<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Organisation and function of the Nucleolar Organiser Regions in human chromosomes.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Floutsakou, Ioanna</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2013-12-20</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/4160">http://hdl.handle.net/10379/4160</a></td>
</tr>
</tbody>
</table>
“Organisation and function of the Nucleolar Organiser Regions in human chromosomes.”

Ioanna Floutsakou

Centre for Chromosome Biology, School of Natural Sciences, National University of Ireland. Galway

A thesis submitted to the National University of Ireland, Galway for the degree of Doctor of Philosophy

September 2013

Supervisor: Professor Brian McStay
## Contents

List of Contents 1
Acknowledgements 5
List of figures and tables 6
Abbreviations 9
Abstract 14

### 1 Introduction

1.1 The nucleolus and its importance 16
1.2 Discovery and initial study of the nucleolous 16
1.3 Ribosomal genes and Nucleolar Organiser Regions (NORs) 17
1.4 Active and inactive NORs 20
1.5 The proportion of active repeats in the NORs 22
1.6 Ribosomal gene transcription 24
1.7 The basal human RNA polymerase I transcription initiation factors 25
1.8 Regulation of rDNA transcription 28
   1.8.1 Cell cycle regulation of rDNA transcription 28
   1.8.2 Growth factor dependent regulation of rDNA transcription 29
   1.8.3 RNA Pol I transcription responds to genotoxic stress 30
   1.8.4 Cancer and regulation of rDNA transcription 31
   1.8.5 Tissue specific regulation of rDNA transcription 32
1.9 Elongation and termination of transcription 33
1.10 Post-transcriptional steps in ribosome biogenesis 35
1.11 General structural features of the nucleolus 36
1.12 The nucleolar proteome 41
1.13 Ribosomal gene chromatin 43
1.14 Silencing of ribosomal genes 45
1.15 Thesis Aim 48

### 2 Materials and Methods

2.1 DNA manipulation
   2.1.1 Plasmid DNA Purification from small cultures 50
   2.1.2 Plasmid DNA purification from large cultures 50
   2.1.3 Glycerol stocks 50
   2.1.4 Determining the concentration and purity of nucleic acids 50
   2.1.5 DNA sequencing 51
   2.1.6 Agarose Gel Electrophoresis 51
   2.1.7 Acrylamide Gel Electrophoresis 51
   2.1.8 Restriction Digests 52
   2.1.9 Extraction of DNA from agarose gels 52
2.1.10 Ligation of DNA fragments 52
2.1.11 Transformation into competent cells 52
2.1.12 Preparation of competent cells 53
2.1.13 PCR 54
2.1.14 Real Time PCR 54
2.1.15 Quantitation of PCR/protein products 55
2.1.16 Purification of PCR products 55
2.1.17 Extraction of genomic DNA from cultured human cells 55

2.2 RNA
2.2.1 RNA extraction from cultured human cells 56
2.2.2 cDNA synthesis 56

2.3 Tissue culture
2.3.1 Cell Lines 57
2.3.2 Maintaining cell lines 58
2.3.3 Generating the DJ clones stable cell line 59
2.3.4 Long term storage and recovery of cell lines from Liquid N₂ 59
2.3.5 Transfections 60
2.3.6 siRNA transfections 60

2.4 Cell Biology Techniques
2.4.1 Immunofluorescence on interphase cells 61
2.4.2 DNA FISH on interphase cells 61
2.4.3 Combined immunofluorescence and FISH on interphase cells 62
2.4.4 Preparing metaphase spreads 62
2.4.5 DNA FISH on metaphase spreads 63
2.4.6 RNA FISH and combined immunofluorescence 63
2.4.7 Combined FISH and silver staining on metaphase spreads 63
2.4.8 Imaging of fixed cells 64

2.5 Nucleolar pellets
2.5.1 Nucleolar isolation from cultured cells 64
2.5.2 Nucleolar DNA combing 65
2.5.3 Preparation of soluble formaldehyde cross-linked nucleolar chromatin 66
2.5.4 Nucleolar chromatin immunoprecipitation (ChIP) 67
2.5.5 Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) 68

2.6 Protein
2.6.1 Harvesting protein samples from cultured cells 70
2.6.2 Western Blotting 71
2.6.3 Protein concentration detection 71
2.6.4 Expression and purification of 6xHis-tagged proteins 72
2.6.5 Affinity purification of Antibodies 75
3 Bromodomain adjacent to zinc finger domain protein 2A (BAZ2A)
3.1 Background 78
3.2 Results
3.2.1 Generate BAZ2A antibody that can be used in both Immuno-FISH and Western Blot experiments 82
3.2.2 α-BAZ2A Ab works both in immunofluorescence and immunoblotting 84
3.2.3 Silencing of the BAZ2A protein in cells to test antibody’s efficiency and whether its effect would be the reactivation of silent NORs 86
3.3 Discussion 92

4 Mapping of the distal and proximal to the rDNA sequences, termed Distal (DJ) and Proximal (PJ) Junctions
4.1 Background 97
4.2 Results
4.2.1 Confirmation of the DJ and PJ sequences are conserved among the acrocentrics in position to the rDNA 105
4.2.2 Explanation for the hybridization pattern of the PJ BACs on the chromosomal metaphases 109
4.2.3 CER satellite 111
4.2.4 Extension of the known DJ sequence with a new identified BAC 112
4.3 Discussion 117

5 Positioning of the Distal (DJ) and Proximal (PJ) Junctions
5.1 Background 126
5.2 Results
5.2.1 DJ and PJ positioning in relation to the nucleolus and the nucleolar caps 128
5.2.2 Ectopic DJ BAC arrays 135
5.3 Discussion 149

6 Chromatin and transcription profiling of the distal junction
6.1 Background 154
6.2 Results
6.2.1 Chromatin profiling of the DJ using nucleolar ChIP and nucleolar FAIRE 156
6.2.2 Transcription profiling 164
6.2.3 3D-RNA FISH 169
6.2.4 Over-expression studies 172
6.3 Discussion 177

7 Final Discussion
Final conclusions and future research 183

References 188
Acknowledgements

I would like to express my gratitude to my supervisor Prof. Brian McStay for the continuous support of my Ph.D study and research, his patience, guidance, motivation, enthusiasm and immense knowledge. You kept pushing me to my limits so I always had to become better, you challenged me and taught me to be strong but at the same time flexible and adaptable. It’s been a wonderful adventure indeed boss.

I thank my fellow lab mates, Jose, Christine, Alice, Mayo, Sofia and Shelly for all their help in the lab but also for their friendship. You’ve made the lab a fun place to work and were always there for me through good times and bad. I’m going to miss you all. I would like to thank my thesis committee Prof. Kevin Sullivan, Dr. Adrian Bracken and Dr. Andrew Flaus for their valuable comments and corrections for the finalization of this thesis. A special thanks goes to Ms Colette McGuinness for her help and support on the submission of my thesis draft. I would also like to thank Dr. George Patrinos and Dinos Bizovitis who have supported and inspired me before and during my Ph.D.

Mike Burke, Paddy Collins, Nikos Palamaris and Lida Laxanokardi I can’t thank you enough for your friendship, support, advice, shoulder to cry on, warm hug, making me laugh when all I wanted to do was cry, listening to my rants, worries and stories, keeping me sane and always being there for me.

Anthony O’Neill, you were God’s gift to me in the most difficult time of my life. You brought the smile back to my face, you are my rock, my teammate and made all this worth it. You are always there to pick up my pieces, put me back together and give me wings to fly again. I couldn’t have done it without you and your love. Thank you a chuisle mo chroi.

Last but not least, I would like to thank my parents Agathi and Apostolos. I owe you not only my existence but everything I am today. You provided for me, protected me, taught me to stand on my own two feet, fight for what I want and what I believe in, work hard but also live life to the fullest, always be kind, ethical and never do wrong to anyone and then you set me free and let me fly away from the nest to find my own way in this world but you never let me feel alone or unprotected, not even for one second. I would like to dedicate this thesis to you for your constant support and endless unconditional love and our late dog Sahara, who was my most loyal, unselfish and kind friend and companion for half my present life.
List of figures and tables

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Components of the rDNA repeat</td>
<td>18</td>
</tr>
<tr>
<td>1.2</td>
<td>Active versus inactive NOR</td>
<td>21</td>
</tr>
<tr>
<td>1.3</td>
<td>Photo of a Miller spreads</td>
<td>26</td>
</tr>
<tr>
<td>1.4</td>
<td>RNA Polymerase I pre-initiation complex (PIC)</td>
<td>27</td>
</tr>
<tr>
<td>1.5</td>
<td>Nucleolar Compartments</td>
<td>38</td>
</tr>
<tr>
<td>1.6</td>
<td>The Nucleolar Remodelling Complex (NoRC)</td>
<td>47</td>
</tr>
<tr>
<td>3.1</td>
<td>Locations of the motif sequences in BAZ2A and TIP5</td>
<td>80</td>
</tr>
<tr>
<td>3.2</td>
<td>Production of recombinant BAZ2A and antibody Characterization</td>
<td>83</td>
</tr>
<tr>
<td>3.3</td>
<td>Immunostaining using BAZ2A antibody (purified and crude serum) reveals BAZ2A protein is not nucleolar</td>
<td>85</td>
</tr>
<tr>
<td>3.4</td>
<td>RNAi mediated depletion of the BAZ2A protein in HT1080 cells</td>
<td>86</td>
</tr>
<tr>
<td>3.5</td>
<td>RNAi mediated depletion of the BAZ2A protein in HeLa cells</td>
<td>87</td>
</tr>
<tr>
<td>3.6</td>
<td>RNAi mediated depletion of the BAZ2A protein in HeLa cells does not have any observable effect on the nucleolar structure</td>
<td>88</td>
</tr>
<tr>
<td>3.7</td>
<td>Immunostaining validates depletion of BAZ2A in siRNA transfected HeLa cells</td>
<td>89</td>
</tr>
<tr>
<td>3.8</td>
<td>Immuno-FISH showing BAZ2A depletion does not affect Pol I transcription in the nucleolus</td>
<td>90</td>
</tr>
<tr>
<td>3.9</td>
<td>Immuno-FISH showing BAZ2A depletion does not cause inactive NOR re-activation</td>
<td>91</td>
</tr>
<tr>
<td>4.1</td>
<td>Mapping of the human rDNA flanking regions</td>
<td>99</td>
</tr>
<tr>
<td>4.2</td>
<td>DNA combing of the human rDNA and DJ flanking region</td>
<td>99</td>
</tr>
<tr>
<td>4.3</td>
<td>Positions of BACs and cosmids mapped on the DJ</td>
<td>100</td>
</tr>
<tr>
<td>4.4</td>
<td>Positions of BACs and cosmids mapped on the PJ</td>
<td>101</td>
</tr>
<tr>
<td>4.5</td>
<td>DJ and PJ sequences are conserved among the acrocentric chromosomes</td>
<td>102</td>
</tr>
<tr>
<td>4.6</td>
<td>PJ flanking region hybridization observed on the DJ side of the chromosomes as well</td>
<td>102</td>
</tr>
<tr>
<td>4.7</td>
<td>BACs used to construct DJ and PJ contigs</td>
<td>104</td>
</tr>
<tr>
<td>4.8</td>
<td>Mapping of the DJ flanking region by FISH on metaphase Spreads of HT1080 cells</td>
<td>105</td>
</tr>
<tr>
<td>4.9</td>
<td>PJ hybridization found on the DJ side possible explanations</td>
<td>106</td>
</tr>
<tr>
<td>4.10</td>
<td>Locations of regions on the acrocentric chromosomes used as probes</td>
<td>107</td>
</tr>
<tr>
<td>4.11</td>
<td>FISH analysis of chromosomes in metaphase from the GM somatic cell hybrids</td>
<td>108</td>
</tr>
<tr>
<td>4.12</td>
<td>Validation of DJ/PJ content of GM monochromosomal cell Hybrids</td>
<td>109</td>
</tr>
<tr>
<td>4.13</td>
<td>Molecular DNA combing</td>
<td>110</td>
</tr>
<tr>
<td>4.14</td>
<td>DNA combing of the human rDNA</td>
<td>110</td>
</tr>
</tbody>
</table>
Figure 4.15 DNA combing of the human rDNA and PJ flanking region
Figure 4.16 48bp CER repeats are found distal to the NOR on all five acrocentric chromosomes
Figure 4.17 Genomic organisation of BAC AL591856
Figure 4.18 BAC AL591856 validation of sequence content from PCR on monochromosomal cell hybrids
Figure 4.19 Schematic representation of the known DJ sequence to Date
Figure 4.20 FISH of normal human chromosomes in metaphase with BAC AL591856 and rDNA probes
Figure 4.21 FISH of normal human chromosomes in metaphase with BAC AL591856 and acrocentric specific probes
Figure 4.22 Cosmids and BACs spanning the DJ-rDNA and PJ-rDNA junctions
Figure 4.23 Conservation of DJ and PJ among the acrocentric chromosomes
Figure 4.24 DJ inverted repeat arms similarity plot
Figure 4.25 Segmental duplication analyses for PJ and DJ
Figure 5.1 PJ localization is more complex than DJ
Figure 5.2 DJ localizes to the nucleolar periphery
Figure 5.3 Evidence that the DJ localizes to the nucleolar periphery
Figure 5.4 DJ perinucleolar localization in HT1080 cells
Figure 5.5 DJ seems to act as a perinucleolar anchor for the rDNA repeats in HT1080 cells
Figure 5.6 DJ seems to act as a perinucleolar anchor for the rDNA repeats in HeLa cells
Figure 5.7 The AL591856 BAC is positioned immediately adjacent to the DJ
Figure 5.8 Positions of BACs relative to the DJ contig used for construction of exogenous DJ BAC arrays
Figure 5.9 DJ arrays have integrated into non-NOR bearing metacentric chromosomes
Figure 5.10 DJBAC3 clone does not contain any active rDNA repeats
Figure 5.11 Ectopic DJ arrays target peri-nucleolar Heterochromatin
Figure 5.12 Positions of BACs relative to the DJ contig used for construction of individual exogenous DJ BAC arrays
Figure 5.13 Clones of individual or coupled BAC insertions show Variable positioning of the exogenous DJs
Figure 5.14 DJ arrays have integrated into non-NOR bearing metacentric chromosomes apart from clone LA14 138F-10-9
Figure 5.15 Clone LA14 138F-10-5 contains unidentified duplicated active NOR entities
Figure 5.16 The unidentified entities of clone LA14 138F-10-5 do not have a centromere
Figure 6.1 Chromatin landscape of DJ
Figure 6.2  H3K4me3 peaks overlap with open chromatin domains identified by DNaseI-His-seq and FAIRE-seq  

Figure 6.3  Nucleolar H3K4me3 ChIP-PCR validates the presence of H3K4me3 in the DJ  

Figure 6.4  Nucleolar FAIRE-PCR validates the presence of open chromatin in the DJ  

Figure 6.5  Nucleolar Pol II ChIP-PCR validates the presence of Pol II in the DJ  

Figure 6.6  Transcription profiling of the DJ  

Figure 6.7  DJ transcript profiling  

Figure 6.8  Nucleolar localization of the disnor transcripts  

Figure 6.9  Disnor238 transcript probe location  

Figure 6.10  DJ transcripts accumulate in perinucleolar foci  

Figure 6.11  Validation of RNA FISH detection method  

Figure 6.12  DJ transcripts exhibit stronger transcript signals in DJBAC clone 2  

Figure 6.13  Validation of RNA FISH detection method in DJBAC clone 2  

Figure 6.14  Schematic representation and position of the CU633906 BAC  

Figure 6.15  NDJ3 exhibits increased expression of disnor238 transcript  

Figure 6.16  Overexpressed disnor238 localizes to the nucleolar periphery in clone NDJ3  

Table 2.1  Bacterial strains  

Table 2.2  DJ transcript primers  

Table 2.3  BAZ2A duplexes  

Table 2.4  DJ primers for mapping and ChIP experiments  

Table 2.5  Primary Antibodies  

Table 2.6  Secondary AAntibodies  

Table 2.7  Constructs generated  

Table 2.8  Cloning vectors  

Table 5.1  Fold increase of DJ sequences present in BAC array clones determined by qPCR  

Table 5.2  Location of the BAC insertions of the clones from a random selection of 20 cells in interphase shown in Fig. 5.13  

Table 5.3  Location of the BAC integration sites as determined by FISH on metaphase shown in Fig. 5.14
Abbreviations and acronyms

~ approximately
°C degrees Celcius
α anti or alpha
β beta
μ micro
A$_{260}$ absorbance at 260nm
A$_{280}$ absorbance at 280nm
Ab antibody
ACF ATP-utilizing chromatin assembly and remodelling factor
AgNORs argyrophilic NORs
AMD Actinomycin D
ARF alternative reading frame
asRNA antisense RNA
BAC bacterial artificial chromosome
BAZ2A bromodomain adjacent to zinc finger domain protein 2A
BLAST Basic Local Alignment Search Tool
bp base pair
BRG1 ATP-dependent helicase SMARCA4) and BRM (highly homologous to BRG1 ATP-ase
BSA bovine serum albumin
CaCl$_2$ calcium chloride
Cdc cell division cycle
Cdk cyclin dependent kinases
cDNA complementary DNA
CER centromeric repeats
CF core factor
ChIP chromatin immunoprecipitation
Chr chromosome
CHRAC chromatin remodelling and assembly complex
ChromHMM chromatin Hidden Markov Model
CRISPR clustered regularly interspaced short palindromic repeats
CSB cockayne syndrome protein B
C-terminus carboxy terminus
CtBP C-terminal binding protein
Da dalton
DAPI 4’-6-diamidino-2-phenylindole
DFC dense fibrillar centre
DHS DNase I Hypersensitive site
DJ distal junction
DMEM Dulbecco's modified eagle's medium
DMD Duchenne Muscular Dystrophy
DMSO dimethyl sulfoxide
DNA deoxyribonucleic acid
DNMT DNA methyltransferases
dNTPs deoxynucleotide triphosphate
ds double stranded
DTT dithiothreitol
ECL enhanced chemiluminescence
EDTA  ethylenediaminetetra-acetic acid disodium salt
EGF  epidermal growth factor
EGFR  EGF receptor
ENCyclopedia of DNA Elements
EM  electron microscope
ERK  extracellular signal regulated kinase
EST  expressed sequence tag
EtBr  ethidium bromide
ETS  external transcribed spacer
FACT  facilitates chromatin transcription
FAIRE  formaldehyde assisted isolation of regulatory elements
FBS  foetal bovine serum
FC  fibrillar centre
FIB  fibrillarin
FISH  fluorescence in situ hybridization
FITC  fluorescein isothiocyanate
g  gram or gravity
GC  granular component
GM12878  B-lymphoblastoid cells
GFC  giant fibrillar centre
GFP  green fluorescent protein
H1 ES  Embryonic stem cells
H3K4me3  H3 trimethylated at lysine residue 4
HAT  histone acetyltransferase
HepG2  Hepatocellular carcinoma cells
HDAC  histone deacetylase
HEK293  human embryonic kidney 293
HEPES  N-2-[hydroxyethyl] piperazine-N-2-ethanesulphonic acid
HIV  human immunodeficiency virus
HMDM2  human double minute 2 homolog
HMEC  Mammary epithelial cells
HMG  high mobility group
HP1  heterochromatin protein 1
HMT  histone methyltransferases
hr/hr  hour/hours
HRP  horseradish peroxidase
HSMM  Skeletal muscle myoblasts
HTLV-1  human T-lymphotropic virus
HUVEC  Umbilical vein endothelial cells
IGF  insulin growth factor
IgG  Immunoglobulin G
IGS  intergenic spacer
IRS  insulin receptor substrate-1
ISWI  imitation SWI
ITS  internal transcribed spacer
JNK2  c-jun N-terminal protein kinases 2
K562  Erythrocytic leukaemia cells
kb  kilo-base
KCI  potassium chloride
kDa  kilodalton
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>litre</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani media/broth</td>
</tr>
<tr>
<td>m</td>
<td>milli</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>MAPK</td>
<td>mitogen-activated protein kinase</td>
</tr>
<tr>
<td>Mb</td>
<td>mega bases</td>
</tr>
<tr>
<td>MDM2</td>
<td>Mouse double minute 2 homolog</td>
</tr>
<tr>
<td>MEM</td>
<td>minimum essential medium</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>magnesium chloride</td>
</tr>
<tr>
<td>min/mins</td>
<td>minute/minutes</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>ml</td>
<td>milliliter</td>
</tr>
<tr>
<td>mol</td>
<td>moles</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mTOR</td>
<td>mammalian TOR</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>n</td>
<td>nano</td>
</tr>
<tr>
<td>NaAc</td>
<td>sodium acetate</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NHEK</td>
<td>Normal epidermal keratinocytes</td>
</tr>
<tr>
<td>NHLF</td>
<td>Normal lung fibroblasts</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>NOR</td>
<td>nucleolar organiser region</td>
</tr>
<tr>
<td>NoRC</td>
<td>nucleolar remodelling complex</td>
</tr>
<tr>
<td>NP40</td>
<td>Nonident-P40</td>
</tr>
<tr>
<td>N-terminus</td>
<td>amino terminus</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>NURF</td>
<td>nucleosome remodelling factor</td>
</tr>
<tr>
<td>O/N</td>
<td>over night</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAF</td>
<td>polymerase associated factor</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCAF</td>
<td>P300/CBP-associated factor</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PFA</td>
<td>paraformaldehyde</td>
</tr>
<tr>
<td>PJ</td>
<td>proximal juncion</td>
</tr>
<tr>
<td>PI3K</td>
<td>phosphatidylinositol-3-kinase</td>
</tr>
<tr>
<td>PIC</td>
<td>Pre-Initiation Complex</td>
</tr>
<tr>
<td>PKI</td>
<td>protein kinase inhibitor</td>
</tr>
<tr>
<td>PNB</td>
<td>pre-nucleolar bodies</td>
</tr>
<tr>
<td>PNC</td>
<td>peri-nucleolar compartment</td>
</tr>
<tr>
<td>Pol I</td>
<td>polymerase I</td>
</tr>
<tr>
<td>Pol II</td>
<td>polymerase II</td>
</tr>
<tr>
<td>Pol III</td>
<td>polymerase III</td>
</tr>
<tr>
<td>pRb</td>
<td>retinoblastoma protein</td>
</tr>
<tr>
<td>PTRF</td>
<td>Pol I transcript release factor</td>
</tr>
<tr>
<td>rDNA</td>
<td>ribosomal DNA</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RNP</td>
<td>riboNucleoProtein particle</td>
</tr>
<tr>
<td>RP</td>
<td>ribosomal protein</td>
</tr>
<tr>
<td>RPE</td>
<td>retinal pigment epithelial</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RRN3</td>
<td>ribosomal RNA polymerase I complementation group 3</td>
</tr>
<tr>
<td>RSK</td>
<td>ribosomal S6 kinase</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>SD</td>
<td>segmental duplication</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>SDS-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Sec/secs</td>
<td>second/seconds</td>
</tr>
<tr>
<td>Seq</td>
<td>sequencing</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td>SL1</td>
<td>selectivity factor 1</td>
</tr>
<tr>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
</tr>
<tr>
<td>snoRNP</td>
<td>small nucleolar ribonucleoprotein</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>SRP</td>
<td>signal recognition particle</td>
</tr>
<tr>
<td>ss</td>
<td>single stranded</td>
</tr>
<tr>
<td>SSC</td>
<td>saline-sodium citrate</td>
</tr>
<tr>
<td>SSU</td>
<td>small sub-unit</td>
</tr>
<tr>
<td>SUMO</td>
<td>small ubiquitin like modifier</td>
</tr>
<tr>
<td>SWI/SNF</td>
<td>SWItch/Sucrose Non Fermentable</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris acetate EDTA</td>
</tr>
<tr>
<td>TAM</td>
<td>TIP5/ARBP/MBD</td>
</tr>
<tr>
<td>TAF</td>
<td>TBP associated factors</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris borate EDTA</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-binding protein</td>
</tr>
<tr>
<td>TE</td>
<td>Tris HCl EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TF</td>
<td>transcription factor</td>
</tr>
<tr>
<td>TFIIH</td>
<td>transcription factor H for Ppol II transcription</td>
</tr>
<tr>
<td>TIF-1A or B</td>
<td>transcription initiation factor for pol I (RRN3) or B (SL1)</td>
</tr>
<tr>
<td>TIP5</td>
<td>TTF-1 interacting protein 5</td>
</tr>
<tr>
<td>TOR</td>
<td>target of rapamycin</td>
</tr>
<tr>
<td>ToRC</td>
<td>Toutatis-containing chromatin remodelling complex</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TTF-1</td>
<td>transcriptional termination factor 1</td>
</tr>
<tr>
<td>Tween</td>
<td>polyoxyethylene-sorbitan monolaurate</td>
</tr>
<tr>
<td>U</td>
<td>units</td>
</tr>
<tr>
<td>UBF</td>
<td>upstream binding factor</td>
</tr>
<tr>
<td>UCE</td>
<td>upstream control element</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine 5-triphosphate</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>v/v</td>
<td>volume to volume</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>W</td>
<td>watt</td>
</tr>
<tr>
<td>WS</td>
<td>Werner Syndrome</td>
</tr>
<tr>
<td>w/v</td>
<td>weight to volume</td>
</tr>
<tr>
<td>XPB and XPD</td>
<td>xeroderma pigmentosum group B and D genes</td>
</tr>
</tbody>
</table>
**ABSTRACT**

