, I. and Smith, A. (2003b). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-92.

Ying, Q. L. and Smith, A. G. (2003). Defined conditions for neural commitment and differentiation. *Methods Enzymol* **365**, 327-41.

Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T. and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech Dev* **45**, 163-71.

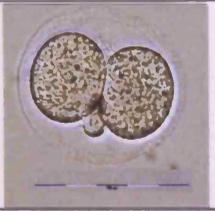
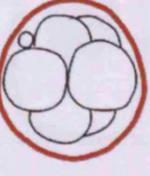
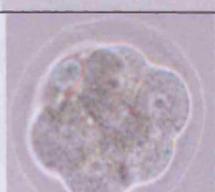
Yoshida, K., Taga, T., Saito, M., Suematsu, S., Kumanogoh, A., Tanaka, T., Fujiwara, H., Hirata, M., Yamagami, T., Nakahata, T. et al. (1996). Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 407-411.

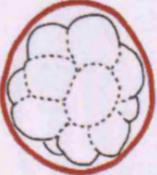
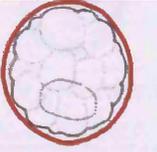
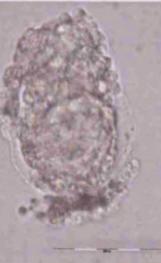
Zeng, F. and Schultz, R. M. (2005). RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. *Dev Biol* **283**, 40-57.

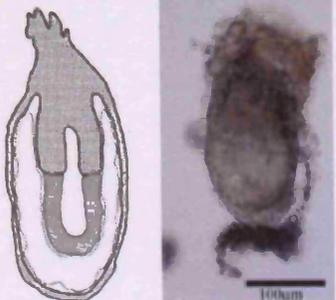
Zhou, X., Sasaki, H., Lowe, L., Hogan, B. L. M. and Kuehn, M. R. (1993). Nodal is a novel TGF-[beta]-like gene expressed in the mouse node during gastrulation. *Nature* **361**, 543.

Ziomek, C. A., Johnson, M. H. and Handyside, A. H. (1982). The developmental potential of mouse 16-cell blastomeres. *J Exp Zool* **221**, 345-55.

Appendix 1 Developmental Stages

Embryonic developmental Stage		Time of isolation in days post coitum to nearest half day	Time of isolation in hours post coitum	Morphological Markers	Use
Oocytes		0.5	6-8	1-cell, zona pellucida, 1 st polar body only.	Reference pool only.
Fertilised eggs 1-cell	 	0.5	8-10	1-cell, zona pellucida, 1 st and 2 nd polar body.	Reference pool
2 cell	 	1.5	32-35	2-cell, zona pellucida, 1 st and 2 nd polar body.	Reference pool
3-4 cell	 	2.0	45-46	4-cells, zona pellucida, 1 st and 2 nd polar body. 1 st polar body may have degenerated	Reference pool
4-8 cell	 	2.5	55-58	Anywhere between 5 and 8 cells. Later embryos with may have started compaction. 1 st polar body may have degenerated	Reference pool

Morula, 8-16 cell	 	3	67-68	Compacted morula, between 8 and 16 cells, exact number unknown due to compaction, zona pellucida	Reference pool
early blastocysts	 	3.5	80-84	Zona pellucida , blastocoel cavity half to fully expanded.	Intact embryos in reference pool. Isolated ICM in samples.
Late blastocysts	 	4.5	103-107	Hatched (no zona pellucida) often collapsed and with maternal cells adhering to distal mural trophoctoderm.	Isolated ICM in samples.
Early post-implantation	 	5.5	132-135	Early post-implantation. Reichert's membrane with maternal cells. Early segmentation of egg cylinder. Early proamniotic cavity not extending into extra-embryonic ectoderm.	Isolated epiblast in samples

Egg Cylinder		6.5	150-154	Segmented egg cylinder. Reichert's membrane with maternal cells. Pro-amniotic cavity. Ectoderm pseudo stratified epithelium.	Isolated embryonic ectoderm in samples.
--------------	---	-----	---------	--	---

Appendix 1: Stages of embryonic development. The drawings were taken from EMAP; the Edinburgh Mouse Atlas Project, <http://genex.hgu.mrc.ac.uk/intro.html>

Appendix 2. Formulae used in Analysis of Labelled Target

Beer's Law $A(x) = \epsilon \text{cm}^{-1} \text{mol}^{-1} * \text{pathlength}(\text{cm}) * \text{concentration} C(\text{mol/L})$,

where A is the absorbance at a wavelength x(nm), ϵ is the extinction coefficient for the substance being analysed ($\text{lcm}^{-1} \text{mol}^{-1}$).