The short-arms of the five acrocentric human chromosomes contain sequences that direct the assembly and function of the nucleolus, one of the key functional domains of the nucleus and ribosomal biogenesis, yet they are absent from the human genome assembly. It is well known that not all 10 NORs are active in the cells but the regulatory mechanisms that regulate their activity have not been identified. Previous reports, suggested that a chromatin remodeling factor, BAZ2A, is a nucleolar protein that plays a role in rDNA silencing and could be implicated in whole NOR inactivation. I have tested this hypothesis and found no evidence in support of it. Instead I focused on describing the genomic architecture of the human NORs. This was driven by the hypothesis that sequences outwith the rDNA could regulate NOR activity. Sequences found distally and proximally to ribosomal gene arrays were found to be conserved among the acrocentric chromosomes and to share a complex genomic architecture similar to other euchromatic regions of the genome. However, they have distinct genomic characteristics. Proximal sequences are almost entirely segmentally duplicated. In contrast, the distal sequence is predominantly unique to the acrocentric short arms, and is dominated by a large inverted repeat. This distal element was found to localize to the periphery of the nucleolus, where it appears to anchor the ribosomal gene repeats. This combined with its complex chromatin structure and production of transcripts that behave like IncRNAs, localizing to their site of synthesis, suggests that this region is involved in nucleolar organization. Further work on these novel genetic elements will shed light on the well-known empirical association of altered nucleolar morphology and function with human pathology.
1. Introduction
1.1 The nucleolus and its importance

In higher eukaryotic cells the nucleolus is a prominent subnuclear compartment whose primary role is ribosomal biogenesis (Busch and Smetana, 1970). Ribosome biogenesis is extremely important to actively dividing cells since they have a vast demand for proteins. For this need to be satisfied, every cell generation must produce a sufficient amount of ribosomes (Alberts et al., 2002). The overall protein synthetic capacity of the cell is determined by the steady state number of ribosomes present (Olson, 2004). Ribosome biogenesis is a complex process largely dependent on the coordinated synthesis of ribosomal proteins, ribosomal RNAs and their processing and assembly into mature ribosomes. To understand the magnitude of the importance of the ribosomal gene transcription one needs to know that it accounts for approximately 40-60% of all cellular transcription and 80% of the steady cellular RNA content (Olson, 2004, Laferte et al., 2006).

The size of the nucleolus has been proposed to be proportional to its rate of ribosomal biogenesis (Melese and Xue, 1995). In cells producing very large amounts of proteins it can occupy even up to 25% of the total nuclear volume, suggesting that its prominence is largely determined by its service to the cell (Alberts et al., 2002). At the heart of the nucleolus is a set of tandemly repeated genes encoding pre-ribosomal RNA (rDNA) that in humans are located in the secondary constrictions of the acrocentric chromosomes. Due to their ability to initiate the formation of nucleoli during interphase, these segments of the chromosomes are called nucleolar organiser regions (NORs) (Roussel et al., 1996). Before a more detailed description on the function of the nucleolus it would be interesting to discuss some historical highlights since its identification that will help understand the main issues regarding the NORs and nucleolar formation.

1.2 Discovery and initial study of the nucleolus

The first reference on the nucleolus dates back to 1781 when Fontana described an ovoid body in the nucleus (Review in Mosgoeller, 2004). In the
1800s, Wagner (1835) and Valentin (1836,1839) were the first to independently refer to the nucleolus in their studies until Montgomery in 1898 hand-drew figures of nuclei and nucleoli. But it wasn’t until Heitz (1931) (Review in Pederson, 2010) and McClintock (1934) independently observed that nucleoli arise at specific chromosomal loci that McClintock called the nucleolus-organising bodies or as nowadays known nucleolar organiser regions (NORs). This finding could now be associated with Montgomery’s figures of different shapes and sizes of nucleoli and show that the nucleolus is not just a part of the nuclear anatomy but derives from a cytogenetic entity. More insight on the function of the nucleolus was given by four separate studies showing that the NORs contained the genes encoding the 18S and 28S ribosomal RNAs (rDNAs). Studies on Xenopus revealed that a mutant lacking nucleoli resulted from the absence of rRNA synthesis and therefore lack of ribosome biogenesis (Brown and Gurdon,1964, Birnstiel et al.,1966). Furthermore experiments on Drosophila showed that the size of the NOR and the amount of rRNA present are directly associated (Ritossa and Spiegelman, 1965). Liau and Perry in 1969 reported from experiments on isolated nucleoli that the rRNAs are initially synthesised in the form of large precursor molecules (Liau and Perry, 1969). These are then processed to give rise to the 18S and 28S rRNAs. The most important finding though around the same period were the Miller spreads, captures of ribosomal genes during transcription (Miller and Beatty, 1969).

1.3 Ribosomal genes and Nucleolar Organiser Regions (NORs)

As mentioned earlier, the chromosomal context around which the nucleolus forms is called the nucleolar organizer region (NOR) (McClintock, 1934). NORs contain the ribosomal DNA repeats (rDNA) that are organized into clusters of tandem arrays and are transcribed by the dedicated RNA Polymerase I (RNA Pol I) machinery encoding the major rRNA species. In humans, NORs are located between the two blocks of satellite sequences on the short arms of the five acrocentric chromosomes (13, 14, 15, 21, and 22) (Henderson et al., 1972, Roussel et al., 1996). The location of the NORs on
the acrocentric chromosomes and the adjacent heterochromatic satellite DNA reinforce their isolation from genes transcribed by RNA Pol II and RNA Pol III (Mais et al., 2005, McStay and Grummt, 2008).

**Figure 1.1 Components of the rDNA repeat.** This is a schematic representation of an individual rDNA repeat. The arrow points to the direction of transcription. The location of the two external transcribed spacers (5’ETS and 3’ETS) is indicated with dotted lines as well as the two internal transcribed spacers (ITS1 and ITS2). The ETS are located outside the coding genes (18S, 28S and 5.8S) while the ITS are located inside the coding region. Outside the coding region lay the non-transcribed regions (NTS)/intergenic spacers (IGS) indicated by dotted lines. (Adapted from Romanova et al., 2006)

The entire human rDNA repeat’s length is approximately 43kb. Each cell contains overall approximately 600 repeats (Schmickel, 1973, Stults et al., 2008). The transcribed region is 13kb long and codes for the 47S precursor RNA (pre-rRNA). The genes encoding the 18S, 5.8S and 28S ribosomal subunits are located between two external transcribed spacer sequences (5’ETS and 3’ETS) and are separated by two internal transcribed spacer sequences (ITS1 and ITS2). The transcribed regions are separated by non-transcribed sequences of approximately 30kb termed intergenic spacers (IGS). These intergenic spacers contain regulatory elements including the RNA Pol I specific rDNA gene promoter, transcription termination sites and transposable elements. (Gonzalez and Sylvester, 1995, Haltiner et al., 1986, Sylvester et al., 2004) (Fig.1.1). No enhancer elements and spacer promoters have so far been identified in the human rDNA repeat as have been found in mouse and *Xenopus* (Labhart and Reeder, 1984, Pikaard et al., 1990). The rDNA promoter is bipartite comprising of a core promoter element and an upstream control element (UCE). The core element is an essential and sufficient signal for RNA Pol I transcription. It determines the transcription
start site and binds the basal transcription factors that then recruit Pol I but does not require the presence of UCE for transcription in vitro. The UCE is important in modulating the efficiency of transcription but not its accuracy and is mostly essential in vivo. Its role is dependent on the direction and distance from the core element (Haltiner et al., 1986, Learned et al., 1986).

Evidence coming from biochemical analysis of the rDNA in different cell lines of different species has shown that at any given time ~ 50% of the rDNA repeats are transcriptionally silent. This suggests that mammals have an excess of rDNA over that required (McStay and Grummt, 2008). Robertsonian translocations provide good evidence that this is the case. In these translocations, the short arms of two acrocentric chromosomes are fused together resulting in the loss of 2 out of 10 NORs. Carriers of such translocations are phenotypically normal (Hamerton et al., 1975, Therman et al., 1989). Furthermore, as shown in several studies the rRNA synthesis rate is not directly related to the number of rDNA repeats. One example is the Saccharomyces cerevisiae (yeast) in which a reduction from the 150 rDNA repeats down to 40 contained in the single chromosome XII doesn’t affect cell growth or rRNA synthesis (Oakes et al., 2006). Another example is the Potorous (rat-kangaroo) in which the NOR is located on the X chromosome so females have twice the number of rDNA repeats. Dosage compensation would be expected but on the contrary both NORs are transcriptionally active escaping X inactivation (Merry et al., 1983). However, all these examples still could not account for as much as 50% transcriptionally inactive rDNA. Therefore, a more likely hypothesis is that the NORs are themselves a mosaic of active and inactive rDNA repeats (McStay and Grummt, 2008).

Pulse-field gel electrophoresis experiments of digested genomic DNA with enzymes that do not cut human rDNA, such as EcoRV and Sse83871, resulted in a large 3 Mb rDNA band and other minor bands of 1 and 2Mb (Sakai et al., 1995). This suggests that human NORs are composed of only rDNA and contain approximately 70 copies of rDNA repeats (Worton et al., 1988, Golzalez and Sylvester, 1997). However, these clusters of rDNA repeat range between 50kb to more than 6Mb between and within human
individuals (Stults et al., 2008). The rDNA repeats are oriented and transcribed in a telomere to centromere fashion (Worton et al., 1988, Golzalez and Sylvester, 1997). However, results using the molecular combing technique on single DNA molecules showed that the rDNAs are distributed on the combed fibre as a mosaic of canonical (telomere-to-centromere) and (~1/3) noncanonical repeats (centromere-to-telomere i.e. inverted) forming palindromic structures (Caburet et al., 2005, McStay and Grummt, 2008).

It is very important to highlight that due to the repetitive nature of the NORs and the adjacent satellite DNA, these chromosome arms are still missing from current drafts of the human genome. Hence, the precise chromosomal context of NORs is very poorly understood (McStay and Grummt, 2008).

1.4 Active vs inactive NORs

In most human cell lines the majority of NORs, but not all, are active, forming between 1 and 3 nucleoli (Savino et al., 2001). The rDNA in active NORs is found to be approximately 10 times less condensed than that of the adjacent satellite repeats. Therefore, active NORs appear as weakly DAPI stained chromatin regions on metaphase chromosomes due to their reduced condensation and are called secondary constrictions (Heliot et al., 1997). The primary constriction is the centromere. During this chromatin state, there is less intense DAPI staining due to the fact that DAPI cannot bind efficiently resulting in this apparent gap on the chromosomes. Interestingly, an axis of condensed DNA, possibly AT-rich, has been observed within the secondary constriction giving rise to speculation that the DNA forms loops during this de-condensed state (Saitoh and Laemmli, 1994).

The decondensed sequences consist of rRNA genes that were transcriptionally active in the previous interphase and that will be active again after mitosis. It has been shown that active NORs can be silver stained due to acidic/argyrophilic domains within the components of the RNA Pol I transcription machinery including the Upstream binding factor (UBF) (Fig1.2).
Therefore, they are also called AgNORs (Sumner, 1982, McStay, 2006). These components remain bound to active NORs during mitosis making silver staining the preferred method to identify and count the active NORs in metaphase chromosomes (Bloom and Goodpasture, 1976, Scheer and Rose, 1984, Roussel et al., 1993, Roussel et al., 1996, Jordan et al., 1996).

**Figure 1.2 Active versus inactive NOR.** This is a schematic representation of active and inactive NORs. A) On the left is the active NOR. The dotted red lines point out the decondensed chromatin that is visualised as a secondary constriction. On the right is the inactive NOR with the condensed chromatin. B) During interphase, nucleoli form around individual active NORs. One distinctive characteristic of active NORs is that they can bind to UBF and therefore it can be used as a staining antibody for immunofluorescence. (Adapted from Brian McStay)

Using this technique it has been established that in human lymphocytes the number of active NORs is 8 out of 10 (Heliot et al., 2000) and in the cancer cell line HeLa is 7 out of 10 (Roussel et al., 1993). Generally, in organisms with multiple NORs, the secondary constrictions are small and difficult to visualise and even though this technique is widely used is not completely reliable. During the course of this research another more reliable way of identifying and counting the active NORs has been established that will be discussed in detail later on in this thesis.
Mais and colleagues (Mais et al., 2005) have created exogenous active NORs. These are constructed artificial arrays containing tandem arrays of *Xenopus* rDNA sequences with high affinity for UBF that were introduced into non-NOR bearing human chromosomes. They lack promoter sequences, so they are transcriptionally silent and therefore are called pseudo-NORs. They appear as secondary constrictions in metaphase chromosomes, are associated with UBF during interphase and stain with silver. This suggests that the secondary constriction appearance of active NORs may be the result of all the argyrophilic proteins binding to the DNA preventing its full condensation (Mais et al., 2005). This suggestion can be further supported by the finding that upon siRNA depletion of UBF in pseudo-NORs, both the secondary constriction appearance and silver staining ability are lost (McStay unpublished observations).

Inactive NORs are condensed; hence do not form secondary constrictions on metaphase chromosomes. Also, they do not retain any of the RNA Pol I machinery components and do not stain with silver nitrate. During interphase in human cells silent NORs appear as condensed foci of rDNA repeats. They can be visualised by fluorescence in situ hybridisation (FISH) and are not associated with the nucleolus nor appear to stain with nucleolar proteins such as UBF (McStay and Grummt, 2008).

Within active NORs not all rDNAs are transcriptionally active. While, individual rDNA repeats can be silenced through promoter inactivation, the silencing mechanism of entire NORs is as yet unknown (Roussel et al., 1996, McStay, 2006, Raska et al., 2004). This suggestion will also be discussed in more detail later on in this thesis.

**1.5 The proportion of active repeats in the NORs**

Apart from not all NORs being active, there is a proportion of rDNA repeats within active NORs that are inactive/silent. This proportion cannot be identified by silver-staining but by the DNA cross-linking reagent psoralen. In
transcriptionally active genes the chromatin is de-condensed and accessible to psoralen cross-linking, whereas in silent heterochromatic genes the chromatin is more condensed and inaccessible. After psoralen cross-linking the active and inactive rRNA genes can be distinguished on an agarose gel. The active rRNA genes migrate slower due to their cross-linking (Conconi et al., 1989, Dammann et al., 1993). Based on many different studies it has now been accepted that approximately half of the rDNA repeats are transcriptionally active in most cell types (Conconi et al., 1989, Grummt and Pikaard, 2003, McStay and Grummt, 2008). However, this number does not add up to the proportion of the rDNA repeats that would in general be inactive in only the inactive NORs at a chromosomal level. This finding suggests that human NORs are a mosaic of active and silent repeats.

Additionally, Caburet and his colleagues (Caburet et al., 2005) using molecular combing have demonstrated that approximately 1/3 of human rDNA repeats have a palindromic rearrangement (not telomere-to-centromere as previously described by Worton et al., 1988). These non-canonical repeats are proposed to be pseudogenes i.e. transcriptionally silent since the non-canonical repeats would lack a promoter at the 5’ end. This suggestion is further strengthened by the finding that the level of these palindromic structures in Werner Syndrome’s (WS) patients’ samples increased approximately to 50% (Caburet et al., 2005). The WS’s protein is localised in the nucleolus (Gray et al., 1998) and is homologous to the sgs1 gene in yeast (Johnson et al., 1999). The product of the WRN gene is a RecQ DNA helicase implicated in DNA replication and DNA repair pathways (Shen and Loeb, 2001). In another study, an increase in the methylation state of the rRNA genes was observed in fibroblasts from WS patients. This could be related to the higher level of palindromic structures and inactivation of transcription units due to rearrangements (Machwe et al., 2000).

An explanation as to why these noncanonical rDNA repeats are nonfunctional and therefore silenced, could be to avoid possible base pairing among antisense transcripts and pre-rRNA that would severely affect ribosomal biogenesis (Caburet et al., 2005, McStay and Grummt, 2008). This is very
important since it is known through analysing sequence polymorphisms within individual rDNA repeats that there is strong recombination activity both intrachromosomally and interchromosomally within the rDNA that evidently leads to concerted evolution. Such crossover events also result in the conservation of the sequences distal and proximal to the NORs that will be discussed in chapter 4 (Worton et al., 1988, Stults et al., 2008).

1.6 Ribosomal Gene Transcription

Three nuclear RNA Polymerases have been identified performing different functions within the cell; RNA Pol I, II and III (Roeder and Rutter, 1969). The RNA Pol I is dedicated solely on the transcription of the 5.8S, 18S and 28S rRNAs that give rise to the 40S (small) and 60S (large) pre-ribosomal subunits. RNA Pol II transcribes protein encoding genes as well as the majority of the small nuclear RNAs (snRNAs) and RNA Pol III transcribes the remaining snRNAs and the 5S rRNA that is part of the large ribosomal subunit (60S) (Korostevlev and Noller, 2007, Paule and White, 2000). It has been found that partial inhibition of RNA Pol I activity results in an irregular accumulation of all the ribosomal components. This phenomenon suggests that the regulation of the RNA Pol I activity holds a critical role in ribosome biogenesis (Laferte et al., 2006). RNA Pol I consists of 13 subunits that all have identical or related counterparts in Pol II (has only 12 subunits) and Pol III (has 17 subunits) (Kwapisz et al., 2008). The mammalian (human) RNA Pol I subunits are: RPA195 (Hannan et al., 1998), RPA135 (Hannan et al., 1998), PAF53 (Hanada et al., 1996), RPA43 (Cavanaugh et al., 2002), RPAC40 (Song et al., 1994), PAF49hASEI (Yamamoto et al., 2004), RPB25 (Hirenmath and Ladias, 1998), RPB14.4 (Hirenmath and Ladias, 1998), RPA16 (Yao et al., 1996), RPB17 (Hirenmath and Ladias, 1998), RPA12.2 (Acc. No. BM59854 image clone 5469419), RPB7 (Shpakovski et al., 1995) and RPB7.6 (Shpakovski et al., 1995).
It is important to mention that rDNA transcription is extremely species specific. This means that, for example, the human rDNA cannot be transcribed in vitro or in vivo by the mouse RNA Pol I transcription machinery and vice versa. This is due to the RNA Pol I transcribing only one class of genes, albeit their repetitive nature, hence allowing it to evolve rapidly and without constraints between species under less severe constraints. This does not apply for RNA Pol II and III since the gene classes they transcribe are more diverse, therefore constraining their evolutionary divergence (Miesfeld and Arnheim, 1984).

The visualization of the heavily transcribed ribosomal genes by RNA Pol I has been successfully captured in photographs under an electron microscope (EM) by Miller and Beatty in 1969. The Miller spreads (Fig. 1.3), that resemble a “Christmas tree”, are generated by treating the cells with low ionic strength buffers (contain detergents); this results to the release and dispersion of chromatin that is then centrifuged onto a carbon-coated grid. The trunk of the tree represents the path of the ribosomal array and the branches of different sizes represent the nascent pre-RNAs extending away from the tree. The terminal knobs at the 5' end of the pre-RNAs contain the small sub-unit (SSU) processome. Furthermore, the Miller spreads also reveal the heavy loading of RNA Pol I on the ribosomal genes, which is approximately one RNA Pol I complex for every approximately 100 nucleotides (Miller and Beatty, 1969, French et al., 2003).

1.7 The basal human RNA polymerase I transcription initiation factors

Initiation of transcription from the human ribosomal genes occurs with the formation of the pre-initiation complex (PIC). The Upstream Binding Factor (UBF; Jantzen et al., 1990) and Selectivity Factor 1 (SL1; Learned et al., 1985) bind to the rDNA promoter, to which then the initiation-competent subfraction of RNA Pol I (Pol I β), that is associated with Rrn3 in humans/TIF-1A in mice (Moorefield et al., 2000, Bodem et al., 2000), is recruited and

Figure 1.3 Photo of the Miller spreads. The Miller spreads reveal the rDNA transcription unit that resembles a Christmas tree. In this structure the rRNA genes are located on the trunk of the tree and the pre-rRNA nascent transcripts are extending out like branches. The 5’ end terminal knobs structure that contain the small subunit processome are indicated by arrowheads. Figure from (Raška, 2003)

Highly abundant UBF contains multiple HMG boxes that are known to be able to bend DNA (Janzen et al., 1990). This motif enables a single dimer of UBF to wrap the DNA in a clock wise direction and form a loop of about $360^0$ every approximately 140bp, bringing the core and UCE units of the promoter in close distance (Stefanovsky et al., 2001). Hence, this change in structure may enable interactions between UBF and SL1 that are bound to the two different parts of the promoter elements, helping the pre-initiation complex formation (Bazett-Jones et al., 1994, Copenhaver et al., 1994) (Fig. 1.4). Primarily, UBF is known to recruit RNA Pol I to the promoter to initiate transcription, stabilizes SL1 and acts as an antirepressor by antagonising non-specific DNA-binding proteins (Kuhn and Grummt, 1992). However, it can also act as a repressor of transcription elongation at the chromatin level due to the $360^0$ DNA loop formation it creates by its association with the promoter. This structure formation termed the enhancesome, appears to block the passage of RNA
Pol I complex. This can be reversed by ERK phosphorylation of UBF’s HMG boxes that by remodeling of this structure, allows continued elongation (Stefanovsky et al., 2006). Finally, it is the prime candidate for maintaining the “open” chromatin form of secondary constrictions in active NORs due to the fact that it remains attached to the promoter during mitosis (Mais et al., 2005).

**Figure 1.4 RNA Polymerase I pre-initiation complex (PIC).** This is a schematic representation of the Pol I PIC. The PIC assembly involves the binding of the homodimeric UBF to SL1, which consists of the TBP and five Pol I specific TBP-associated factors (TAFs). Rrn3/TIF-IA polypeptide interacts with SL1 and is essential for the recruitment of initiation competent form of Pol I for transcription to commence. (Adapted from White, 2005)

The mouse homologue of SL1 is TIF-IB (Clos et al., 1986). SL1 interacts with the rDNA promoter in a highly species specific manner; this specificity is in extension observed in ribosomal gene transcription (Heix and Grummt, 1995). It is a multisubunit complex consisting of a TATA-binding protein (TBP) and TBP-associated factors\_1 (TAF\_110, TAF\_63, and TAF\_48). More specifically, TAF\_110 contacts the rDNA promoter, TAF\_63 binds to the DNA via a zinc finger and TAF\_48 interacts with UBF (Comai et al. 1992). Recently another two TAFs have been identified to be part of the SL1 complex, TAF\_112 and TAF\_41. TAF\_112 binds tightly to the RNA Pol I branch as well as on the rDNA promoter (Denissov et al., 2007). TAF\_41 interacts with the RNA Pol I branch as well as with UBF and when downregulated, the rDNA promoter occupancy of SL1 and Pol I decreases (Gorski et al., 2007).
1.8 Regulation of rDNA transcription

The rate of rDNA transcription is regulated by short- and long-term regulation. Short-term regulation involves changing the rate of transcription from active rDNA repeats by regulating the key components of the RNA Pol I machinery that were described in the above section (for reviews see Grummt, 2003, Russell and Zomerdijk, 2005, Moss, 2007). Long-term regulation on the other hand, involves changing the ratio of active to inactive rDNA gene repeats through epigenetic changes on the rDNA chromatin (for review see McStay and Grummt, 2008).

1.8.1 Cell cycle regulation of rDNA transcription

It has been found that rDNA transcription is a process regulated by the cell cycle. Transcription is silenced during mitosis and progressively increases during G1 reaching maximum levels at S and G2 (Klein and Grummt, 1999). Transcriptional silencing and reactivation during mitosis are controlled mostly by posttranslational modifications of the RNA Pol I machinery key factors. When the cell enters mitosis, SL1 is inactivated by phosphorylation of TAF110 at threonine 852 (T852) by cyclin dependant kinase (Cdk) 1/cyclin B. This phosphorylation inhibits SL1 ability to interact with UBF and form the PIC (Heix et al., 1998, Kuhn et al., 1998). UBF is also silenced by phosphorylation during mitosis (Klein and Grummt, 1999). As the cell cycle progresses and cells exit mitosis and enter G1 and S phase, rDNA transcription is gradually restored reaching its maximum level. Transcription is restored by the dephosphorylation of TAF110 at T852 of SL1 by Cdc14B phosphatase, enabling SL1 to interact with UBF (Drygin et al., 2010). Furthermore, reactivation is further established by the hyperphosphorylation of UBF by Cdk4/cyclin D and Cdk2/cyclin E at serine 484 (S484) at G1 and later during S phase by Cdk2/cyclin E and Cdk2/cyclin A at S388. Phosphorylation of S388 in UBF promotes the interaction between UBF and the RNA Pol I subunit PAF53 (Voit et al., 1999, Voit and Grummt, 2001) which is further
enhanced by P300/CBP-associated factor (PCAF)-dependent acetylation of UBF (Meraner et al., 2006).

1.8.2 Growth factor dependent regulation of rDNA transcription

The rate of rDNA transcription is regulated by mitotic stimulations and nutrient availability. The mitogen-activated protein kinase (MAPK) signalling pathway has been found to regulate RNA Pol I transcription by targeting the key components of the nucleolar transcription machinery. For example, in response to the external stimuli EGF (external growth factor), the EGFR (EGF receptor) is activated and initiates the MAPK signalling cascade that includes Raf, MEK1/2 and ERK1/2. ERK1/2 phosphorylates and activates the ribosomal S6 kinase (RSK). They then both phosphorylate Rrn3/TIF-IA at two serine residues (Ser633 and Ser649) respectively, resulting in the upregulation of rDNA transcription and cell proliferation (Zhao et al., 2003, Wang et al., 2001). Furthermore, the activation of the MAPK pathway results in ERK1/2 dependant phosphorylation of UBF at Thr117 and Thr201 up-regulating rDNA transcription (Stefanovsky et al., 2001b).

An example of rDNA transcription rate regulation in response to nutrient availability is the target of rapamycin (TOR) kinase pathway. TOR proteins are members of the phosphatidylinositol 3-kinase (P13K) family and have been found to be associated with nutrient-dependent regulation of cell-growth and proliferation in mammals and yeast. Upon nutrient starvation or rapamycin inhibition of the TOR pathway it has been found that rDNA synthesis is reduced resulting in reduced production of ribosomes. Inhibition of the mammalian TOR (mTOR) results in inactivation of Rrn3/TIF-IA (Claypool et al., 2004, Mayer et al., 2004). The activation of protein phosphatase 2A (PP2A) causes dephosphorylation of Rrn3/TIF-IA at Ser44 which correlates with increased phosphorylation of S199 by an unknown kinase and results in the inactivation of Rrn3/TIF-IA. The inactive Rrn3/TIF-IA cannot interact with SL1/TIF-IB or RNA Pol I anymore so the PIC formation is obstructed and the rDNA transcription is inhibited (Mayer et al., 2004). Also, S6-kinase
dependant phosphorylation at the C-terminal tail of UBF promotes its association with SL1/TIF-IB up-regulating rDNA transcription (Hannan et al., 2003). Finally, the insulin growth factor (IGF1) facilitates the binding of the insulin receptor substrate-1 (IRS-1) and the phosphoinositide 3-kinase (PI3K). This results in direct phosphorylation of UBF and upregulation of rDNA transcription (Drakas et al., 2004).

1.8.3 RNA Pol I transcription responds to genotoxic stress

Cell stress response and cell cycle arrest have been suggested to be involved in the regulation of the nucleolus. Nucleolar segregation is a phenomenon observed due to genotoxic stress such as UV irradiation and inhibition of RNA Pol I transcription by actinomycin D (AMD) treatment described in more detail in section 1.11 and chapter 5 (Flickinger, 1968, Chen and Jiang, 2004).

An example of nucleolar stress response is the activation and stabilization of the tumour suppressor protein p53. Under normal conditions p53 is short lived and its cellular levels are kept low through interactions with the E3 ubiquitin ligase MDM2/HDM2 that directs it for degradation. Upon exogenous stress to the cell by DNA damage, hypoxia, heat shock etc, both the large (L) and small (S) ribosomal proteins relocate from the nucleolus to the nucleoplasm (Boulon et al., 2010, Zhang and Lu, 2009). There they interact with MDM2(mouse)/HDM2(human) preventing p53 degradation. Hence, p53 is stabilized resulting in either cell cycle arrest or apoptosis. Another trigger into p53-dependent cell cycle arrest upon stress is the predominantly nucleolar p14ARF (alternative reading frame) protein’s increased expression. This results in p14ARF to associate with MDM2/HDM2 and physically isolate it to the nucleolus. This results in preventing p53 ubiquitylation, nuclear export and degradation (Weber et al., 1999). On the other hand, work by Llanos and colleagues have shown that p14ARF can directly inhibit the ubiquitin ligase activity of MDM2/HDM2 and stabilize p53 (Llanos et al., 2001). The general consensus is that the accumulation of active p53 down-regulates rDNA transcription indirectly. However one study has shown that this down-
regulation in transcription is a direct effect of p53 accumulation interfering with SL1/TIF-IB and UBF interactions in vitro hence preventing the PIC formation (Zhai and Comai, 2000).

Furthermore, the down-regulation of rDNA transcription in response to stress has been shown to be regulated by Rrn3/TIF-IA. In mice, TIF-IA in response to stress stimuli has been found to be phosphorylated at Thr200 by c-Jun N-terminal kinase-2 (JNK2). This phosphorylation prevents its interaction with SL1/TIF-IB and RNA Pol I and subsequently inhibits rDNA transcription. In mouse embryonic fibroblasts TIF-IA decrease leads to nucleoli disruption, cell cycle arrest, up-regulation of p53 protein and apoptosis (Mayer et al., 2005, Yuan et al., 2005).

1.8.4 Cancer and regulation of rDNA transcription

Increased rDNA transcription is tightly associated with cancer (Ruggero and Pandolfi, 2003). Hence, proto-oncogenes and tumour suppressors have been shown to directly target key factors in the RNA Pol I machinery and regulate rDNA transcription. For example, the proto-oncogene c-Myc has been found to up-regulate rDNA transcription by increasing the UBF levels in the cell, whereas Mad1, a c-myc antagonist, downregulates the UBF expression and consequently rDNA transcription (Poortinga et al., 2004). Down-regulation of rDNA transcription during cell differentiation is associated with reduction of c-Myc levels. This is achieved by the down-regulation of the expression levels of its key targets, UBF, Rrn3/TIF-IA and the RNA Pol I subunits (Poortinga et al., 2004). The hyper-activation on the other hand of the rDNA transcription in c-Myc caused cancers is due to up-regulation of the above key targets. c-Myc has also been shown to bind to the rDNA repeat in vivo suggesting that it can directly modulate rDNA transcription (Arabi et al., 2005, Grandori et al., 2005). It seems more likely however that c-Myc indirectly directs rDNA transcription by modulating the levels of the transcription machinery as described above.
Other tumour suppressors, apart from p53, as the word implies suppress the growth of tumours and at the same time have the ability to down-regulate RNA Pol I transcription. For example, the tumour suppressor retinoblastoma protein (pRb) and the related pocket proteins p107 and p130, can suppress cell growth and cell proliferation. pRb accumulates in the nucleolus upon cell confluence, cell cycle arrest or during differentiation and down-regulates rDNA transcription (Cavanaugh et al., 1995, Hannan et al., 2000b). One study suggests the reason for the down-regulation is the interaction between pRb and UBF causing dissociation of UBF from the rDNA (Voit et al., 1997). Another study suggests that the pRb/UBF interaction inhibits UBF’s association with SL1/TIF-IB thus obstructing the PIC formation (Hannan et al., 2000a). Finally, the pRb related pocket protein p130, but not the p107, has a similar to pRb function by binding to UBF and represses rDNA transcription (Hannan et al., 2000a, Ciarmatori et al., 2001).

It has been observed that UBF up-regulation correlates with increased levels of rDNA transcription and hypertrophic growth in neonatal and adult cardiomyocytes (Brandenburger et al., 2003). The opposite effect is observed in cells during differentiation such as of the L6 myoblasts changing to myotubes. During this phase the rDNA transcription rate is down-regulated and the UBF absolute levels are decreased (Larson et al., 1993). Other examples include the F9 embryonal carcinoma cells to primitive endoderm cells (Datta et al., 1997, Alzuherri and White, 1999), 3T3-L1 preadipocyte cells to adipocyte cells (Li et al., 2006b) and murine granulocyte differentiation (Poortinga et al., 2004).

1.8.5 Tissue-specific regulation of rDNA transcription

It has been observed that in multicellular organisms not all cells share the same need for rRNAs, hence the rate of rDNA transcription differs in different cell lines. For example, in developing mouse oocytes the rate of rDNA transcription doubles but the number of rRNA genes remains stable. A major role for this phenomenon to occur is played by basonuclin. Basonuclin is a
transcription regulator that affects rDNA transcription and is mainly expressed in keratinocytes and gametocytes. It accumulates in the nucleolus and remains bound to the rDNA throughout the cell cycle (Tseng et al., 1999, Tian et al., 2001, Zhang et al., 2007).

The Runt-related transcription factor, Runx2, is also a cell-type specific rDNA transcription regulator. It is essential in osteoblast differentiation and skeletal morphogenesis. Like basonuclin, it accumulates in the nucleolus and remains bound to the rDNA throughout the cell cycle. It regulates rRNA synthesis by repressing rDNA transcription (Young et al., 2007).

1.9 Elongation and termination of transcription

The RNA Pol I transcription elongation process is unfortunately still poorly understood. Some studies suggest that proteins with chromatin-remodelling functions such as FACT (facilitates chromatin transcription) and histone chaperones such as nucleolin and nucleophosmin may promote Pol I transcription elongation (Rickards et al., 2007, Murano et al., 2008, Birch et al., 2009). This is consistent with findings showing that subunits of the histone chaperone FACT are associated with mammalian Pol I as well as the transcribed part of the rDNA and when depleted by RNA interference the production of pre-rRNA levels are significantly reduced (Birch et al., 2009). Another protein with chromatin remodelling function is CSB (Cockayne syndrome protein B) (Beerens et al., 2005). CSB is found to localise in the nucleolus at sites of active rDNA transcription and is part of a protein complex containing Pol I, TFIH and basal Pol I transcription initiation factors (Bradsher et al., 2002). CSB in association with G9a, a histone methyltransferase, have been suggested to promote Pol I transcription elongation by imprinting a specific chromatin mark (H3K9me3) recognised by chromatin-modifying enzymes or elongation factors hence enabling transcription through chromatin (McStay and Grummt, 2008). Mutations in the CSB protein result in the Cockayne Syndrome that among skeletal and neurological abnormalities includes markedly reduced rDNA transcription levels (Bradsher et al., 2002).
Interestingly, mutations in subunits (XPB and XPD) of TFIIH, mentioned earlier, which is mainly a transcription factor for RNA Pol II acting as an elongation factor for RNA Pol I, also lead to Cockayne Syndrome (Assfalg et al., 2012). UBF has been associated with Pol I transcription elongation by binding across the whole rDNA repeat including the transcribed region (O’Sullivan et al., 2002) and enhancing elongation by recruiting Pol I to the promoter as mentioned earlier (Kuhn and Grummt, 1992). Furthermore, as mentioned before, UBF can act as a repressor of elongation at the chromatin level. Enhancement of elongation can be observed by ERK (MAPK-kinase) mediated phosphorylation of UBF (Stefanovsky et al., 2006). Finally, UBF has been found to stimulate promoter escape and RNA Pol I clearance leading to the RNA Pol I becoming a stable elongation complex forming full length transcripts from a previously static complex producing abortive transcripts (Panov et al., 2006).

Termination of transcription is a multistep process conserved during evolution among humans, mouse, rat, yeast and *Xenopus*. The steps involved in termination are halting of RNA Pol I, release of both pre-rRNA and RNA Pol I and the 3’-end processing of the primary transcript. As mentioned earlier the IGS in humans contains multiple (11) termination sites downstream from the 3’ end of the pre-rRNA coding region [Bartsch et al., 1987, Pfleiderer et al., 1990]. In mice, the termination sites are recognised by the transcription terminator factor 1 (TTF-1) (Grummt et al., 1985). When TTF-1 binds to the termination sites, it bends the DNA which forces the RNA Pol I to halt, resulting in its conformational change through its interaction with PTRF (Pol I transcript release factor). Transcription termination is then achieved by the simultaneous release of RNA Pol I and rRNA transcripts (Jansa et al., 1998, Jansa and Grummt, 1999). Another terminator element in mice is T₀. This is a single termination site located immediately upstream of the rDNA promoter. It has been shown that T₀ can stimulate and stabilise transcription in vivo implying a dual role for TTF-1 (Grummt et al., 1986, McStay and Reeder, 1986, Henderson and Solner-Webb, 1986).
1.10 Post-transcriptional steps in ribosome biogenesis

Ribosome biogenesis occurs in the nucleolus but it involves both nucleoplasmic and cytoplasmic activities. Therefore, it is a complex but very well coordinated series of events whose initial step, rDNA transcription, is responsible for up to 60% of the total cellular transcription activity (Warner, 1999). Synthesis is followed by the processing and modification of the pre-rRNA and its assembly with ribosomal proteins, including transient interactions with non-ribosomal proteins (Tschochner and Hurt, 2003).

The 80S eukaryotic ribosome comprises of a small 40S and a large 60S subunit. The small 40S subunit contains the 18S rRNA as well as some ribosomal proteins. The large 60S subunit contains the 5.8S and 28S rRNAs, the imported 5S rRNA synthesised in the nucleoplasm by RNA Pol III and some imported ribosomal proteins synthesised in the cytoplasm (Tschochner and Hurt, 2003). The biogenesis of these ribosomal subunits starts with the synthesis of the 47S pre-rRNA transcripts by the RNA Pol I machinery as described earlier. The generation of the mature 18S, 5.8S and 28S rRNAs involves the pre-rRNA’s processing and modification with the help of a number of small nucleolar ribonucleoprotein (snoRNPs) complexes.

These snoRNP complexes consist of non-coding RNA, the small nucleolar RNA (snoRNA) and are categorised in two classes; the box C/D snoRNAs that direct 2’-O-ribose methylation and the H/ACA snoRNAs directing pseudouridylation (Ganot and Bortolin, 1997, Kiss-Laszlo et al., 1996, Ni et al., 1997, Tycowski et al., 1996). The box C/D snoRNAs are associated with the core proteins NOP58, NOP56, 15.5K and the methyltransferase fibrillarin (Watkins et al., 2000) and the H/ACA snoRNAs with Nop10, Gar1, Nhpr and the pseudouridylase Nap57 (dyskerin) (Meier, 2005). Pseudouridylation and 2’-O-ribose methylation are two highly abundant modifications. Approximately 200 modified sites have been found in vertebrate ribosomes, further strengthening the notion that the pre-rRNA processing is a very complex process (Bleichert and Baserga, 2011).
The mature ribosomes do not contain any snoRNAs. RNA helicases are assumed to help the snoRNAs and their associated proteins to be released from the pre-rRNA transcript after pre-rRNA folding, cleavage and processing of the 18S, 5.8S and 28S rRNAs. There has not been any evidence yet of the RNA helicases directly unwinding the snoRNA-pre-rRNA duplex. However, there have been results showing that upon depletion of RNA helicases some snoRNAs remain associated with pre-ribosomes (Bohsack et al., 2008, Liang and Fournier, 2006, Srivastava et al., 2010). Other models have also been suggested for the release of snoRNAs. One suggestion is the conformational changes the pre-rRNA undergoes during its processing, could be dissociating it from snoRNAs. Another suggestion is that this dissociation could be the result of the ribosomal proteins and/or non-helicase processing factors binding to the pre-rRNA. The most probable mechanism though seems to be a combination of RNA helicases and some indirect mechanisms such as described above (Bleichert and Baserga, 2011). Finally, the mature 40S and 60S ribosome subunits are exported to the cytoplasm with the help of adaptor proteins, where they form the functional 80S ribosome by binding to mRNA (Olson, 2004).