The pathlength in a standard cuvette is 1cm $C(\text{mol/L}) = A/\epsilon$ mol of dye incorporated = $A/\epsilon * \text{vol}(\text{L})$

pmol of dye incorporated = $A/\epsilon * \text{vol}(\mu\text{l}) * 10^{-6}$

for Cy3TM, $\epsilon=150,000$ and absorbance maximum is 550nm

pmol of dye incorporated = $A_{(550)}/0.15 * \text{vol}(\mu\text{l})$

and for Cy5TM, $\epsilon= 250,000$ and absorbance maximum is 650nm

pmol of dye incorporated = $A_{(650)}/0.25 * \text{vol}(\mu\text{l})$

The average molecular weight of a deoxynucleotide is 324.5 and assumes average cDNA is 1kb

$\epsilon=8919$ for single stranded DNA at the absorbance maximum of 260nm

concentration of single stranded DNA = $A_{260} * 324000 / 8919 \text{g/l}$

$$= A_{260} * 37 \mu\text{g/ml}$$

Frequency of incorporation per 1000 nucleotides (FOI)

For Cy3 FOI= ng. Cy3TM/ng cDNA = $A_{550}/\epsilon_{550} * 324 / A_{260} * 37$

$$= A_{550}/A_{260} * 58.5$$

For Cy5TM FOI = $(A_{650}/\epsilon_{650} * 324) / A_{260} * 37$

$$= A_{650}/A_{260} * 35.$$

Appendix 3: Commands for Analysis in Limma

This is a protocol for the analysis of files into Limma and uses the files from the DNase experiment described in Chapter 5. The typed commands are in red. The base package R and limma were downloaded from the web.

```
> library(limma)
```

```
> limmaUsersGuide()
```

```
[1] "C:/PROGRA~1/R/rw2011/library/limma/doc/usersguide.pdf"
```

Go to file pulldown menu and select Change Directory. Go to where your files for limma analysis are stored.

```
> spottypes <- readSpotTypes()
```

```
> objects()
```

```
[1] "spottypes" "targets"
```

```
> show(spottypes)
```

	SpotType	Gene ID	Color
1	gene	*	black
2	landmark	Landmark	green
3	scorecard	Scorecard*	red
4	Rich	Rich*	blue
5	BLANK	BLANK	yellow
6	empty	empty	purple

```
7 null orange
```

```
> targets<-readTargets()
```

```
> show(targets)
```

SlideNumber	FileNameCy3
1	11 Cyanine3_052405_1119_Exp4a15pmol65pmt.txt
2	12 Cyanine3_052405_1139_Exp4a15pmol65pmt.txt
3	13 Cyanine3_060305_1113_Exp4a15pmol65pmt.txt
4	14 Cyanine3_060305_1137_Exp4a15pmol65pmt.txt

SlideNumber	FileNameCy5	Cy3	Cy5
1	Cyanine5_052405_1119_Exp4a15pmol65pmt.txt	Dnase control	
2	Cyanine5_052405_1139_Exp4a15pmol65pmt.txt	Dnase control	
3	Cyanine5_060305_1113_Exp4a15pmol65pmt.txt	control	Dnase
4	Cyanine5_060305_1137_Exp4a15pmol65pmt.txt	control	Dnase

```
> files <- targets[,c("FileNameCy3","FileNameCy5")]
```

```
> RG <- read.maimages(files, source="imagene")
```

```
Read header information
```

```
Read Cyanine3_052405_1119_Exp4a15pmol65pmt.txt
```

```
Read Cyanine5_052405_1119_Exp4a15pmol65pmt.txt
```

```
Read Cyanine3_052405_1139_Exp4a15pmol65pmt.txt
```

```
Read Cyanine5_052405_1139_Exp4a15pmol65pmt.txt
```

```
Read Cyanine3_060305_1113_Exp4a15pmol65pmt.txt
```

```
Read Cyanine5_060305_1113_Exp4a15pmol65pmt.txt
```

```
Read Cyanine3_060305_1137_Exp4a15pmol65pmt.txt
```

```
Read Cyanine5_060305_1137_Exp4a15pmol65pmt.txt
```

```
> RG$genes$Status <- controlStatus(spottypes, RG)
```

Matching patterns for: Gene ID

Found 16128 gene

Found 48 landmark

Found 230 scorecard

Found 48 Rich

Found 10 BLANK

Found 65 empty

Found 480 null

Setting attributes: values Color

An imageplot of the background intensities allows one to visualise any defects in the slide.

```
> imageplot(log2(RG$Rb[,1]), RG$printer, low="white", high="red")
```

```
> imageplot(log2(RG$Gb[,1]), RG$printer, low="white", high="green")
```

The MA plot of each slide is also useful particularly for visualising intensity bias on the array.

```
> plotMA(RG,array=1)
```