1.11 General structural features of the nucleolus

Around the 1980s and 1990s with the use of Electron Microscopy (EM) it was revealed that the nucleolus is the best example of nuclear functional compartmentalisation and the most prominent sub-nuclear compartment. Nucleoli vary in size and shape in different cell types but also within a single cell. The nucleolus is not surrounded by a membrane but by a shell of heterochromatin, presumably from the acrocentric chromosomes, that isolates it from the rest of the nucleoplasm (Nemeth and Langst, 2011, Sullivan et al., 2001).

EM has revealed that in most higher eukaryotes the nucleolus comprises of three compartments: a) the fibrillar centre (FC), b) the dense fibrillar centre (DFC) and c) the granular component (GC). The FC is lightly stained and its
size varies between 0.1-1.0µm in diameter and is surrounded by the DFC which is more intensely stained. The FC and DFC are embedded in the GC that appears to be filled with granules sized between 15-20nm that are probably representing pre-ribosomes through various stages of their maturation. The FCs number seems to double from G1 to G2 and even though their locations are different within the nucleoli, they seem to stay connected by a network of DFC (Junera et al., 1995, Hernandez-Verdun, 2006, Raska et al., 2006) (Fig 1.5).

There has been considerable debate on how this structural partition of the nucleolus corresponds to functionality due to their considerable variation between different species, cell types even between individual cells. Different experimental techniques showed different nucleolar compartments as the location of rDNA transcription mainly due to the DFC’s highly compact structure (Stanek et al., 2001). For example, using EM the Miller spreads cannot be identified since the Miller spreads do not exist in vivo (Raska, 2003). In-situ hybridization and immunocytochemical experimental approaches may fail as well due to lack of knowledge whether they reflect the location of active genes since high-resolution images cannot be established from tomographic data (Ploton et al., 2004). But different kinds of cell treatments may provide more conclusive results due to a clearer signal. For example, non-isotopic transcription mapping after the incorporation of modified nucleotides, showed a low signal in FC suggesting the absence of highly active genes. In contrast, a high signal was present in DFC. Therefore, the most favoured hypothesis is that the transcribing genes are clustered and located in the DFC or at the border line between the FC and DFC interphase (Gonzalez-Melendi et al., 2000, Koberna et al., 2002, Casafont et al., 2006).
Figure 1.5 Nucleolar compartments. Left panel is a differential interference contrast (DIC) image of a HeLa cell with the white arrows indicating the prominent nucleoli. (Figure from Boisvert et al., 2007). The panel on the right is an enlarged image of a HeLa nucleolus visualized by transmission EM. The three compartments comprising the nucleolus are indicated clearly visible. The FC is indicated with an asterisk, it surrounded by the DFC and embedded in the GC. (Figure from Sirri et al., 2008).

It has been shown that the size and number of FCs depends on the metabolic and transcriptional activities of the cells. Notably in more active neuronal cells a giant FC among small FCs has been identified called GFC. It has been shown that the FC is enriched in rDNA, RNA Pol I subunits and the UBF (Goessens, 1984, Scheer and Rose, 1984, Jordan, 1991, Casafont et al., 2007). As mentioned earlier, the nascent transcripts seem to appear initially at the junction between the FC and DFC and accumulate at the DFC. This was further established from experiments on GFCs where no transcripts could be detected (Casafont et al., 2007, Cmarko et al., 2000, Hozak et al., 1994, Shaw and Jordan, 1995). The nascent rRNA transcript starts being processed at the site of transcription at the DFC and continues its migration of the RNA to the GC. The nucleolar proteins that facilitate the processing of the nascent rRNA transcripts such as fibrillarin, nucleolin and the snoRNAs are also found within the DFC (Dragon et al., 2002, Ginisty et al., 1998, Ochs et al., 1985). The GC is the location of the mature rRNA and the ribosomal protein assembly before the subunits are exported to the cytoplasm through the nuclear membrane pores. The proteins that are associated with these later
stages of processing are localised in the GC such as nucleophosmin/B23, Nop52 and nucleostemin (Tsai and McKay, 2002, Savino et al., 1999, Spector et al., 1984).

In higher eukaryotes the nucleolus is assembled in early G1 and disassembled at the end of G2 when transcription of rDNA is arrested, the nuclear envelope collapses and the chromatin condenses into chromosomes. Repression of the RNA Pol I rDNA transcription during mitosis is maintained by CDK1-cyclin B kinase, that is commonly known to be involved in early mitotic events. It has been shown that inhibition of the CDK1-cyclin B kinase pathway in mitotic cells induces resumption of rDNA transcription. This is not sufficient to restore proper nucleolar assembly that includes proper pre-rRNA processing and relocalisation of the processing machinery to the rDNA transcription sites (Sirri et al., 2002). During the disassembling time the nucleolar proteins involved in pre-RNA processing components of the DFC and GC, leave the nucleolus during prophase and are redistributed across the surface of all the chromosomes ensuring their segregation to daughter cells (Gautier et al., 1992). Some factors also remain associated with the rDNA repeats during mitosis, such as UBF, Rm3/TTF-I, SL1/TIF-1B and Treacle (Valdez et al., 2004, Jordan et al., 1996, Roussel et al., 1996, Sirri et al., 1999, Gonzales et al., 2005). It has been also suggested that RNA Pol I remains associated with the rDNA during mitosis. For example from GFP fused RNA Pol I subunits experiments it was shown that RPA43-GFP localises to NORs during mitosis whereas GFP-RPA39/40 associates with the NORS only during prophase, late anaphase and telophase (Chen et al., 2005). This is consistent with the differences in function of these two subunits. RPA43 is a regulatory subunit suggested to play a role in bringing the initiation complex and other RNA Pol I subunits together (Cavanaugh et al., 2002, Yuan et al., 2002) whereas RPA39/40 is suggested to be assembled into the RNA Pol I complex at a later stage (Dundr et al., 2002). However, we have seen no evidence of Pol I on mitotic cells with antibodies against RPA195, PAF53, RPA43 and PAF49 (McStay unpublished). When the nucleolus reforms the nucleolar machineries associated with transcription of the rDNA and processing of the rRNA are targeted to the sites of rRNA
synthesis. These machineries are inherited from the previous interphase (Hernandez-Verdun et al., 2002).

During re-assembly of the nucleolus in early telophase, the rRNA processing proteins maintained in the chromosomal periphery are concentrated in foci termed pre-nucleolar bodies (PNBs). PNBs are distinguished from NORs by the fact that they are smaller in size, do not contain any rDNA and contain U3 snRNPs (snoRNAs). The recruitment of PNBs along with partially processed pre-rRNAs to sites where rRNA transcription has initiated assists the transfer of the essential nucleolar components to the NORs to conclude nucleogenesis (Ochs et al., 1985, Jimenez-Garcia et al., 1994, Savino et al., 1999, 2001, Dundr et al., 2000). rDNA transcription is resumed simultaneously for all active NORs by their association with the newly reactivated RNA Pol I machinery by the CdK1 activity in telophase. This phenomenon results in the formation of a small nucleolus around each individual active NOR. Presumably, each large nucleolus contains multiple NORs but has never been formally proven (Savino et al., 2001).

Nucleolar segregation can also be induced with the treatment of the cells with small amounts (0.01-0.04μg/ml) of AMD. AMD is a wellknown rDNA transcriptional inhibitor of RNA Pol I that is followed by the sorting and rearranging of the nuclear proteins and RNAs into nuclear subdomains. The proposed mechanism of inhibition is by AMD binding to DNA at the transcription initiation complex, immobilizing it and preventing the elongation of transcriptjion (Sobell, 1985). During RNA Pol I arrest there are sequential changes to the nucleolar organisation. The three regions comprising the nucleolus (FC, DFC, GC) segregate into three distinct but in close proximity domains that move to the periphery of the nucleolus. Even though they retain most of their original protein and RNA components, they no longer intermingle. These domains are termed the “central body” consisting of the GC and are associated with the newly formed nucleolar “caps” consisting of the FC/DFC. Simultaneously, the GC proteins are released into the nucleoplasm. Most nucleoplasmic proteins retain their localization, but a significant amount, many of which are rRNA binding, also enter the nucleolar
caps along with the rDNA. Furthermore, a large heterochromatic domain is formed that surrounds the segregated nucleolus (Hadjiolova et al., 1995, Shav-Tal et al., 2005, Hernandez-Verdun, 2006). It has been shown that RNA Pol I transcription inhibition does not prevent the exchange of some nucleolar components such as nucleolin but does decrease the exchange rate of UBF (Chen and Huang, 2001).

1.12 The nucleolar proteome

In recent years combining the ability to purify nucleoli and the use of mass spectrometry to identify and analyse proteins in large scale, there have been great steps on the characterisation of the human nucleolar proteome. Interestingly, in comparison to the yeast proteome there is 90% homology suggesting that the nucleolar proteome is highly conserved through evolution within the eukaryotic kingdom (Andersen et al., 2005). To date the human nucleolus has been found to be associated with approximately 5,000 proteins (Ahmad et al., 2009). These proteins have been found to co-purify with nucleoli isolated from human cells. Interestingly, even though a high proportion of them are associated with ribosomal biogenesis, many of them are not present in high abundance in nucleoli and/or have functions in other locations in the cell. Examples include many pre-mRNA processing factors and proteins that are involved in cell-cycle control as well as DNA replication and repair (Boisvert et al., 2010, Andersen et al., 2002, Scherl et al., 2002, Andersen et al., 2005, Boisvert et al., 2007). Also, upon AMD treatment and inhibition of transcription in cells, it was found that the level of factors associated with ribosomal biogenesis was depleted from nucleoli but some other proteins’ levels associated with the nucleolus were increased even up to ten times. By redistributing proteins between the nucleolus and the nucleoplasm during this state shows that the nucleolus is not only a ribosome synthesis machine that breaks down upon inhibition of transcription (Andersen et al., 2005).
Some proteins found to be located in the nucleolus are also found to be involved in the regulation of every step of the mRNA metabolism such as their synthesis, splicing, editing, nuclear export, and also their translation and degradation (Scherl et al., 2002, Andersen et al., 2002). Evidence now shows that the nucleolus is involved in the processing and maturation of other classes of cellular RNA apart from rRNA. For example it has been found that the nucleolus is related to RNA modifications and protein assembly for multiple RNP complexes that are transcribed by RNA Pol III, such as small nuclear (sn)RNPs, 5S rRNA, some tRNAs and RNAses P RNA, the signal recognition particle (SRP) RNA and microRNAs (miRNAs). This also suggests that even though transcription by Pol III occurs in the nucleus, the processing and maturation of these transcripts occurs in a common location in the nucleolus (Boisvert et al., 2007).

It has also been found that viral infections can alter the nucleolar morphology and the nucleolar proteome. Viral encoded protein localisation to the nucleolus is achieved either by nucleolar targeting signal sequences in the viral proteins or by interactions between the viral proteins and endogenous nucleolar proteins (Boisvert et al., 2007, Olson, 2010, Lam et al., 2010). Three examples for the latter include the endogenous B23 nucleolar protein. This protein has been found to interact with the adeno-associated viral protein Rep (Bevington et al., 2007), the Rev protein of the human immunodeficiency virus (HIV) (Fankhauser et al., 1991) and the Rex protein of human T-lymphotropic virus (HTLV-1) (Adachi et al., 1993). The importance of this association of localisation is the discovery that HIV replication is inhibited by the expression of a nucleolar localising Rev element termed U16RBE (Michienzi et al., 2006). Also, adenovirus infections induce a specific removal of UBF from the nucleolus, independently to the other components of the rDNA transcription complex, and together with B23 are turned into viral replication centres (Lawrence et al., 2006). Various proteins have been found to associate with the nucleolus at different stages of the cell cycle which suggests that the nucleolus also plays a role in cell cycle regulation (Di Bacco et al., 2006, Boisvert et al., 2007).
Evidently, the nucleolus apart from its primary function which is ribosomal biogenesis, it is also involved in other functions such as cell cycle regulation, senescence, stress responses and coordination of the biogenesis of other classes of functional RNPs (Andersen et al., 2005, Boisvert et al., 2007, Olson, 2010). In conclusion, the nucleolus is a dynamic and multifunctional structure in the cell. However, further study on the structure and function of the nucleolus under a range of growth conditions and cell-cycle stages is needed in order to evaluate fully its biological roles. The evidence showing a large number of nucleolar proteins identified in proteomic analyses that are encoded by still uncharacterized open reading frames suggests that it is highly likely that further functions of the nucleolus might be uncovered in the future.

1.13 Ribosomal gene chromatin

Chromatin is formed by the packaging of eukaryotic DNA with proteins. The fundamental repeating unit is the nucleosome. The nucleosome comprises DNA sequence of approximately 145-147bp wrapped around a histone octomer consisting of the histones H2A, H2B, H3 and H4 (Luger et al., 1997). The result of this binding is two complete turns of the DNA sequence around the octomer stabilising the structure. A more stabilized, higher chromatin structure such as the 30 nm fibres, are formed by the binding of the linker DNA to the linker histone H1. The more chromatin condensation steps these 30 nm fibres undergo, the more compact the structure becomes resulting in a final and now visible chromatin structure that is the metaphase chromosomes (Luger et al., 2012). However, there seems to be some contention recently on whether the 30 nm fibre exists in vivo or if it is just an artefact (Fussner et al., 2012). Chromatin is broadly divided into two states in cells, silent heterochromatin and active euchromatin. These can be distinguished by differentiations in their DNA methylation status and histone modifications.

Euchromatin’s main features are: having a less compacted structure, being rich in unique sequences and generally gene-rich. Heterochromatin’s main
features are: its scarcity of genes, its tightly compacted structure throughout cell cycle and its high content of repetitive sequences (Zhou and Berger, 2004, Cook and Karpen, 1994). Heterochromatin is also important for the nucleolar structure. As mentioned earlier in this chapter, the nucleolus is surrounded by a heterochromatic shell that consists of satellite DNA surrounding the NORs and silent rDNA clusters on active or inactive NORs (Sullivan et al., 2001). In addition, it has been found that heterochromatin from non-acrocentric (non-NOR-bearing) chromosomes also associates with nucleoli. Chromosomes 1 and 9 contain a large block of pericentromeric chromatin and some satellites are found there (Manuelidis and Borden, 1988).

A very interesting finding has also been the requirement of the inactive X chromosome’s association with nucleoli in order to facilitate heterochromatin formation post-replication (Zhang et al., 2007). The ribosomal genes can be found in both chromatin states, which are nucleosomes or actively transcribed rDNA (McStay and Grummt, 2008).

Even though rDNA repeats are the most transcriptionally active genes in eukaryotic cells, nucleoli are surrounded by heterochromatin. This statement is further strengthened from findings showing that disruption of this heterochromatin correlates with enhanced instability of the rDNA array, nucleolar disintegration and cellular senescence (Peng and Karpen, 2007). Key players in rDNA stability have been shown to be epigenetic factors affecting the rDNA chromatin structure, such as DNA methylation and histone modification enzymes (McStay and Grummt, 2008). These are described in detail in the following section. It is worth mentioning here that even though it is not known how the “open” chromatin form is established or the precise nature of its conformation, the absence of UBF from silent NORs but its presence in pseudoNORs and active NORs suggests that it is necessary for this establishment (Mais et al., 2005).
1.14 Silencing of ribosomal genes

To date most work on silencing has been performed in plants and mouse cell lines. As mentioned earlier, only indirect methods (such as DNA methylation is linked with silencing) can be used to distinguish between individual active and silent rDNA repeats. These techniques help determine the DNA methylation status of the rDNA repeats as well as the specific histone modifications associated with their chromatin. The most widely used method for distinguishing active from silent rDNA genes is chromatin immunoprecipitation (ChIP) with the help of methylation-sensitive restriction enzymes (Lawrence et al., 2004).

The general consensus modification signatures from several studies correlated with transcriptional activity are DNA hypomethylation, hyperacetylation of histones H3 and H4 and di- and trimethylation of histone H3K4 and methylation of histone H3K9. Modification signatures correlated with transcriptional silence on the other hand are DNA hypermethylation, hypoacetylation of histones H3 an H4 and di- and trimethylation of histones H3K27 and H3K9 (Earley et al., 2006, Ohm et al., 2007, Meissner et al., 2008, Law and Jacobsen, 2010). In general, the way histone methylation is proposed to cause structural changes to chromatin that affects transcription is by putting negative- or positive-acting marks that then recruit effector proteins. One example of an effector protein is heterochromatin protein 1 (HP1) that binds to methylated H3K9 via its chromodomain. HP1 then recruits the histone methyltransferase SU(VAR)39H1 through protein-protein interactions. SU(VAR)39H1 methylates H3K9 in adjacent nucleosomes hence inducing heterochromatin spreading (Bannister et al., 2001, Stewart et al., 2005, Munari et al., 2012). Nevertheless, the above statement is quite generalised since the division between active and silent histone modifications is not strict. The H3K9me3 modifications for example, have been found to be conserved at the transcriptional regions of a mammalian RNA Pol II genes (McStay and Grummt, 2008, Pikaard and Lawrence, 2002, Santoro and Grummt, 2005).
DNA methylation and transcriptional inhibition, is proposed to be performed in a similar way as above. The recruited repressors by DNA methylation bind to specific sites containing methylated CpG dinucleotides. DNA methyltransferases (DNMT1, DNMT3a and DNMT3b) and DNA binding proteins that recognise methylated cytosine sites interact then with histone deacetylase (HDAC) corepressors and histone methyltransferases (HMTs) (Klose and Bird, 2006, Kondo, 2009, Denis et al., 2011, Handy et al., 2011). Transcription is inhibited by methylation from the recruitment of histone deacetylases through methyl-CpG binding proteins or members of the methyl-CpG binding domain (MBD) promoter family (Brown and Szyf, 2007, Ghoshal et al., 2004). Additionally, cross-linking experiments have shown that DNA methylated sites are primarily present in promoters and enhancers of silent genes (Stancheva et al., 1997). Furthermore, methylation of one CpG dinucleotide at position -133 is enough to disrupt binding of UBF preventing the PIC formation in mice (Fig. 1.6) (Santoro and Grummt, 2001). It is important to mention that in humans, methylation of the rDNA is a more complex process since there are 25 CpGs in the promoter. The most usual pattern seen in human rDNA promoters’ methylation is a mosaic of a few to most CpGs methylated (Ghoshal et al., 2004).

Finally as mentioned before, approximately 1/3 of the human rDNA repeats show a non-canonical rearrangement that suggests the possibility of these repeats constituting a major part of methylated rRNA genes (Caburet et al., 2005, Brock and Bird, 1997). Human cancer cell lines compared to normal cells have been shown to undergo hypomethylation at the rDNA promoter (Ghoshal et al., 2004). Recently it has been shown that in contrast to the current belief that silent rDNA is associated with hypermethylation, hypermethylation is rarely observed in primary cell lines (Sinkkonen et al., 2010). The only exception to this are established transformed cell lines as this and additional studies have shown. Examples of transformed cell lines and researches include HeLa, human embryonic kidney (HEK) 293 (Brown and Szyf, 2007, Sinkkonen et al., 2010), mouse NIH 3T3 (Nemeth et al., 2008, Strohner et al., 2001), aged cells (Oakes et al., 2003) and non-dividing neuronal cells (McGowan et al., 2008). Therefore, rDNA methylation might
actually be required for maintaining a long term silent status instead of being required for its establishment in the first instance or this is just a cell line artefact.

A silencing mechanism of ribosomal genes has been proposed for mice. According to this mechanism, TTF-1 can cause nucleosome repositioning at the promoter by binding to $T_0$, resulting in the recruitment of the ATP-dependent nucleolar remodelling complex (NoRC) (Fig. 1.6). NoRC has been proposed to be composed of two subunits, TIP5 (TTF-1 interacting protein 5) in mice /BAZ2A (Bromodomain adjacent to zinc finger domain protein 2A) in humans (Jones et al., 2000) and the ATP-ase SNF2h in mammals/ ISWI in Drosophila (Strohner et al., 2001). Drosophila ISWI can interact with other proteins giving three different chromatin-remodelling complexes: NURF (nucleosome remodelling factor) (Tsukiyama et al., 1995), CHRAC (chromatin remodelling and assembly complex) (Varga-Weisz et al., 1997) and ACF (ATP-utilising chromatin remodelling and assembly Factor) (Ito et al., 1997). Homologues of these complexes have been found in yeast (Tsukiyama et al., 1999), Xenopus (Guschin et al., 2000) and humans (Poot et al., 2000).

![Figure 1.6 The Nucleolar remodelling complex (NoRC).](image)

This is a schematic representation of the mechanism by which NoRC is proposed to silence rDNA repeats. NoRC assembly is initiated on the rDNA promoter by the binding of TIP5/Snf2h to TTF-1. Repression is promoted with the recruitment of HDAC1 and DNMT, which results in silencing from histone H4 deacetylation, histone H3K9 dimethylation, DNA methylation and prevention of UBF binding to the promoter. The nucleosomes have histone acetyl groups and DNA methylated at CpG position 133. (Figure from Matthews and Olson, 2006)
NoRC is then understood to recruit the DNA methyltransferases DNMT1 and DNMT3b as well as the histone deacetylases HDAC1 and HDAC2 resulting in de novo DNA methylation and deacetylation at histone H4 (Santoro et al., 2002, Zhou et al., 2002). Since NoRC is associated with only the late replicating rDNAs, the repressed chromatin state has the potential to be maintained throughout cell division. Overexpression of TIP5 in mice was found to reduce the number and size of nucleoli as well as changing some rRNA genes’ replication times from early to late (Li et al., 2005).

However, even though NoRC might have the potential to silence individual rDNA repeats, it has not been shown that has the potential to change the activity status of whole NORs.

1.15 Thesis Aim

The main aim of my thesis was to elucidate the mechanisms that control the activity status of entire NORs. The first aim was to investigate BAZ2A, a component of NoRC, and especially whether it has the ability to influence rDNA transcription silencing. The next step, was to investigate the possibility that novel regulatory elements could influence an entire NORs activity, such as sequences outside of the NOR. These sequences would be the mapped and their position in reference to the nucleolus would be identified. Furthermore, their chromatin status would be investigated and would be assessed on whether they are functional and if they have the ability to influence the nucleolar morphology and/or regulate the activity status of the rDNA array.
2. Materials and methods
2.1 DNA manipulation

2.1.1 Plasmid DNA Purification from small cultures

Single colonies were picked from agar plates and grown in 10mls LB medium with the addition of the appropriate antibiotic selection and placed in a shaking incubator at 37°C O/N. The cells were centrifuged at 4000xg for 15min, 4°C. DNA was extracted and purified from the cells using the NucleoSpin® Plasmid kit (Machery-Nagel Cat No. 740588.250) as per manufacturer’s instructions.

2.1.2 Plasmid DNA Purification from large cultures

From an agar plate a single colony was picked and grown in 10mls LB medium with the addition of the appropriate antibiotic selection and placed in a shaking incubator at 37°C O/N. This single culture was transferred to 400ml LB medium with the addition of the appropriate antibiotic and grown in a shaking incubator at 37°C O/N. The cells were then centrifuged at 4000xg for 15min, 4°C. DNA was extracted and purified from the cells using the NucleoBound® Xtra Maxi (Machery-Nagel Cat No. 740414.50) as per manufacturer’s instructions.

2.1.3 Glycerol Stocks

An 800μl aliquot from an O/N cell culture were added to a screw cap tube containing 200μl of pre-warmed Glycerol (100%) to give a final concentration of 20% Glycerol. The samples were vortexed and stored at -80°C.

2.1.4 Determining the Concentration and Purity of Nucleic acids

The absorption of the DNA samples diluted in TE was measured using a Picodrop (Picodrop Limited). The purity of the sample is indicated as a sharp single peak and a A260/A280 ratio reading around 1.9.
2.1.5 DNA sequencing
Automated DNA sequencing was performed by Source BioScience (LifeScience) and the results were visualised using 4peaks (mekentosj.com). The sequence contigs were assembled and proof read using DNA Strider 1.4 for Lion operating system on Macintosh.

2.1.6 Agarose Gel Electrophoresis
Agarose gels were prepared by dissolving agarose at concentrations between 0.8%-2% (w/v) in 1xTAE (0.04M Tris-acetate, 1mM EDTA) or 1xTBE (90mM Tris, 90mM Boric acid, 2mM EDTA) and brought to boiling point in a microwave oven. EtBr was added to the melted agarose (0.5μg/ml) before the gel was poured into a gel cast containing a well comb and was allowed to set at RT. Gel loading buffer (40% sucrose (w/v), 0.25% bromophenol blue (w/v) or 0.25% xylene cyanol (w/v) in distilled H₂O) was added at one-tenth volume for each sample. The samples were then electrophoresed next to a DNA ladder (HyperLadder Bioline) at 80-100 V. The DNA bands were visualised on a UV transilluminator (Gbox imager, Syngene) and the images were captured using GeneSnap software (Syngene).

2.1.7 Acrylamide Gel Electrophoresis
6% acrylamide gels were prepared by diluting 4ml of 10xTBE buffer and 6ml of 40% acrylamide/bisacrylamide (Appligene) in 30ml deionised H₂O. Approximately, 6ml of the above mix is enough for one gel. Before pouring the gel, 4μl/ml of 20% ammonium persulphate and 2μl/ml of TEMED (Sigma) were also added to the solution. The solution was mixed quickly and the gel was poured between two glass plates with a comb in place. The gel was allowed to set and clamped into an upright gel tank filled with 1xTBE buffer. The gel was pre-run at 150V for 5mins before loading the samples and was run at 150V until the loading dye had almost reached the bottom of the gel. The gels were stained in 30ml of 1xTBE running buffer containing 20μl of 10mg/ml ethidium bromide for 15mins and the bands were visualized as previously described.
2.1.8 Restriction Digests
Restriction digests were performed using restriction enzymes purchased from NEB (New England BioLabs®) or Roche. Usually 2 units of the enzyme was used per 1μg DNA and the buffer recommended by each company. The sample was incubated at the optimum temperature required for each enzyme for approximately 3 hours.

2.1.9 Extraction of DNA from Agarose gels
Following restriction digests, the DNA bands were visualised on a UV transilluminator (Benchtop 3UV™ UVP) at 302nm. The selected bands were excised from the gel using a sterile scalpel and transferred to an eppendorf tube. DNA was purified from the extracted bands using NucleoBound® Extract II (Machery-Nagel Cat No. 740609.50) as per manufacturer’s instructions.

2.1.10 Ligation of DNA fragments
Standard 20μl ligation reactions contained 2μl 10x ligation buffer, 1μl T4 DNA ligase, ~50ng of vector DNA and insert DNA at a 3 fold molar excess. Reactions were incubated at RT for 3hrs. Ligations for cloning of PCR products performed using the pGEM-T Easy Vector Systems (Promega) were carried out as per manufacturer’s instructions, using an insert DNA concentration at a 5 fold molar excess.

2.1.11 Transformation into competent cells
A 50μl aliquot of competent cells was thawed on ice before the addition of DNA or 10μl of ligation mix to the cells. Following an incubation on ice for 15 min the cells were heat shocked at 42°C for 30 seconds and chilled on ice for 2 min. 1ml of LB media (10g/L Bacto tryptone, 5g/L Bacto yeast extract, 170mM NaCl pH7.0) was added to the cells that were then placed on a shaking incubator at 37°C 1200rpm for 1h. One-tenth of the transformation was plated out on an agar plate containing the appropriate antibiotic. The remaining volume was centrifuged at 4000 rpm for 5 mins, the cell pellet was
resuspended in the left over volume of LB media after the tube’s inversion and plated out on an agar plate containing the appropriate antibiotic.

**Table 2.1 Bacterial Strains**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Description</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH10B</td>
<td>F– mrcA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu) 7697 galU galK rpsL nupG λ–</td>
<td>Endonuclease deficient (endA) to prevent non-specific digestion yielding cleaner DNA and recombinase deficient (recA) to reduce unwanted recombination and improve insert stability</td>
<td>Standard Cloning</td>
</tr>
<tr>
<td>DH5α</td>
<td>F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1</td>
<td>Endonuclease deficient (endA), recombinase deficient (recA), contains HsdR for efficient transformation of unmethylated DNA</td>
<td>Cloning repetitive DNA</td>
</tr>
<tr>
<td>BL21-Al</td>
<td>F– ompT hsdSB(rB–, mB–) gal dcm araB:T7RNAP-tetA</td>
<td>Encodes T7 RNA polymerase under the control of the arabinose-inducible araBAD promoter</td>
<td>Cloning for protein expression</td>
</tr>
</tbody>
</table>

**2.1.12 Preparation of competent cells**

A single colony was picked from an agar plate to inoculate 10ml LB plus the appropriate antibiotic and was incubated in a shaking incubator at 37°C until the OD$_{600}$ was 0.5. This culture was then used to inoculate a 200ml LB culture plus appropriate antibiotic which was also incubated in a shaking incubator at 37°C until the OD$_{600}$ was 0.5. The culture was then placed on ice for 5 mins and centrifuged at 4000xg for 5min, 4°C. The cell pellet was resuspended in 80ml TFB I (30mM Potassium acetate, 100mM RbCl$_2$, 10mM CaCl$_2$.2H$_2$O, 50mM MnCl$_2$.4H$_2$O, 15% Glycerol (v/v), pH5.8 with acetic acid) and placed on ice for 15min. The cells were then centrifuged at 4000xg for 10 mins, 4°C and the cell pellet was resuspended in 8ml TFB II (10mM Mops, 75mM CaCl$_2$.2H$_2$O, 10mM RbCl$_2$, 15% Glycerol (v/v), pH6.5 with 1M KOH) and
placed on ice for 30min. The cells were then aliquoted into eppendorf tubes, snap frozen in liquid nitrogen and stored at -80°C.

2.1.13 PCR
Polymerase chain reactions were performed using either the proofreading DNA polymerase Pfu or Taq, both isolated and prepared in the lab by Brian McStay. A typical PCR reaction using Pfu contained ~50ng of the DNA template, 1U/50µl Pfu Pol, 0.2mM dNTPs, 0.2µM of each primer, 1M Betaine and a 1:10 dilution of 10xPfu Buffer (200mM Tris-HCl pH8.8, 20mM MgSO₄, 100mM KCl, 100mM (NH₄)₂SO₄, 1% Triton X-100, 1mg/ml BSA nuclease free).

The PCR conditions were generally 30 cycles as follows:
Denature- 95°C 30 seconds
Anneal- temperature gradient of 40-60°C for 30 seconds
Extension- 72°C 2 min/kb

A typical PCR reaction using Taq contained ~100ng of the DNA template, 0.2mM dNTPs, 5U/50µl Taq Pol, 0.2µM of each primer, 1M Betaine and a 1:10 dilution of 10xTaq Buffer (500mM KCl, 100Mm Tris-HCl [pH9.0], 15mM MgCl₂ and 1% TritonX-100).

The PCR conditions were generally 30 cycles as follows:
Denature- 95°C 15 secs
Anneal- temperature gradient of 60-70°C for 15 secs
Extension- 72°C 1 min/kb

Matercycler® Gradient (Eppendorf) was used to preform all PCR reactions.

2.1.14 Real Time PCR
The PCR reactions were performed using a DNA Engine Opticon® 2 System for Real-Time PCR Detection (Bio-Rad). Typical reactions (25µl) were composed of 23µl master mix that was exactly as the mix for the Taq Pol reactions described above, with the addition of 0.5µl 50mM MgCl₂ and a 1/17000 dilution of SYBR Green (Biogene). 2µl of the cDNA sample was added to the reaction.

Typical PCR conditions were 35 cycles as follows:
Denature- 95°C 15 secs
Anneal- temperature gradient of 60°C for 15 secs
Extension- 72°C 30secs
Each cycle was followed by a plate read before the final step which was followed by a melting curve from 65°C to 95°C with a read every 0.5°C held for 1sec.
The specificity of the products was determined by a combination of the melting curve profiles and gel electrophoresis. The standard curves for quantification were generated using serially diluted DNA from previously amplified PCR products for wild type cells. The results were normalised against the control.

2.1.15 Quantitation of PCR/protein products
Quantitation of products was performed by measuring the intensity (height) of the bands and normalizing against the background noise using GeneTools (Syngene) as per the manufacturer’s instructions.

2.1.16 Purification of PCR products
PCR products were purified using NucleoBond® Extract II (Machery-Nagel Cat No. 740609.50) as per manufacturer’s instructions.

2.1.17 Extraction of Genomic DNA from cultured human cells
After the cells grown in a T175 flask reached ~90% confluence, the media was removed and the cells were washed twice with 5ml PBS. 5ml TE (20mM Tris pH8, 2mM EDTA) were added to the flask followed by SDS to a final concentration of 0.5% and proteinase K to a final concentration of 0.3mg/ml. Following an O/N incubation of the flask at 37°C , the solution was then transferred to a 15ml tube and 1/10th volume 3M Sodium Acetate was added. An equal volume of Phenol/Chloroform was then added to the tube followed by rotation for 20mins at RT. The sample was centrifuged at 4000g for 10min and the aqueous layer was transferred to a fresh tube. 500µg RNase I was added to the sample before incubation at 37°C for 30min. After the phenol/chloroform extraction was repeated, the aqueous layer was transferred to a fresh tube, 2½ volumes of 100% ethanol was added and the sample was
incubated at -20°C for 20min to precipitate the DNA. The DNA was collected with a Pasteur pipette and desalted by dipping into 70% ethanol (v/v). the DNA was finally resuspended in the appropriate volume of TE (10mM Tris pH8, 0.1mM EDTA) depending on the amount of precipitated DNA.

2.2 RNA

2.2.1 RNA extraction from cultured human cells

After the cells were grown in a T175 flask until ~90% confluent, they were scraped off in ice-cold PBS and pelleted at 1200rpm for 5mins. The cell pellet was resuspended in 1ml TRIsure™ (Bioline) and incubated at RT for 5mins. 0.2ml of chloroform was added per 1ml of TRIsure™ and the tube was shaken vigorously for 15 seconds, incubated at RT for 15mins and centrifuged at 12000g for 15min at 4°C. The aqueous phase was transferred to a new eppendorf tube, 0.5ml isopropanol per 1ml TRIsure™ was added to the sample which was then incubated at RT for 10min and centrifuged at 12000g for 10min at 4°C. After all the supernatant was removed using a pipette, the RNA pellet was washed with 1ml ice-cold 70% ethanol and centrifuged at 12000g for 5min at 4°C. The pellet was then dissolved in 50-100µl Hyclone H₂O and the RNA sample was stored at -20°C.

2.2.2 cDNA Synthesis

cDNA was prepared from total cellular or nucleolar RNA samples with a Protoscript M-MuLV First Strand cDNA synthesis kit (New England Biolabs) using either random hexamer or oligo-dT primers as per manufacturer’s instructions. The presence of specific cDNAs using Reverse Transcriptase were detected by PCR using a Taq Polymerase protocol and were run in 1.5% TBE agarose gels.
Table 2.2 DJ transcript primers

<table>
<thead>
<tr>
<th>Transcript</th>
<th>DJ location/Exon and number if applicable</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>disnor138</td>
<td>137.6 forward</td>
<td>CCCCATCGTTTGGTCGCAGG</td>
</tr>
<tr>
<td>disnor138</td>
<td>137.4 reverse</td>
<td>GCCCCGTGCATGCTGGAATTG</td>
</tr>
<tr>
<td>disnor187 (AK026938)</td>
<td>155.0/exon3/forward</td>
<td>ACAGACCTGGAGACATGCTACAC</td>
</tr>
<tr>
<td>disnor187 (AK026938)</td>
<td>150.7/exon4/reverse</td>
<td>CCCGGCTCTAGTGATGCATCC</td>
</tr>
<tr>
<td>disnor 238 (BX647680)</td>
<td>238.5/exon1/forward</td>
<td>ACAGCCCCGTGTCCCTGCG</td>
</tr>
<tr>
<td>disnor 238 (BX647680)</td>
<td>242.6/exon3/reverse</td>
<td>CTGCTTCACCAAGCTCCCA</td>
</tr>
<tr>
<td>disnor 238 (BX647680)</td>
<td>DJ242.5/forward/(rest of exon 3)</td>
<td>GAGACCTTCCTTCCCCACGGGTT</td>
</tr>
<tr>
<td>disnor 238 (BX647680)</td>
<td>DJ244.3/reverse/(rest of exon 3)</td>
<td>TCGCTGGGGGAATCGGGGATG</td>
</tr>
<tr>
<td>12GAPDH</td>
<td>Control/forward</td>
<td>GAAGGTCGGAGTCAACGGATT</td>
</tr>
<tr>
<td>334GAPDH</td>
<td>Control/reverse</td>
<td>ATGAGCCCCAGCCTTCTCCAT</td>
</tr>
</tbody>
</table>

2.3 Tissue Culture

2.3.1 Cell Lines

The HT1080, male human fibrosarcoma cell line (Rasheed et al., 1974), was grown in Dulbeco’s MEM+GlutaMAX-1 (+ 4.5 g/L glucose; GIBCO) supplemented with 10% foetal bovine serum (v/v) (BioSera) and 5U/mL (100μg/ml) of penicillin/streptomycin (Sigma).

The DJ-clone stable cell lines that derived from HT1080 were all cultured in the above media with the addition of 5μg/mL blasticidinS (Melford).

The karyotypically normal non-transformed female human hTERT-immortalised retinal pigment epithelial (hTERT-RPE-1) cell line (Bodnar et al., 1998) was grown in DMEM/Nutrient Mixture F-12 Ham (+ 15mM HEPES, sodium bicarbonate, without L-Glutamine, (Sigma)) supplemented with 10% foetal bovine serum (v/v) (BioSera), 2mM L-Glutamine (v/v)(Sigma), 0.348% Sodium bicarbonate (v/v)(Sigma) and 5U/mL (100μg/ml) of penicillin/streptomycin (Sigma).
The HeLa-S3, female human cervix adenocarcinoma cell line (Puck and Fisher, 1956), was grown in Dulbecco’s MEM+GlutaMAX-1 (+ 4.5 g/L glucose; Gibco) supplemented with 10% fetal bovine serum (v/v) (BioSera), 1% (100X) MEM non-essential amino acid solution (Sigma) and 5U/mL (100μg/ml) of penicillin/streptomycin (Sigma).

The A9-13, A9-14, A9-15 and A9-22, mouse/human monochromosomal hybrid cell lines (Cuthbert et al., 1995) (Source: Newbold RF, Brunel University) each containing a single human chromosome, were grown in Dulbecco’s MEM+GlutaMAX-1 (+ 4.5 g/L glucose; Gibco) supplemented with 10% fetal bovine serum (v/v) (BioSera), 5U/mL (100μg/ml) of penicillin/streptomycin (Sigma) and 400U/ml (50mg/ml) Hygromycin B (Gibco).

The GM09142 and GM10063 mouse/human monochromosomal hybrid cell lines (Worton et al., 1984) (Source: Coriell Institute) each contain a human chromosome from the translocation between chromosomes X and 21. The GM09142 were grown in Alpha Modification MEM (Eagle), 15% fetal bovine serum (v/v) (BioSera), 1% L-Glutamine 200mM (v/v)(Sigma) and 5U/mL (100μg/ml) of penicillin/streptomycin (Sigma). The GM10063 were grown in Alpha Modification MEM (Eagle), 15% foetal bovine serum (v/v) (BioSera), 1% L-Glutamine 200mM (v/v)(Sigma), 5U/mL (100μg/ml) of penicillin/streptomycin (Sigma) and 10ml HAT media supplement (50X) Hybri-Max for a final concentration of 0.1 mM sodium hypoxanthine, 0.4 μM aminopterin, 16 μM thymidine (Sigma).

2.3.2 Maintaining cell lines
The cells were cultured in T75 or T175 culture flasks. To transfer the cells the media was removed and the cells were washed with 1XPBS (Phosphate Buffered Saline). They were then trypsinised with a final concentration of 1xTrypsin (Sigma) and left at 37°C for 5min. The flask was gently tapped to detach cells from the bottom and an equal volume of media was added to inhibit trypsin. The cells were then transferred to 15ml tubes and centrifuged.
at 1200rpm for 5 min and the pellet was resuspended in 5ml medium and when single cell suspension was achieved they were re-seeded at required dilution into new flasks containing media.

2.3.3 Generating the DJ clones stable cell line
Transfections were performed using a standard calcium phosphate protocol in 150mmx20mm plates. HT1080 cells were transfected with 20μg of bacmid CH507-145C22 (CU633906, Chr 21) or with the same final concentration of the three bacmids RP11-337M7 (AL592188, Chr 22), CH507-535F5 (CT476834, Chr 21) and RP11-272E10 (AC011841) separately or in couples or all three together. A blasticidin selection marker in a 200:1 w/w ratio was also used to aid selection which was carried out using DMEM medium supplemented with 5μg/mL blasticidin (GIBCO). The colonies were picked approximately 12 days later using cloning discs (SIGMA) and transferred to 6-well plates. The clones were maintained in blasticidin containing media and screened by DNA FISH on metaphase and interphase cells to determine the size of insertion. The clones kept for further research were the ones that the exogenous DJ was clearly distinguished from the endogenous. Some negative clones were also kept for controls.

2.3.4 Long Term Storage and Recovery of Cell Lines from Liquid N₂
After trypsinising the cells in a T175 flask and centrifuging them at 1200 rpm for 5min, the cell pellet was resuspended in 3mls of freezing medium (10% DMSO (v/v) 90% FBS (v/v)) and 1ml cell suspension aliquots transferred to cryotubes. The tubes were then placed in Nalgene Cryo 1°C and stored at -80°C O/N before transferred to liquid nitrogen tanks.
To recover the cells from liquid nitrogen, the cryotube containing frozen cells was thawed quickly at RT and the cells were transferred to 15ml tube containing 5mls media and centrifuged at 1200rpm for 5min. The cell pellet was resuspended in 5mls media and transferred to a T75 flask.
2.3.5 Transfections

To generate stable cell lines transfections were performed using a standard calcium phosphate protocol in 150mm x20mm plates. The cells were seeded the day before so that the cell density was approximately 60%. 2xHEBS (270mM NaCl, 10mM KCl, 1.4mM Na₂HPO₄, 10mM Dextrose, 42mM HEPES pH 7.05) and CaCl₂ (2.5M in 10mM HEPES pH 7.2) were thawed at RT. TE (pH 7.15) was also allowed to warm at RT. For each 15cm plate, 800µl 2xHEBS was placed in a sterile 15ml tube. 20µg of DNA was added to an eppendorf tube as well as 100ng of pJRC41 that contains a blasticidin resistance marker. The volume was brought up to 720µl with TE (pH 7.15). The sample was mixed by pipetting a few times before 80µl of 2.5M CaCl₂ was added to the tube. Mixing was repeated before the sample was added dropwise to the 15ml tube. The DNA/CaPO₄ solution was incubated at RT for approximately 10min before added to the cells dropwise. The dish was agitated to ensure mixing and placed in the incubator (37°C/5% CO₂). After 7 hours of incubation a fine precipitate should be clearly visible under the light microscope. The media was then removed and the cells were washed with 1X PBS until the precipitate was no longer visible. Fresh warm media was added to the dish which was placed again in the incubator. Following 24 hrs of transfection the appropriate selection marker was added to the media at a concentration 5µg/mL BlasticidinS. Colonies were picked approximately 10 days later using 5mm cloning discs (SIGMA) and were transferred to 6-well plates containing the appropriate media as above(5µg/mL blasticidinS).