The scale on the x and y can be set using:

```
> plotMA(RG,array=1,xlim=c(4,16),ylim=c(-3,3))
```

otherwise limma will set to fit data. Lines showing the unregulated $M=0$ and 2-fold regulated cut off can be added by:

```
> abline(0,0,col="blue")
```

```
> abline(-1,0,col="red")
```

```
> abline(1,0,col="red")
```

A plot of printtiploess loess shows the loess curves for each print tip

```
> plotPrintTipLoess(RG,array=1)
```

Background correction and Normalisation

A) Normalise within arrays

1 No background correction, print tip loess normalisation

```
> RGb1<-backgroundCorrect(RG,method="none")
```

```
> Norm1<-normalizeWithinArrays(RGb1)
```

```
> plotMA(Norm1,array=1)
```

2 Background correction with "normexp" function, print tip loess normalisation

```
> RGb2<-backgroundCorrect(RG,method="normexp",offset=50)
```

```
> Norm2<-normalizeWithinArrays(RGb2)
```

```
> plotMA(Norm2,array=1)
```

3 Background correction with "normexp" function, robust spline normalisation

```
> RGb2<-backgroundCorrect(RG,method="normexp",offset=50)
```

```
> Norm3<-normalizeWithinArrays(RGb2,method="robustspline")
```

```
> plotMA(Norm3,array=1)
```

B) Normalise between arrays

Boxplot and Density plot to display arrays using Norm2

```
> boxplot(Norm2$M~col(Norm2$M)) or
> boxplot(Norm2$M~col(Norm2$M),ylim=c(-4,3))
> plotDensities(Norm2)
```

1 Scale

```
> Norm4 <- normalizeBetweenArrays(Norm2)
> boxplot(Norm4$M~col(Norm4$M))
> plotDensities(Norm4)
```

2 Aquantile

```
> Norm5Aq <- normalizeBetweenArrays(Norm2, method="Aquantile")
> plotDensities(Norm5Aq)
> boxplot(Norm5Aq$M~col(Norm5Aq$M))
```

3 Quantile

```
> Norm6q <- normalizeBetweenArrays(Norm2, method="quantile")
> plotDensities(Norm6q)
> boxplot(Norm6q$M~col(Norm6q$M))
```

Make the linear model

```
> design <- modelMatrix(targets,ref="control")
```

show the design

```
> design
```

For the experiment described in Chapter 5, this should be

Dnase

1 -1

2 -1

3 1

4 1

The linear model is fitted to the data and the statistics calculated.

```
> fit <- lmFit(Norm5Aq, design)
```

```
> fit <- eBayes(fit)
```

The output is read out with the decimal places set to three. The number= argument can be omitted if the top ten is required as 10 is the default.

```
> options(digits=3)
```

```
> topTable(fit, number=10, adjust="fdr")
```

Appendix 4

Protocol for Analysis of

Microarray Data using the NIA

Anova Package

Appendix 4 Analysis of array data using NIA tool

The scanned images were stored as pairs of 16 bit tiff files. These were analysed using the program “Spot” from CSIRO following the instructions provided (<http://www.cmis.csiro.au/iap/Spot/spotmanual.htm>). The program was implemented on a LINUX machine with 4G RAM and run under R. Image files were analysed in batches corresponding to their scanning batch.

The data output as “.spot” files were imported into Limma and normalised within arrays by “Print tip group loess” and between arrays by “scale”.

Data were exported as M and A values. M is defined as $\log_2(R/G)$ and A as $0.5(\log_2(R) + \log_2(G))$ where R is Cy5™ and G is Cy3™.

$$M = \log_2 R - \log_2 G$$

$$A = 0.5(\log_2 R + \log_2 G)$$

$$\log_2 G = \log_2 R - M$$

$$\log_2 G = 2A - \log_2 R$$

$$2\log_2 G = 2A - M$$

$$\log_2 G = A - M/2$$

$$\log_2 R = M + \log_2 G$$

$$\log_2 R = 2A - \log_2 G$$

$$2\log_2 R = 2A + M$$

$$\log_2 R = A + M/2$$

Appendix 4: Protocol for Analysis using NIA Anova

These values were then assembled into a tab delimited text file for entry into NIA anova. (<http://lgsun.grc.nia.nih.gov/ANOVA/help.html#format> *The first row consists of column headers. The first column has gene IDs which can be numerical or textual.*)

For input into Cluster the ratio was calculated for each spot as $2^{(\log_2(G) - \log_2(R))}$ in a tab delimited file.