2.3.6 siRNA tranfections

Transfections of the BaZ2A siRNA oligo into the HT1080 or HeLa cell lines were performed using siRNA duplexes purchased from DharmaFECT as per manufacturer’s instructions. The siRNA duplexes were transfected using a final concentration of 20 nM with DharmaFECT 1 reagent (Thermo Scientific) at 24 and 72-h after seeding. The cells were then harvested 24 hours after the second round of siRNA transfection.
Table 2.3 BAZ2A duplexes

| BaZ2A - combination of SMARTpool duplexes | GAUAAGACACCAAGACGUAAUU 5'PUACGUCUUGGUGUCUUUAUCUU 5'PAUUGCCUGAAUCCUCGACGUU GUGGAGCAUAUGGUAAUUU 5'PAUACCAUAUAGUCCUCACUU GGAGACGGCGUUGGGAUAUUU 5'PUAUCCCCAAGCGUCUCCCUU |
| Negative control duplex | 5'GUAAACAAUGAGAGACCGGCdTdT3' dTdTCAUUGUUACUGCUGCCG5' |

2.4 Cell Biology Techniques

2.4.1 Immunofluorescence on Interphase cells

The cells were grown on Superfrost® Plus microscopic slides (Scientific Laboratory Supplies) for at least 24hr. After the removal of the media the cells were fixed with 4% paraformaldehyde (PFA) (w/v) in PBS for 10min at RT. The cells were then rinsed with PBS for 10minX3 at RT. They were then permeabilized on the slides with 0.5% Saponin (w/v) and 0.5% Triton X-100 (v/v) in PBS for 10min at RT before being incubated in 20% glycerol/PBS (v/v) for 2 hours at 4°C. The slides were then snap frozen in liquid N₂ and stored at -80°C. The slides were thawed in PBS for 10min at RT and the antibodies were diluted from 1:50 to 1:200 in 1% bovine serum albumin (BSA) (w/v) in PBS. 100μl of the primary unconjugated antibody was applied to each slide and incubated for 60min in a humidity chamber at 37°C. The slide was then rinsed for 10minX3 with PBS at RT followed by incubation with 100μl of the conjugated secondary antibody for 60min in a humidity chamber at 37°C. Finally the cells were rinsed for 10minX3 with PBS at RT and the slides were mounted in Vectashield plus DAPI (Vector Laboratories).

2.4.2 DNA FISH on Interphase cells

The cells were grown on Superfrost® Plus microscopic slides (Scientific Laboratory Supplies) for at least 24hr. After the removal of the media the cells were fixed with 4% paraformaldehyde (PFA) (w/v) in PBS for 10min at RT. The cells were then rinsed with PBS for 10minX3 at RT. They were then
permeabilised on the slides with 0.5% Saponin (w/v) and 0.5% Triton X-100 (v/v) in PBS for 10min at RT before being incubated in 20% glycerol/PBS (v/v) for 2 hours at 4°C. Hybridisation probes were labelled with spectrum green or red dUTP using Nick Translation kit (Vysis) as per manufacturer’s instructions. Before hybridisation the slides were thawed in PBS for 10min at RT and then depurinated in 0.1M HCl for 10min at RT and washed for 5minX2 in PBS at RT. The cells were equilibrated with 200μl of 50% deionised formamide/2xSSC for 15min in a humidity chamber at 37°C before depurination. The precipitated DNA probe (50ng/slide) was resuspended in Hybrisol VII (MP Biomedicals) (25μl/slide) and applied to a coverslip on a 37°C block which was then applied to the cells. The sides of the coverslip were sealed with rubber cement (Marabu-Fixogum) before the slides were denatured at 73°C for 12mins and allowed to hybridise for at least 18h in a humidity chamber at 37°C. After hybridisation the slides were washed in 50% formamide/2xSSC at 42°C for 5minX3 following a wash in 0.1x SSC at 60°C for5minX3, they were quickly dried and mounted in Vectashield plus DAPI (Vector Laboratories).

2.4.3 Combined Immunofluorescence and FISH on Interphase cells
After the last wash of DNA FISH, the slides were washed in PBS for 5min at RT followed by antibody staining. When using Fibrillarin monoclonal antibody, the antibody staining was performed with primary and secondary antibodies, followed by fixation with 2% PFA in PBS for 10min, prior to depurination and hybridisation.

2.4.4 Preparing Metaphase spreads
The cells were grown in a T175 flask to approximately 70% confluence and were treated with Demecolcine (Sigma) at a final concentration of 0.1μg/ml for 1hr. The mitotic cells were recovered by a soft shake off of the flask, were transferred to a 50ml tube and centrifuged at 1700rpm for 7mins at 4°C. The cell pellet was resuspended in 3mls of hypotonic solution (75mM KCl) and incubated at 37°C for 10mins. 3-5 drops of ice cold Camoy’s fixative solution (3:1 (v/v) methanol/glacial acetic acid) were added to the cells and they were
again centrifuged at 1700rpm for 7mins at 4°C. The cell pellet was resuspended in 7mls ice cold fixative solution and the cells were centrifuged at 1700rpm for 7mins at 4°C. This step was repeated one more time before the cell pellet was resuspend in 3mls fixative solution. The mitotic cells were dropped using a Pasteur pipette from approximately 10cm high onto the superfrost slides and were allowed to air dry before hybridisation.

2.4.5 DNA FISH on metaphase spreads
The precipitated probe (50ng/slide) was resuspended in Hybrisol VII (Qbiogene) (25μl/slide) and applied to a coverslip on a 37°C block which was then applied to the cells. The sides of the coverslip were sealed with rubber cement (Marabu-Fixogum) before the slides were denatured at 73°C for 3mins and allowed to hybridise for 24-48hrs in a humidity chamber at 37°C. After hybridisation the slides were washed in 0.4xSSC/0.3% NP-40 for 2mins at 73°C followed by 2xSSC/0.3% NP-40 at RT. The slides were mounted in Vectashield plus DAPI (Vector Laboratories). The slides of human normal male metaphase chromosome spreads (Applied Genetics) were denatured in 70% formamide/2X SSC at 73°C for 5mins, dehydrated through a 70%-100% ethanol series and air dried. The probe (50 ng/slide) combined with human COT-1 DNA (10μg/slide) in 20μL/slide Hybrisol VII (MP Biomedicals) was denatured at 73°C for 5mins, added to the slides and allowed to hybridize for 24-48 h at in a humidity chamber 37°C. The 5’FITC labelled CER probe was added in the probe mix at the end before the denaturation of the slides.

2.4.6 RNA FISH and Combined Immunofluorescence
The protocols for RNA FISH and immuno-RNA FISH experiments differ to the DNA FISH only in that the slides were neither depurinated nor denatured before the application of the denatured probes.

2.4.7 Combined FISH and silver staining on metaphase spreads
FISH on metaphase spreads were carried out as outlined above. After the post hybridisation washes the slides were washed in PBS for 5min at RT and were allowed to air dry. 100μl of 2% Gelatine/0.1% formic acid was added to
the slide followed by the addition an equal volume of 50% silver nitrate. A coverslip was applied on the slide and incubated on a 60°C block until the solution had turned golden/dark brown. The slide was then rinsed briefly in water, was allowed to air dry and mounted in Vectashield plus DAPI (Vector Laboratories).

2.4.8 Imaging of fixed cells
Z-stacks of fluorescent images were captured and merged using a Photometric Coolsnap HQ camera and Volocity 5 imaging software (Improvision) with a 63x Plan Apochromat Zeiss objective mounted on a Zeiss Axioplan2 imaging microscope. The number and the depth of Z-stacks taken per image were calculated by establishing the midpoint of the cell by moving the microscope’s focus upwards and downwards until the cell was in focus. This midpoint was then set as zero in the software and the upper and lower limits were set by moving the focus again until the cell was out of focus. The average depth for interphase cells was ~4µm and the depth limit of each Z-stack was 0.2µm. Usually the Z-stack depth for interphase cells was ~50 Z-stacks and ~10 for metaphase chromosome spreads. Once the channels for imaging were selected i.e. DAPI, FITC and Rhodamine, the image was captured and deconvolved using the Volocity 5 software. Z-stacks of fluorescent images were captured using a Photometric Coolsnap HQ camera and Volocity 5 imaging software (Improvision) with a 63 Plan Apochromat Zeiss objective mounted on a Zeiss Axioplan2 imaging microscope. In some cases extended focus projections of deconvolved Z-stacks (iterative restoration) are presented in other cases individual focal planes are shown.

2.5 Nucleolar Pellets
2.5.1 Nucleolar Isolation from cultured cells
The cells (10X15cm dishes) were grown to approximately 90% confluence and were washed in PBS, scrapped off, resuspended in 40mls PBS in a 50ml tube and centrifuged at 1200rpm for 5min at 4°C. The cell pellet was washed in 25ml PBS and centrifuged again at 1200rpm for 5min at 4°C. The wash was repeated once more before the cell pellet was resuspended in 7ml of
Buffer A (10mM HEPES-KOH [pH7.9], 1.5mM MgCl₂, 10Mm KCl, 0.5mM DTT plus protease inhibitors). After incubation on ice for 5min the nuclei were released using a dounce tissue homogenizer with a tight pestle. Their release was microscopically monitored. The released nuclei were transferred to a new 15ml tube and were centrifuged at 1200rpm for 5 mins at 4°C. The pellet was resuspended in 3ml high-magnesium buffer, S1 Buffer (0.25M sucrose, 10mM MgCl₂ plus protease inhibitors), was layered on top of 3ml low-magnesium buffer S2 Buffer (0.35M sucrose, 0.5mM MgCl₂ plus protease inhibitors) and was centrifuged at 1430g for 5 mins at 4°C. The pellet was then resuspended in 5mls S2 Buffer and sonicated 5x10sec intervals at 10amp using Soniprep 150 (MSE) with a fine probe. The solution was then layered on top of S3 Buffer (0.88M sucrose, 0.5mM MgCl₂ plus protease inhibitors) and was centrifuged at 3000g for 10 mins at 4°C. The pellet was resuspended in 20/2TE (20mM Tris [pH8.0], 2mM EDTA) of 1-2X10⁶ cell equivalents/ml, 850µl were aliquoted in each tube and stored at -80°C (Leung et al., 2004).

2.5.2 Nucleolar DNA Combing
Combing was performed as per manufacturer’s instructions provided by Genomic Vision with a few modifications stated below. Briefly, the resuspended nucleoli were mixed with an equal volume of 1.5% low melting point agarose in PBS at 50°C and the mixture was pipetted into a plug mould (BioRad, 100 µL/slot) which was refrigerated for at least 1 hr before removal of the plugs. The embedded nucleoli were deproteinised in 250µl/plug ESP buffer [450mM EDTA, 1% sarkosyl, 0.25mg/ml Proteinase K (Eurobio Laboratories)] at 50°C O/N. The agarose plug was then melted at 68°C for 20 mins and the agarose was removed with the addition of 2µl of β-agarase (1U/µl, New England Biolabs) and an incubation at 42°C O/N. The encapsulated high molecular weight nucleolar DNA was combed onto silanized coverslips (Genomic Vision Paris) using a Molecular Combing apparatus (Genomic Vision Paris). Initially, until commercial coverslips became available, I tried to clean and silanise the coverslips (22x22 mm) myself using a liquid-phase silanisation protocol by Goldar and colleagues (Goldar et al., 2008). However this procedure was aborted since there was no major difference in the results and due to it being a dangerous procedure by
using the explosive “piranha solution” (hydrogen peroxide and sulphuric acid). The coverslips were then hybridised with biotin (Invitrogen) and digoxigenin (Roche) labelled probes as per the manufacturer’s instructions. Hybridisation and detection was performed as described previously in Caburet et al., 2005. The rDNA fragments are very small and highly repetitive which was causing background hybridization problems. Therefore the protocol was modified by using purified nucleoli instead of cells to make the agarose plugs. Also, the concentration per plug (90μl) was increased from $10^4$ cells/μl to 2.5x$10^6$ nucleoli/μl. This way I got more fibres per coverslip and less background hybridization.

2.5.3 Preparation of soluble formaldehyde cross-linked nucleolar chromatin

The cells (10X15cm dishes) were grown to approximately 90% confluence and were cross-linked by 0.2% formaldehyde for 10 mins at RT on a rocker. The cells were washed in PBS, scrapped off, resuspended in 40mls PBS in a 50ml tube and centrifuged at 1200rpm for 5 mins at 4°C. The cell pellet was washed in 25ml PBS and centrifuged again at 1200rpm for 5 mins at 4°C. The wash was repeated once more before the cell pellet was resuspended in 7ml of Buffer A (10mM HEPES-KOH [pH7.9], 1.5mM MgCl$_2$, 10Mm KCl, 0.5mM DTT plus protease inhibitors). After incubation on ice for 5 mins the nuclei were released using a dounce tissue homogenizer with a tight pestle. Their release was microscopically monitored. The released nuclei were transferred to a new 15ml tube and were centrifuged at 1200rpm for 5 mins at 4°C. The pellet was resuspended in 3ml high-magnesium buffer, S1 Buffer (0.25M sucrose, 10mM MgCl$_2$ plus protease inhibitors), was layered on top of 3ml low-magnesium buffer S2 Buffer (0.35M sucrose, 0.5mM MgCl$_2$ plus protease inhibitors) and was centrifuged at 1430g for 5 mins at 4°C. The pellet was then resuspended in 5ml S2 Buffer and sonicated 5x10 sec intervals at 10amp using Soniprep 150 (MSE) with a fine probe. The solution was then layered on top of S3 Buffer (0.88M sucrose, 0.5mM MgCl$_2$ plus protease inhibitors) and was centrifuged at 3000g for 10 mins at 4°C. The pellet was resuspended in 0.2ml of 20/2TE (20mM Tris [pH8.0], 2mM EDTA) following an addition of 1/10$^{th}$ of the volume of 20% sodium dodecyl sulfate (SDS).
Following incubation at 37°C for 15 mins, 1ml 20/2TE was added and the solution was sonicated for another 2x10 sec bursts at 5amp. The sample was centrifuged again at 14,000rpm for 1 min at 4°C. The resulting sheared nucleolar chromatin was used immediately in nucleolar ChIP assays or 100µl aliquots were snap frozen in liquid nitrogen and stored at -80°C.

The formaldehyde cross-links were reversed by adding 2.5µl Proteinase K (10mg/ml, Roche) to 50µl of nucleolar sheared chromatin and incubating the mix at 62°C O/N. 10µl 3M Sodium acetate was added to the mixture. Phenol-chloroform extraction was performed before the DNA was recovered by ethanol precipitation and was resuspended in 50µl of TE (10 mM Tris [pH 8.0], 0.1 mM EDTA) and incubated at 37°C for approximately 20 mins. This would be the input control for ChIP and FAIRE. Using specific primers as controls Taq polymerase PCR was performed on the 2µl of the sample which was then run on a 1.5% TBE agarose gel to check that the nucleolar chromatin was of the appropriate size range using a UV transilluminator (Gbox imager, Syngene).

**2.5.4 Nucleolar Chromatin Immunoprecipitation (ChIP)**

Pre-equilibration of 5µl protein A or G Dynabeads (10mg/ml, Dynal) in 20µl PBS containing 0.1% IgG free bovine serum albumin (BSA), 100µg/ml of Herring Sperm DNA was performed in a shaking incubator for 1 hr at 4°C at 900rpm. The beads were magnetically recovered and incubated in a shaking incubator for 2 hrs at 4°C at 800rpm after the addition of 20µl PBS with the appropriate Antibody (α-H3K4me3 or α-CTCF or α-UBF or α-Pol I or α-Pol II) (25mg/ml). The beads were magnetically recovered, washed in 20µl PBSX2 and were added 150µl (90µl of the nucleolar sheared chromatin diluted in 585µl IP Dilution Buffer (14.5mM Tris-HCl [pH7.6], 0.75% Triton X-100 and 182.5mM NaCl)). The sample was then incubated in a shaking incubator at 4°C at 700rpm O/N. Once again the beads were magnetically recovered and washed X2 in 150µl IP Wash Buffer I (20mM Tris-HCl [pH8.0], 2mM EDTA, 0.2% SDS, 0.5% Triton X-100, and 150mM NaCl) for 5 mins at 4°C at 700rpm on a shaking incubator. The beads were magnetically recovered and washed
X2 in 150µl IP Wash Buffer II (20mM Tris-HCl [pH8.0], 2mM EDTA, 0.2% SDS, 0.5% Triton X-100, and 500mM NaCl) for 5 mins at 4°C at 700rpm on a shaking incubator. The immunoprecipitated material was eluted from the beads with X2 washes in 25µl each of 20/2TE containing 2% SDS at 37°C for 10 mins. Another 50µl 20/2TE was added to the pooled eluates and the proteins were digested with 5µl proteinase K (14mg/ml; Roche) at 62°C O/N to reverse the formaldehyde cross-links. 1/10th 3M Sodium acetate was added to the mixture which was then extracted with phenol-chloroform. 1 µl Glycogen (10mg/ml) was added as a carrier and the DNA was recovered by ethanol precipitation and resuspended in 50µl of TE for PCR analysis.

Table 2.4 DJ primers for mapping and ChIP experiments

<table>
<thead>
<tr>
<th>Kb distal to rDNA</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4</td>
<td>ACAACGCAAGGAAAAGCGACACC</td>
<td>GGCAACGTGGGACTTGACCCT</td>
</tr>
<tr>
<td>40.6</td>
<td>TGCTGATCCCCTTTTTGTC</td>
<td>GAAGGTGGTGTCGGTGAGAT</td>
</tr>
<tr>
<td>93.9</td>
<td>CCCTTCAGTGCTCACACACCG</td>
<td>CCAGCAGCTGGATCGCAAGG</td>
</tr>
<tr>
<td>106.1</td>
<td>TCTCCAGTGACACCTTGCTGCT</td>
<td>GACGTGAAACGAGTCGCAGCC</td>
</tr>
<tr>
<td>111 (AL581856)</td>
<td>TCTGGAGTGGTTCATAACAGTC</td>
<td>CTGAGTGTACTGATAGCCCAAGC</td>
</tr>
<tr>
<td>138.0</td>
<td>TCCCCGCAACACGCAGCGAGG</td>
<td>AATCCGACATGCACGGG</td>
</tr>
<tr>
<td>153.5 and 275.9</td>
<td>GTGTCATACATATGTCAC</td>
<td>CCCATTCTCTCCGAGCTGTC</td>
</tr>
<tr>
<td>196.4 and 229.2</td>
<td>GCCCAATGCCTGCCGCAAC</td>
<td>CGCTCAAGAGCCCAGCAC</td>
</tr>
<tr>
<td>204.1</td>
<td>AACAGGCACCTGCATTGGGGGA</td>
<td>AGCACACTCAAAGAGGGTCAAGG</td>
</tr>
<tr>
<td>292.0</td>
<td>TGGGTCTAGGGAGAAGTCCGC</td>
<td>TGCGACCAAAGGGCTGGGAG</td>
</tr>
<tr>
<td>340.2</td>
<td>GGCATTCGCTCCTGTTGTCCT</td>
<td>CACAGGGCCGGGCATTGTCAG</td>
</tr>
<tr>
<td>Kb proximal to rDNA ~9 kb</td>
<td>CATACTACCTTGTCCTCCAG</td>
<td>GTCTGTTCAACATTTCCATG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rDNA repeat</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>5ETS</td>
<td>GGCCTGTTGAGGACTGAG</td>
<td>ACGTGCCTACCCAGAGGAGCAG</td>
</tr>
<tr>
<td>18S</td>
<td>CGACGCCAAGCTCTGACG</td>
<td>CTCTCCCCAATCGAACCCTGA</td>
</tr>
</tbody>
</table>

2.5.5 Formaldehyde-Assisted Isolation of Regultory Elements (FAIRE)

10µl 3M Sodium acetate was added to the 50µl of nucleolar sheared chromatin and phenol chloroform extraction was performed. The DNA was recovered by ethanol precipitation, resuspended in 50µl of TE (10 mM Tris
[pH 8.0], 0.1 mM EDTA) and incubated at 37°C for approximately 20min. The samples were then used for PCR analysis.

Table 2.5 Primary Antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Species raised in</th>
<th>Antigen</th>
<th>Source/Company</th>
<th>Dilution for Immunoflorescence (IF)/Western Blot (WB)/Chromatin Immunoprecipitation (ChIP)</th>
</tr>
</thead>
</table>
| α-hBaz2A   | sheep             | N-terminal encoding from amino acids 158 to 489 Baz2A expressed protein (Ioanna Floutsakou) | McStay Lab and Diagnostic Scotland Edinburgh | IF:1/50  
WB: 1/50                                                                                     |
| α-hPAF49   | Sheep             | Full-length baculovirus expressed recombinant protein (C.Mais) | McStay Lab and Diagnostic Scotland Edinburgh | IF:1/200  
ChIP: 1/200                                                                                 |
| α-hRPA43   | Sheep             | Full-length baculovirus expressed recombinant protein (C.Mais) | McStay Lab and Diagnostic Scotland Edinburgh | IF:1/200  
ChIP: 1/200                                                                                 |
| α-Fibrillarin | Mouse monoclonal | Provided by U Sheer, Wuerzburg                  | IF:1/200                                    |
| α-NOP52    | rabbit            | Novis                                         | IF:1/200                                    |
| α-hUBF     | sheep             | Full-length baculovirus expressed recombinant protein (J.Wright) | McStay Lab and Diagnostic Scotland Edinburgh | IF:1/200                                                   |
| Histone H3K4me³ | Mouse monoclonal | Millipore                                     | ChIP:1/200                                  |
| Mouse IgG control | Mouse     | Millipore                                     | ChIP:1/200                                  |
| Rabbit IgG control | Rabbit   |                                               | ChIP:1/200                                  |
| Streptavidin/Cy3 | Rockland  |                                               | Combing IF: 1/25                            |
| α-digoxigenin | sheep             | Roche                                         | Combing IF: 1/50                            |

69
Table 2.6 Secondary Antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Species raised in</th>
<th>Source/Company</th>
<th>Dilution for Immunofluorescence/Western Blot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhodamine α-rabbit</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200</td>
</tr>
<tr>
<td>Rhodamine α-mouse</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200</td>
</tr>
<tr>
<td>Rhodamine α-sheep</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200</td>
</tr>
<tr>
<td>Rhodamine α-human</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200</td>
</tr>
<tr>
<td>FITC α-mouse</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200</td>
</tr>
<tr>
<td>FITC α-sheep</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200 Combing IF: 1/25 Combing: Rockland</td>
</tr>
<tr>
<td>FITC α-human</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>IF:1/200</td>
</tr>
<tr>
<td>FITC α-donkey</td>
<td>Rabbit</td>
<td>Rockland</td>
<td>Combing IF: 1/50</td>
</tr>
<tr>
<td>α-sheep HRP</td>
<td>Mouse monoclonal</td>
<td>Sigma-Aldrich</td>
<td>WB:1/2000</td>
</tr>
<tr>
<td>α-rabbit HRP</td>
<td>Donkey</td>
<td>Jackson ImmunoResearch</td>
<td>WB:1/5000</td>
</tr>
<tr>
<td>Biotin α-streptavidin</td>
<td>rabbit</td>
<td>Combing: Rockland</td>
<td>Combing IF: 1/50</td>
</tr>
</tbody>
</table>

2.6 Protein

2.6.1 Harvesting Protein Samples from cultured cells

The cells were grown in 6cm dishes until they were fully confluent. They were then scraped off the plate and washed with 0.5mls PBS before transferred to an eppendorf tube. The plate was washed once more with another 0.5mls PBS that was transferred in the same eppendorf. The cells were then centrifuged at 1200rpm for 5mins at 4°C, and the pellet was washed with PBS. The cells were centrifuged again and the final cell pellet was resuspended in approximately 400μl of Laemmli Buffer (Bio-Rad) depending on the cell pellet size. To sheer high molecular weight DNA, the samples were boiled for 10mins at 95°C followed by sonication in a Bioruptor waterbath.
sonicator for 3-5x30sec bursts at the highest magnitude until the samples were not viscous.

2.6.2 Western Blotting

Electrophoresis of the proteins was performed on a NuPAGE® Novex 4-12% Bis-Tris gels (Invitrogen™) in an Xcell SureLock™ Mini-cell (Invitrogen™) filled with NuPAGE® running buffer (Invitrogen™) as per manufacturer’s instructions. The protein gels were either stained using Coomassie brilliant blue (PhastGel® BlueR (Sigma), 10% acetic acid, 40% Methanol) and destained before analysing in the transilluminator in Destain Buffer (40%Methanol, 10% Acetic Acid and 50% De-ionised H2O) or were transferred to a Hybond™ECL™ nitrocellulose membrane (Amersham) using an xCell II™blot module (Invitrogen™) filled with NuPAGE® transfer buffer (Invitrogen™) as per manufacturer’s instructions. Following blocking in 5% powder milk (Marvel) (w/v) in PBS at 4°C O/N, the membrane was incubated with the primary antibody diluted in 5% powder milk (Marvel) (w/v) at 4°C O/N. The membrane was then washed for 10minX3 in PBS at RT and incubated in the secondary antibody coupled to HRP (horseradish peroxidase) diluted in 5% powder milk (Marvel) (w/v) for 1h at RT. The membrane was washed again for 10minX3 in PBS at RT and equal volumes of Western Lightning™ Plus-ECL Enhanced Luminol Reagent and Oxidising Reagent (PerkinElmer) were mixed and applied to the membrane for 5min. The chemiluminescence signals were detected using a CCD Syngene G-Box chemi XT16 camera, the images were captured using GeneSnap software from Syngene and the quantification was performed using GeneTools software from Syngene.

2.6.3 Protein concentration detection

Protein concentrations were determined spectrophotometrically using the Bradford Protein assay dye reagent concentrate (Bio-Rad) as per manufacturer’s instructions.
2.6.4 Expression and purification of 6xHis-tagged proteins

The part of the Baz2A open reading frame from the N-terminal previously cloned into vector pENTR4 was cloned into the His-tagged vector pDEST17 as described in Table 2.8. The plasmid was then transformed into DH10 cells competent cells. The positive clone was selected after the transformants were screened by restriction digests with the appropriate enzymes that were analysed on an agarose gel. BL21 competent cells were then transformed using DNA from the positive clone and grown on agar plates plus the appropriate antibiotic. To confirm the expression of the protein, one clone was inoculated in 50ml LB plus the appropriate antibiotic and grown at 37°C in a shaking incubator until it reached an OD of 0.4. 5ml were kept as an uninduced control. 1/100 20% L-Arabinose was added in the rest 45ml and the tube was incubated at 37°C in a shaking incubator for another 2hrs. After transferring into two eppendorf tubes 1ml of the culture, the cells were centrifuged at 4000Xg at 4°C for 10min and the pellet was stored at -20°C. The cells in the eppendorf tubes were also centrifuged at 6000Xg at 4°C for 10min and the pellet of the one tube was resuspended in 5ml of Native Protein Extraction Buffer (0.1M Tris pH7.9, 10% glycerol (v/v), 0.5M NaCl, 1% NP-40 (v/v), 1mM DTT and half a tablet protease inhibitors [complete protease inhibitor minus EDTA cocktail tablets, Roche]) and sonicated in a Bioruptor waterbath sonicator for 3-5x30sec bursts at the highest magnitude until the sample was not viscous. Equal volume of Laemmli Buffer (Bio-Rad) was added to 100µl of the lysate before it was boiled at 95°C for 10min and kept on ice till loaded on the protein gel. The remaining 5ml were aliquoted in 1ml eppendorf tubes and centrifuged at 6000xg for 10min at 4°C. 100µl of the supernatant which contained the soluble protein was transferred to a new eppendorf tube and an equal volume of Laemmli Buffer was added before it was boiled at 95°C for 10min. 200µl of 1X Laemmli Buffer was added In the pellet of the second eppendorf tube as well as in the uninduced control pellet. Electrophoresis and analysis of the protein samples was performed as described in section 2.1.14. The protein’s sequence was verified by DNA sequencing. The protein was purified by NiNTA affinity chromatography under native conditions.
The plasmid expressing the protein was extracted from a large culture as described earlier. The pellet was resuspended in 25mls Native Protein Extraction Buffer (0.1M Tris pH7.9, 10% glycerol (v/v), 0.5M NaCl, 1% NP-40 (v/v), 1mM DTT and 25µl 1000xPMSF) and sonicated in a sonicator using the thick probe for 3-5x30sec bursts at 13 microns with a 30sec wait, until the sample was clearer. 2ml Ni-NTA nickel agarose beads (Qiagen) were washed twice by the addition of 7ml Native Extraction Buffer and centrifugation at 1000xg for 1min at 4°C. The sonicated solution was centrifuged at 4000xg for 15min at 4°C to pellet the cell debris. The supernatant was added to pre-equilibrated Ni-NTA nickel agarose bead and rotated for 1hr at 4°C. After centrifugation at 1000xg for 30sec the supernatant was transferred to a new 50ml tube. The beads were washed twice with 25ml 10mM Imidazole in Native Extraction Buffer and centrifuged at 1000xg for 30sec. The beads were then resuspended in 5ml 10mM Imidazole and transferred into a column (Bio-Rad Poly-Prep® chromatography column) and the his-tagged Baz2A protein was eluted by increasing the Imidazole concentration to 250mM. The eluted protein was collected in 1ml fractions and 5µl of each were tested by Bradford assay to calculate the protein concentration. The peak fractions were pooled together and dialysed into CB-100 Buffer (20mM Hepes pH7.9, 20% glycerol (v/v), 0.2mM EDTA, 0.1mM DTT, 0.1mM PMSF containing 100mM KCl) O/N at 4°C. Dialysis was performed with the use of a Dialysis Tubing membrane (GIBCO-BRL) and 2 dialysis clips to close both ends of the tube that was rotated in the CB100 buffer O/N at 4°C. The following day the buffer was renewed and the tube was left in the buffer for another 8hr. The protein was then transferred into screw-cap eppendorfs and its concentration was calculated by Bradford assay. 1ml i.e. 4mg of protein was sent to Diagnostic Scotland Edinburgh, for an antibody to BaZ2A to be raised in sheep.
<table>
<thead>
<tr>
<th>Construct</th>
<th>How construct was generated</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baz2ApENTR4DEST17</strong></td>
<td>LR recombination reaction between Baz2A and pDEST17 (His-tag)</td>
<td>Expression of N terminally his-tagged BaZ2A</td>
</tr>
<tr>
<td><strong>His-tagged (Ioanna Floutsakou)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>pDJ93HS SK-</strong></td>
<td>PCR performed on CT476834 BAC DNA using primers designed against the 138 repeat block. The PCR product was ligated into the pBluescript SK- (Stratagene) by standard ligation</td>
<td>To identify function of the ACRO138 repeats block in human cells</td>
</tr>
<tr>
<td><strong>pDJ138HS SK-</strong></td>
<td>PCR performed on CT476834 BAC DNA using primers designed against the 138 repeat block. The PCR product was ligated into the pBluescript SK- (Stratagene) by standard ligation</td>
<td>To identify function of the ACRO138 repeats block in human cells</td>
</tr>
<tr>
<td><strong>pDJ290HS SK-</strong></td>
<td>PCR performed on AC011841 BAC DNA using primers designed against the 138 repeat block. The PCR product was ligated into the pBluescript SK- (Stratagene) by standard ligation</td>
<td>To identify function of the ACRO138 repeats block in human cells</td>
</tr>
</tbody>
</table>
Table 2.8 Cloning Vectors

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBluescript SK- (Stratagene)</td>
<td>LacZ, f1(-), ColE1 origin, Ampicillin⁸</td>
<td>DNA cloning</td>
</tr>
<tr>
<td>pGEM®-T-Easy (Promega)</td>
<td>T7 promoter, SP6 promoter, f1 origin, LacZ, Ampicillin⁸</td>
<td>Cloning of PCR products</td>
</tr>
<tr>
<td>pENTR4 (Invitrogen)</td>
<td>Gateway™ Entry vector containing attL sites, pUC origin, Kanamycin⁸</td>
<td>Site specific recombination with Gateway™ Destination vector</td>
</tr>
<tr>
<td>pDEST17 (6xHIS-tag)</td>
<td>6xHIS-tag arabinose inducible E.coli expression vector</td>
<td>Expression of recombinant protein</td>
</tr>
<tr>
<td>pCEB</td>
<td>β-gal reporter plasmid</td>
<td>Assess of transfection efficiency</td>
</tr>
<tr>
<td>pJRC41 (Chubb et al, 2002)</td>
<td>Contains a blasticidin resistance cassette</td>
<td>Used in co-transfection experiments for the generation of stable cell lines</td>
</tr>
</tbody>
</table>

2.6.5 Affinity Purification of Antibodies

4mg of the relevant antigen was dialysed in 0.2M NaCO₃/0.5M NaCl pH8.3 using a Slide-A-Lyser® G₂ Dialysis Cassette (Thermo Scientific) as per manufacturer’s instructions. After dialysed, the antigen was coupled onto 1ml HiTrap™ NHS-activated HP column (Amersham Biosciences) as per manufacturer’s instructions. The antibody was then bound and eluted using the ImmunoPure® Gentle Binding (Pierce) and Elution Buffers (Pierce). The column was attached to a pump to aid circulation of the liquids and was equilibrated with 2ml of ImmunoPure® Gentle Binding Buffer (Pierce). 4mls of serum was mixed with an equal volume of ImmunoPure® Gentle Binding Buffer (Pierce) was centrifuged at max speed for 30sec and the supernatant was injected in the column and circulated for about 10min. The column was
then washed with 2ml Binding Buffer (Pierce)X3 and the antibody was eluted with the addition of approximately 10ml of Gentle Elution Buffer (Pierce). After testing the concentration by Bradford assay, the peak fractions were pooled together and the membrane with the elution mix was dialysed against 1xPBS O/N at 4°C.
3. Bromodomain adjacent to zinc finger domain protein 2A (BAZ2A)
3.1 Background

The first publication on the bromodomain adjacent to zinc finger domain protein 2A (BAZ2A) was by Jones and his colleagues in 2000 (Jones et al., 2000). They identified a novel bromodomain family containing four novel genes. These novel genes, BAZ1A, BAZ1B, BAZ2A and BAZ2B, seemed to be widely expressed and were similar to Drosophila Acf1 and Caenorhabditis elegans ZK783.4. They were found to localize to human chromosomes 14q12-q13, 7q11-q21, 12q24.3-qter and 2q23-q24, respectively.

The bromodomain is characterized as a structural domain conserved in all mammals. It has been found to be 110 amino acids long with a central motif of 60 amino acids. Interestingly, the 20 amino-terminal residues between bromodomain repeats located within a single protein show less sequence similarity to each other than to similarly located bromodomains in other proteins. This suggests the existence of possible specific functional features that could hypothetically affect their localization to specific nuclear compartments (Jeanmougin et al., 1997). The bromodomain is often found to associate with other common motifs such as the PHD finger that is adjacent to the amino terminus of the bromodomain in the TIF protein family (Venturini et al., 1999) and in Acf1 (Ito et al., 1999).

The bromodomain is found in most proteins with histone acetyltransferase (HAT) activity. Nuclear HAT proteins promote transcription by acetylating residues on the histone tails (Kouzarides, 1999). This way, the association between the histone and the DNA is weakened so that the DNA is free to associate with transcription factors. (Winston and Allis, 1999). The bromodomain has been found to recognize acetylated lysine ($K_{AC}$) residues, as for example the ones on the N-terminal tails of histones, which is often a requirement for protein-histone association and chromatin remodelling (Zeng and Zhou, 2002, Filippakopoulos et al., 2012). This association between the bromodomain and chromatin is further strengthened by the fact that some bromodomain proteins have been found to be subunits of chromatin remodelling complexes. Examples include the Drosophila Acf1 protein complexes with ISWI (imitation SWI) in ACF (ATP-utilizing chromatin
assembly and remodelling factor) (Ito et al., 1999) and the BRG1 (ATP-dependent helicase SMARCA4) and BRM (highly homologous to BRG1 ATP-ase) components of the mammalian SWI/SNF (SWItch/Sucrose Non Fermentable) complexes (Muchardt et al., 1996).

As mentioned in Chapter 1, TIP5 is one of the two subunits comprising the nucleolar remodelling complex (NoRC) and has been reported to co-localize with UBF in the nucleolus of mouse fibroblast NIH3T3 cells (Strohner et al., 2001). NoRC is the proposed silencing mechanism of rDNA genes in mice (Santoro et al., 2002, Nemeth et al., 2004). The proposed steps of NoRC’s mediated repression of Pol I transcription are: 1) NoRC is targeted to the nucleolus by the interaction between its TIP5 subunit with TTF-I that is bound to the rDNA promoter-proximal terminator $T_0$ site (Langst et al., 1997, Strohner et al., 2001), 2) it remodels the nucleosomes at the promoter and represses its function through changes in its chromatin conformation by the recruitment of DNA methyltransferases DNMT1 and DNMT3b, histone deacetylases HDAC1 and HDAC2 for de novo DNA methylation and histone deacetylation at histone H4 respectively (Santoro et al., 2002, Zhou et al., 2002, Strohner et al., 2004) making the promoter inaccessible for PIC formation (Li et al., 2006). It has also been proposed that since NoRC is associated only with the late replicating rDNAs that this repressed chromatin state may be maintained during cell division (Li et al., 2005). Experiments where TIP5 was over-expressed in NIH3T3 cells resulted in repression of rDNA transcription in a dose-dependent manner, suggesting a repressive rather than stimulatory role of NoRc on rDNA transcription (Santoro et al., 2002, Zhou et al., 2002). Additionally, over-expression of TIP5 in NIH3T3 cells was found to reduce the number and size of nucleoli, cell proliferation level was decreased as well as changing some rRNA genes’ replication times from early to late (Li et al., 2005). Experiments where TIP5 was knocked-down in NIH3T3 cells resulted in impaired rDNA silencing enhancing rRNA production and cell proliferation (Santoro et al., 2009). Additionally, it induced genome instability at the rDNA locus as well as the nearby centric (heterochromatin structure of repetitive minor satellite) and pericentric (major satellite) DNA sequences (Guetg et al., 2010). Although it has been proposed that BAZ2A is
the human homologue of TIP5 localized in the nucleolus and performing a similar role in human cells there is currently no direct evidence confirming this.

It is important to mention here that even though TIP5 and BAZ2A have similar sequences, are not identical. BAZ2A contains the ZB2 motif that is absent in TIP5 and TIP5 contains the TAM (TIP5/ARBP/MBD) domain as well as four AT-hooks that are absent in BAZ2A (Jones et al., 2000, Strohner et al., 2001). The TAM and AT-hooks is a motif that assists binding to the minor groove in DNA (Aravind and Landsman, 1998). The AT-hooks are thought to be transcription co-regulators by being able to modify the DNA’s architecture in such a way for it to be accessible to promoters and transcription factors (Strohner et al., 2001, Zillner et al., 2013). However, it is not known whether the above differences (Fig. 3.1) in the two sequences influence different functions on these two genes.

Figure 3.1 Locations of the motif sequences in BAZ2A and TIP5. This is a schematic representation of the motif content in proteins BAZ2A and TIP5. They both contain, at similar locations, the motifs LH (pink), BAZ1 (blue), BAZ2 (purple), WAKZ (yellow), PHD finger (green) and Bromodomain (red). However, they are different in that BAZ2A contains a ZB2 domain (gray) that is absent from TIP5 and TIP5 uniquely contains a TAM domain (orange) and AT-hooks (black). The significance of these differences is not known. Not to scale. (Adopted from Jones et al., 1999 and Grummt and Pikaard, 2003).

The second subunit comprising NoRC is the ATP-ase SNF2h in mammals/ISWI in Drosophila (Strohner et al., 2001). As mentioned in Chapter 1, Drosophila ISWI has the ability to interact with other proteins giving rise to three different chromatin-remodelling complexes, homologues of which have been found in yeast (Tsukiyama et al., 1999), Xenopus (Guschin et al., 2000) and humans (Poot et al., 2000, Bochar et al., 2000). These complexes are: NURF (nucleosome remodelling factor) (Tsukiyama and Wu, 1995), CHRAC (chromatin remodelling and assembly complex) (Varga-Weisz et al., 1997)
and ACF (ATP-utilising chromatin remodeling and assembly Factor) (Ito et al., 1997). Interestingly, experiments on Drosophila have shown that Toutatis (Tou) the ortholog of TIP5 (Fauvarque et al., 2001) along with ISWI but also CtBP (C-terminal binding protein) form the Toutatis-containing chromatin remodelling complex (ToRC). Surprisingly, it has been shown that ToRC has the ability to remodel the chromatin structure and regulate RNA Pol II transcription. Even more surprisingly, it does not localize in the nucleolus, instead it localizes extensively with CtBP in the nucleus. However, using an antibody from the middle of Tou instead of the C terminus, it was found that a splice variant of C-terminally truncated Tou can form an alternative protein complex that does not contain CtBP and is targeted to the nucleolus (Emelyanov et al., 2012). It has been found that for proper localization in vivo CtBP’s association is required as well as to achieve optimal biochemical activity (Guetg et al., 2012). Yet the most surprising findings from the same research came from immunoprecipitation experiments on HeLa cells. Using a BAZ2A antibody it was found that it associates with CtBP, which is recruited to genes that are transcribed by RNA Pol II and has the ability to regulate them. This finding also suggests that there is no nucleolar localisation (Emelyanov et al., 2012).

In conclusion, NoRC might have the potential to silence individual rDNA repeats in mice but remains an open question whether it has the potential to influence the activity status of whole NORs in humans.

The aim of the work described in this chapter is to assess the role of BAZ2A in the silencing of rDNA/NORs in human cells. Specifically, I generated a BAZ2A antibody that can be used in both Immuno-FISH and Western Blot experiments. I determined the localization of BAZ2A in human cells. Finally I performed siRNA mediate depletion of BAZ2A to assess its role in regulation of rDNA and NORs.
3.2 Results

3.2.1 Generate BAZ2A Antibody that can be used in both Immuno-FISH and Western Blot experiments.

As no commercial antibodies were available that were suitable for both cell biology and western blotting I chose to generate my own antibody against the recombinant protein. Since BAZ2A is a large protein, I used a 1 kb fragment (encoding from amino acids 158 to 489) from the N-terminus of the protein. This fragment was isolated from the BAZ2ApENTR2B plasmid that Brian McStay had generated in the past through an Ncol/XbaI enzyme digest. The BAZ2A fragment was ligated into a pENTR4 vector and this new construct was named BAZ2ApENTR4. After verification and sequencing of this Gateway entry vector I next performed an LR reaction to shuttle the BAZ2A insert into the Gateway destination vector pDEST17. This vector was a 6xHis-tagged plasmid, where 6xHis-tag amino acid motif is commonly used to assist high affinity purification of recombinant proteins (Hengen, 1995). Following verification the resulting plasmid was transformed into a BL21 Arabinose Indusible E.Coli cell strain (Life technologies) from which several colonies were chosen in order to see if the protein was expressed. From the protein gel I observed that the N-terminally 6xHis-tagged BAZ2A fusion protein is indeed expressed and soluble and therefore easy to extract (Fig. 3.2 A). However, the protein seemed to be of a higher molecular weight (52 kDa) than expected (42 kDa) even when the 6xHis-tagged plasmid vector’s molecular weight (2.5 kDa) was added in the equation. Therefore, the coding sequences of the BAZ2A plasmid were verified by sequencing. A possible reason for our observation is that this protein is highly acidic with a pKi of 3.9 and runs much slower on the protein gel. The soluble protein was purified using Nickel NTA agarose beads after extraction (as described in Materials and Methods) and was run on a protein gel (Fig.3.2 B). From the results it was shown that I was able to extract and purify the BAZ2A antigen. The smaller fragments in Fig. 3.2 B are presumed to be break down products of the purified protein.
Figure 3.2 Production of recombinant BAZ2A and antibody characterization. All samples were loaded on NuPAGE gels, stained with coomassie blue and de-stained with Destain Buffer. The black arrows indicate the protein sizes in kD on a protein marker. A) Samples were extracted from uninduced cells, induced with L-Arabinose, insoluble and soluble sample taken from the supernatant of the centrifuged insoluble sample. The protein gel shows (indicated with a red arrow) the ability of the new 6xHis-tagged BAZ2A construct to express the BAZ2A protein which is also shown to be soluble. B) Samples were extracted from induced cells were purified using Nickel NTA agarose beads equilibrated with Native Extraction Buffer, the supernatant being the insoluble sample and after the addition of 10 mM Imidizole the eluted sample being the soluble protein product. The protein gel shows (indicated with a red arrow) the purified protein. C) Samples are: BAZ2A purified bound protein product was dialyzed x100 using CB-100 Buffer through a dialysis tubing (Gibco BRL) and purified bound not dialyzed. The protein gel shows (indicated with a red arrow) the BAZ2A dialyzed antigen sent to immunize the sheep. D) Samples are different quantities of α-BAZ2A purified IgG (13μl, 20μl and 7μl) (purification procedure described in Chapter 2) and 3rd bleed crude serum (1μl, 1μl and 0.5μl). The protein gel shows (indicated with a red arrow) the purified α-BAZ2A Ab’s specificity in contrast to the crude serum’s. The quantity of the Ab to be used for best possible results is ~10μl.