Annotation was taken from the file NIA-15k-GeneID.tsv

The following pages show the settings used for the NIA anova analysis

NIA Array Analysis

Developed at the [National Institute on Aging \(NIA/NIH\)](#), Laboratory of Genetics, Baltimore MD

- False discovery rate (FDR)
- ANOVA with error variance correction
- 3D PCA/SVD-biplot
- PCA import for experiment comparison
- Pattern matching
- Optional permutation test
- Server-based software

Public Data Sets

Description of NIA Array Analysis tool

- [General description](#)
- [Input data format](#)
- [Gene annotations](#)
- [ANOVA](#)
- [Hierarchical clustering](#)
- [Principal Component Analysis \(PCA/SVD/biplot\)](#)
- [Pattern matching](#)

Related links

- [NIA Mouse Gene Index](#)
- [NIA 44K mouse array](#)
- [NIA Mouse cDNA project](#)
- [SAM \(significance analysis of microarrays\)](#)
- [TIGR array analysis software](#)
- [Microarray Data Analysis resources](#)

Reference:
Sharov, A.A., Dudekula, D.B., Ko, M.S.H. 2005. A web-based tool for principal component and significance analysis of microarray data. Bioinformatics (submitted)

User login:
Password:

You can log in as "guest" or [Sign in](#) to get a password (no restriction)

Report problems to Alexei Sharov (sharoval@mail.nih.gov)

Done



NIA Array Analysis Tool

HELP: [General description of NIA Array Analysis Tool](#)

[Arrayjoin Tool](#) (Windows) for building the input file from multiple scanner files ([DOWNLOAD](#))

Upload Data File: [Browse...](#) [Upload](#) [see file format](#) [Normalize](#) [Split-normalize](#)
 upload as: Data name (use 1-word; if blank, = filename)
 None Log-transformation in input file Computer guess Lower cutoff value

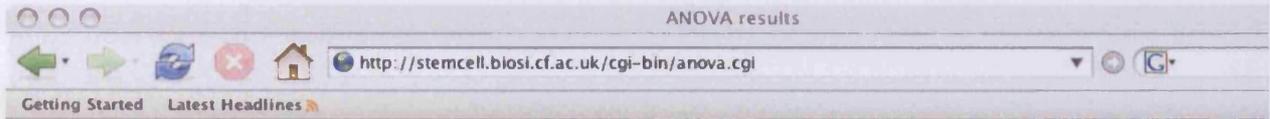
Select data file for analysis

<input type="text" value="2802new"/>	Description: <input type="text" value="2802new"/>	View data	Delete file
<input type="text" value="Analysis"/>	Array annotation	View array	Add array
<input type="text" value="annotationforanova"/>	Threshold z-value to remove outliers		
<input type="text" value="8"/>	Error model		
<input type="text" value="Bayesian"/>	Size of sliding window for averaging error variances		
<input type="text" value="50"/>	Proportion of highest variance values to be removed before variance averaging		
<input type="text" value="0.05"/>	Desirable degrees of freedom for Bayesian error model		
<input type="text" value="10"/>	FDR false discovery rate threshold		
<input type="text" value="0.01"/>	Number of permutations		
<input type="text" value="100"/>	Number of colors in the array		
<input type="text" value="2-color arrays"/>	Minimum log-intensity cutoff (-9999 = none)		
<input type="text" value="-9999"/>	Maximum log-intensity cutoff (-9999 = none)		
<input type="text" value="-9999"/>	Include reference into ANOVA analysis		
<input type="text" value="don't analyze reference"/>	Reference name		
<input type="text" value="control"/>	No cross-channel correction		
<input type="checkbox"/>			

Table of tissues/experiments (see [help](#)) [Show design](#) N data columns=

#	No. of re- plications	Name of tissue/experiment
1	4	ICM_88hpc
2	6	ICM_105hpc
3	6	DICM_136hpc
4	5	DICM180hpc
5	5	EmEct_Dy6.5
6	6	EmbEctDy5.5
7	6	IMT_11
8		
9		

Appendix 4: Protocol for Analysis using NIA Anova



Results

[Download](#)

[Download](#)

[Download](#)

Results of ANOVA

Input file

Means (significant genes only)

[TIGR-MEV](#) file (Stanford format)

Search Term:

In:

Go [Cluster tissues](#) Find specific genes

Significantly specific

Minimum fold change:

Go [Plot the error function](#)

Go [Plot the dendrogram for replications](#)

Go [Make correlation matrix](#)

Principal Component Analysis (PCA)

Covariance Matrix type for PCA Center values

Select the number of principal components

Fold-change threshold for clusters

Correlation threshold for clusters

Genes to analyze

Gene number to display

PCA by

Pair-wise comparison of means

Compare with

Use Use and fold-change threshold

Pattern-matching

Type (or paste) a probe, OR a column of probes, OR a column of gene symbols (if present in the annotation file), OR a table with 2 columns: tissue_name and expression (separated by tabs or "=")
If multiple probes, then the pattern is averaged. Only exact matches are found.
If a gene has multiple gene symbols, only the first one is considered.

Plot genes listed above only