Finally the protein was dialyzed (Fig. 3.2 C) before its concentration was measured using a Bradford assay. Approximately 3.2 mls (4mg/ml) of antigen
was sent to Diagnostic Scotland Edinburgh to immunize a sheep to generate the desirable Ab. Briefly, the process is as follows: 1. a pre-immune blood sample is taken from the sheep before it is injected with the antigen, 2. at 4 weeks the sheep is given a second injection of the antigen and on week 5 the first bleed is drawn from the sheep, 3. on week 8 the sheep is injected with the antigen again and on week 9 the second bleed is drawn from the sheep, 4. finally, on week 12 the sheep is injected with the antigen one last time and the third bleed is drawn. Each antigen injection is 0.5 mls, so the sheep was injected with 2 mgs of the antigen overall. The sheep is then ceased. 250 mls of the pre-immune and all the bleed samples were then sent to our lab. I used the third bleed to affinity purify the Ab which was used in my further experiments (Fig. 3.2 D).

3.2.2 α-BAZ2A Ab works both in immunofluorescence and immunoblotting.

Western Blot results with BAZ2A Antibody staining of HeLa cell lysate gave the expected band of 200.9 kDa (Fig. 3.3 top). This proved that the Ab is working and identifying the correct size protein. This purified α-BZ2A Ab was used for immunostaining experiments to identify the BAZ2A protein’s localization in the cells. I immunostained RPE-I, HeLa and HT1080 AMD treated cell interphases with α-BAZ2A and α-Fibrillarin (nucleolar protein) antibodies. BAZ2A staining on the RPE-I cells was performed using the crude serum of the α-BAZ2A Ab. The results showed that the α-BAZ2A Ab works and that the affinity purified version used for the HT1080 and HeLa cells shows higher specificity (Fig. 3.3 bottom). However, in contrast with previous reports from other labs, BAZ2A does not seem to localise in the nucleolus but to the nucleus. The staining was also performed on cells fixed with methanol instead of PFA giving the exact same result as above (data not shown). Notably, Brian McStay obtained an aliquot of the α-TIP5 antibody (Stronher et al., 2001) but using a wide variety of staining procedures consistently observed the BAZ2A in human cells being non-nucleolar.
Figure 3.3 Immunostaining using BAZ2A antibody (purified and crude serum) reveals BAZ2A protein is not nucleolar. Top figure shows HeLa cell lysate sample loaded on a NuPAGE gel, transferred to membrane and probed with the α-BAZ2A antibody. The arrows indicate the protein sizes in kilo Dalton with the help of a protein marker. The immunoblotting revealed the ability of my α-BAZ2A antibody to recognize the expected size BAZ2A protein (indicated with red arrow) in HeLa cells. Bottom figure shows RPE-I, HeLa and HT1080 cells fixed, permeabilized and immunostained as outlined in chapter 2. BAZ2A is visualized by incubation with α-BAZ2A antibodies (purified and crude serum) followed by incubation with rhodamine coupled α-sheep antibody and fibrillarin visualized with an α-fibrillarin antibody followed by incubation with FITC coupled α-mouse antibody. The results revealed the BAZ2A protein localizing in the nucleus of RPE-I (A), HeLa (B) and HT1080 AMD treated cells (C). Furthermore, staining with the purified α-BAZ2A Ab (B and C) is more specific than with the crude serum (A).
3.2.3 Silencing of the BAZ2A protein in cells to test antibody’s efficiency and whether its effect would be the reactivation of silent NORs.

In order to see that the α-BAZ2A Ab actually recognizes to the correct protein, an siRNA approach was used to knock it down in growing HT1080 cells. The Western Blot of the protein extracted from the siRNA transfected HT1080 cells and the negative control showed that BAZ2A was indeed silenced (Fig 3.4 A). The protein gel showed that both samples loaded were of the same concentration (Fig. 3.4 B).

Figure 3.4 RNAi mediated depletion of the BAZ2A protein in HT1080 cells. HT1080 cells were mock transfected or transfected with BAZ2A siRNAs. Transfections were performed in 6-well plates using DharmaFECT1 reagent and a final concentration of 20nM BAZ2A siRNAs. Two rounds of siRNA transfections were performed 24hrs and 72hrs after cells were seeded. Mock treated cells and BAZ2A siRNA transfected cells were harvested 24hrs after second transfection and lysed in Laemmlli sample buffer. Samples were loaded on NuPAGE gel, transferred to membrane and probed with the α-BAZ2A antibody (A). The protein gel (B) reveals equal quantity of protein loaded for BAZ2A depleted and control samples. Immunoblotting with α-BAZ2A antibody revealed that the BAZ2A siRNA transfection efficiently reduced cellular levels of BAZ2A in the HT1080 cells (A). The efficiency of the α-BAZ2A Ab was also validated. The arrows point out the sizes in kilo Dalton.
Since all NORs are active in HT1080 cells (Colleran and McStay unpublished) but HeLa cells have been demonstrated to contain 2-3 silent NORs (McStay and Grummt, 2008), BAZ2A depletion was also performed in HeLa cells. The Western Blot results were comparable to the HT1080 cells showing that BAZ2A was indeed silenced (Fig. 3.5 A). Since there is no staining on the knocked-down sample but there is staining on the negative control, α-BAZ2A Ab’s efficiency is established. The protein gel results revealed that all loaded samples were of similar protein concentration (Fig.3.5 B).

Figure 3.5 RNAi mediated depletion of the BAZ2A protein in HeLa cells. A brief description on the method followed is the same as in Fig.3.9. Samples of HeLa cell lysate, siRNA depleted HeLa cells and mock treated negative control HeLa cells, were loaded on NuPAGE gel, transferred to membrane and probed with the α-BAZ2A antibody (A). The protein gel (B) reveals equal quantity of protein loaded for cell lysate, BAZ2A depleted and control samples. Immunoblotting with α-BAZ2A antibody revealed that the BAZ2A siRNA transfection efficiently reduced cellular levels of BAZ2A in the HeLa cells (A). The arrows point out the sizes in kilo Dalton.

Finally, I stained the same protein membrane with both α-BAZ2A and α-RPA43 Abs. The upper half of the protein membrane with HeLa siRNA transfected and mock transfected cells was stained with the α-BAZ2A Ab (200.9 kDa) and the lower half with the dedicated Pol I α-RPA43 Ab (43kD)
which is found in the cytosol, nucleus and nucleolus. The results showed that BAZ2A was indeed knocked-down but this depletion did not have any observable effect on nucleolar structure since RPA43 (Pol I subunit) seems unaffected (Fig. 3.6). Quantification of the band intensities (normalised against the background noise) showed comparable levels for RPA43 and a 10 fold decrease in the BAZ2A signal for the siRNA treated HeLa cells in comparison to the HeLa mock siRNA.

Figure 3.6 RNAi mediated depletion of the BAZ2A protein in HeLa cells does not have any observable effect on the nucleolar structure. Half of the protein membrane with HeLa mock and siRNA transfected cell samples was stained with the α-BAZ2A Ab (top bands) and the other half with the dedicated Pol I α-RPA43 Ab (lower bands). The results revealed that BAZ2A was indeed knocked-down in the siRNA transfected cells but this depletion did not affect RPA43 protein in the same cells, showing that this depletion did not seem to affect the nucleolar structure.

Immunostaining with the α-BAZ2A and α-Fibrillarin Abs on mock (Fig. 3.7 A) and siRNA transfected HeLa cells, where BAZ2A was silenced, showed a vast reduction of the BAZ2A staining in the BAZ2A depleted cells (Fig 3.7 B). The visualization was performed using the same exposure settings for both mock and siRNA transfected cells. It could be suggested that the staining reduction is comparable to the 10 fold reduction shown on the protein membrane. The results showed that BAZ2A is indeed silenced in the RNAi depleted cells and seems to be localised in the nucleus and not in the nucleolus in the negative control mock transfected cells.
Figure 3.7 Immunostaining validates depletion of BAZ2A in siRNA transfected HeLa cells. The cells were fixed, permeabilized and immunostained as outlined in chapter 2. BAZ2A is visualized by incubation with α-BAZ2A antibody followed by incubation with rhodamine coupled α-sheep antibody and fibrillarin visualized with an α-fibrillarin antibody followed by incubation with FITC coupled α-mouse antibody. The results revealed the BAZ2A protein was indeed depleted in the siRNA transfected HeLa cells (lower panels). The staining intensity was vastly reduced compared to that of the mock transfected cells (upper panels). It was again shown that the BAZ2A is localizing in the nucleus of the negative control mock transfected cells (upper panels). Same exposure settings were used for both mock and siRNA transfected cells.

Since I’ve shown that BAZ2A is knocked-down in HeLa cells, my next approach was to see if there was any reactivation of silent NORs between mock (Fig. 3.8 A) and siRNA transfected HeLa cells. In the immuno-FISH experiments I hybridized with the 11.9 kb (EcoRI rDNA fragment from the IGS) probe and stained with an α-UBF Ab. There was no dramatic reorganization of the rDNA in the nucleolus and silent NORs not associated with the nucleolus were also present showing no reactivation of NORs due to BAZ2A depletion (Fig. 3.8 B).
Figure 3.8 Immuno-FISH showing BAZ2A depletion does not affect Pol I transcription in the nucleolus. The cells were fixed, permeabilized and immunostained and hybridized as outlined in chapter 2. UFB is visualized by incubation with α-UBF antibody followed by incubation with rhodamine coupled α-sheep antibody and 11.9 kb EcoRI rDNA fragment is visualized by nick-translation FITC-labeling. The results revealed that the UBF expression levels were comparable between BAZ2A depleted (B) and mock transfected control HeLa cells (A). It was also clear that inactive NORs (green spots) were not associated with the nucleolus nor UBF as normal.

Immuno-FISH using an α-Nop52 Ab, a nucleolar protein, and 11.9 kb EcoRI rDNA fragment on siRNA and mock transfected HeLa cells, revealed the same results as above. Inactive NORs not associated with the nucleoli, were not reactivated when BAZ2A was depleted. Nop52 was chosen to repeat the above experiment since it is technically a better Ab and easier for immuno-FISH experiments. Therefore, we could provide a field of cells showing the same result (Fig. 3.9).
Figure 3.9 Immuno-FISH showing BAZ2A depletion does not cause inactive NOR re-activation. The HeLa cells were fixed, permeabilized and immunostained as outlined in chapter 2. Nop52 is visualized by incubation with α-Nop52 antibody followed by incubation with rhodamine coupled α-sheep antibody and 11.9 kb EcoRI rDNA fragment is visualized by nick-translation FITC-labeling. The results revealed that the inactive NOR numbers were comparable between BAZ2A depleted (B) and mock transfected HeLa cells (A). It is clear there is no re-activation of inactive NORs upon BAZ2A depletion and that inactive NORs (green spots) are not associated with the nucleolus exactly as in HeLa wild type cells.
3.3 Discussion

In several published research papers in the last decade it has been reported that the nucleolar remodelling complex (NoRC) is a silencing mechanism on individual rDNA repeats. It has also been implied in these same studies but not proven yet that it could potentially influence the activity of whole NORs. All research has been performed on the mouse homologue of the human BAZ2A, TIP5, the one of the two subunits forming NoRC and predominantly on mouse fibroblast NIH3T3 cells (Strohner et al., 2001, Santoro et al., 2002, Zhou et al., 2002, Nemeth et al., 2004, Strohner et al., 2004, Li et al., 2005, Li et al., 2006, Santoro et al., 2009, Guetg et al., 2010, Zllner et al., 2013).

Therefore, I decided to test this hypothesis in order to identify the mechanism that controls the activation and silencing of whole NORs in human cells. Initially, I generated a new BAZ2A antibody since there was no proof that the commercially available ones would work also in immuno-FISH experiments apart from Western Blots. Due to the fact that the BAZ2A protein is large, I decided to use a 1 kb fragment from the N-terminus of the protein. The produced protein was soluble and could be purified on Nickel NTA agarose beads. Approximately 3.2 mls (4mg/ml) of the antigen was sent to Diagnostic Scotland Edinburgh for the BAZ2A antibody to be grown in sheep. The third bleed was used to affinity purify the antibody that was used for the following experiments. A Western Blot was used to stain HeLa cell lysate to confirm that the Ab was working and that it identified the correct size of the BAZ2A protein. The Ab identified the correct protein at the expected approximately 200.9 kDa. I then immunostained RPE-I, HeLa and HT1080 (AMD treated) cell interphases with α-BAZ2A and α-fibrillarin Abs. The results not only showed that the BAZ2A Ab was working also in vivo but opposite to previous reports, BAZ2A is a nuclear protein excluded from the nucleolus. Therefore, the next step was to knock-down BAZ2A in growing cells using the siRNA approach in order to prove the Ab’s efficiency but also to observe any silencing effects. I first silenced BAZ2A in HT1080 cells in which all the NORs are active (Colleran and McStay unpublished). From the Western Blot it was
obvious that BAZ2A was indeed knocked-down. Then, I knocked-down BAZ2A in HeLa cells that are known to contain 2-3 silent NORs (McStay and Grummt, 2008). The Western Blot results were comparable to these of the HT1080 cells. Finally, to show that BAZ2A depletion does not affect the levels or distribution of the Pol I machinery in the cells, I stained the same protein membrane with α-BAZ2A and α-RPA43 (Pol I subunit) Abs. The results showed that BAZ2A was indeed knocked-down but this depletion had no observable effect on the nucleolar structure.

After validating the BAZ2A Ab’s efficiency, the next step was to show that BAZ2A is a nuclear protein that is excluded from the nucleolus. The hypothesis was that if BAZ2A is not found in the nucleolus, where it supposedly binds to hSNF2, it cannot be part of NoRC and it cannot influence the NORs activity. Therefore, BAZ2A localization was studied in HeLa siRNA transfected and negative control mock transfected cell interphases. The interphases were stained using α-BAZ2A and α-fibrillarin (a nucleolar protein) Abs and the results showed that BAZ2A staining was considerably reduced in the siRNA transfected cells compared to the mock transfected cells. Also, from the mock transfected cells it was clear that BAZ2A is not a nucleolar protein as shown from the negative control cell staining. All the results point out that BAZ2A is localized in the nucleus and is not a nucleolar protein. Furthermore, immuno-FISH was used to stain HeLa siRNA and mock transfected cells with α-UBF Ab and hybridize with the 11.9 kb rDNA fragment. Since HeLa cells have 2-3 inactive NORs, I wanted to investigate if there was any change in the levels or distribution of the Pol I machinery in the cells where BAZ2A was depleted. However, the immuno-FISH results showed no dramatic difference in the rDNA levels between the siRNA and the mock transfected cells. It was observed that the inactive NORs were not associated with the nucleolus or UBF even in the siRNA treated cells. Hence, I used immuno-FISH again to stain HeLa siRNA and mock transfected cell interphases with α-NOP52, a technically better Ab than UBF, and hybridize with the 11.9 kb rDNA fragment. Again, the hypothesis was that if BAZ2A is part of a silencing mechanism for Pol I transcription then its depletion would allow the reactivation of silent NORs. The results showed no reactivation of
silent NORs in the siRNA transfected HeLa cells and the inactive NOR numbers were comparable between these and the mock transfected cells. These results contradict previous reports regarding BAZ2A (Strohner et al., 2001).

Taking into consideration all of the above results, it is safe to say that at least in human cells BAZ2A does not seem to be a nucleolar protein nor to influence NOR activity. As mentioned earlier in this chapter, there has been a recent research published in *Drosophila* suggesting that ToRC (NoRC in *Drosophila*) does not localize in the nucleolus but in the nucleus. Importantly, only affects Pol II transcription. Furthermore, immunoprecipitation experiments on HeLa cells using a α-BAZ2A antibody showed that it associates with CtBP (C-terminal-binding protein), which is recruited to genes that are transcribed by RNA Pol II and that has the ability to regulate them. CtBP is the third subunit found to comprise ToRC that in its absence a splice variant of C-terminally truncated Tou can form an alternative protein complex targeted to the nucleolus (Emelyan ov et al., 2012). In all other published studies in the last decade reporting that NoRC is a nucleolar silencer of rDNA repeats complex by repressing Pol I transcription, an association with CtBP has not been neither considered nor found (Strohner et al., 2001, Santoro et al., 2002, Zhou et al., 2002, Nemeth et al., 2004, Strohner et al., 2004, Li et al., 2005, Li et al., 2006, Santoro et al., 2009, Guetg et al., 2010).

CtBP proteins have been shown to form functional interactions with several cellular proteins with diverging roles in intracellular signalling as well as transcriptional control. In the nucleus they act as transcriptional corepressors (Bergman et al., 2006, Chinnadurai, 2007, Chinnadurai, 2009, Turner and Crossley, 2001) and in the cytoplasm are involved in the Golgi and endolytic membrane fission (Bonazzi et al., 2005, Chinnadurai, 2007, Corda et al., 2006). As transcriptional co-repressors their primary function is to act as platform for the recruitment of chromatin-modifying enzymes such as histone deacetylases (HDACs), histone methyltransferases and polycomb group proteins, to DNA-binding transcription factors (Kuppuswamy et al., 2008, Shi et al., 2003). It has also been found that disruption of these genes in
*Drosophila* leads to embryonic defects in segmentation consistent with transcriptional repression (Nibu and Levine, 2001, Poortinga *et al*., 1998) and their depletion in mice leads to early postnatal death (Grootecaes *et al*., 2003, Hildebrand and Soriano, 2002).

Considering all of the above information, BAZ2A seems to play a different role in human cells since it is not nucleolar and presumably is involved in Pol II transcription in *Drosophila* and HeLa cells as mentioned in the Emelyanov research earlier (Emelyanov *et al*., 2012).
4. Mapping of the distal and proximal to the rDNA sequences, termed Distal (DJ) and Proximal (PJ) Junctions.
4.1 Background

The heterochromatic nature of the acrocentric short-arms makes them difficult to study from a genomic perspective; particularly as they are thought to be highly repetitive (Wang et al., 1999). As a consequence, the regions flanking the rDNA have been a low priority for genomic analysis (International human genome sequencing consortium, 2004) and these five chromosome short arms are not annotated in the current human genome assembly. Thus, while the genomics of human rDNA is relatively-well studied, with a complete rDNA repeat sequence, 43 kb of IGS and transcribed region sequences, having been assembled from the sequence of sub-cloned rDNA fragments (Gonzalez and Sylvester, 1995), almost nothing is known of the regions that flank the rDNA. The central role of the nucleolus in growth regulation coupled with the potential for regions adjacent to the rDNA to contribute to NOR function prompted us to investigate the genomic context in which the rDNA resides.

The first reference on the DNA region flanking the rDNA on the telomeric or distal site termed the distal junction (DJ) was by Worton and colleagues in 1988 (Worton et al., 1988). From a translocation in a female patient with Duchenne Muscular Dystrophy (DMD) carrying a translocation between the p-arms of chromosomes X and 21 (t(X;21) (p21;p12). In the process of cloning this translocation junction causing DMD, two clones were found to contain the telomeric end of the p-arm of chromosome 21 along with 3-5 rDNA repeats. These clones were used to determine the orientation of the whole rDNA cluster on the p-arm of the acrocentric chromosome 21. The orientation of the rDNA is in such way that transcription occurs from the telomeric towards the centromeric part of the chromosome. It was also found that the break point between rDNA and non rDNA is at ~3.7 kb upstream of the first repeat unit’s transcription initiation site (Worton et al., 1988). Almost a decade later, in 1997, Gonzalez and Sylvester published a study extending the known DJ sequence to 8.3 kb, annotated now in the GenBank database as U67616, showing it is conserved among the acrocentric chromosomes and also conserved in all primates (Gonzalez and Sylvester, 1997).
Defining the DNA region flanking the rDNA on the centromeric or proximal site termed the proximal junction (PJ) proved to be more problematic and the only publication to date was by Sakai and his colleagues in 1995 (Sakai et al., 1995). They identified 493 bp of PJ sequence on chromosome 22 and found that the PJ break point with rDNA lies within the internal transcribed spacer (ITS) that separates the 18S and 28S coding sequences. They also found by sequence analysis that the ITS is connected to several 147 bp repeat elements that are intervened by a 68 bp unique sequence. Even though this 147 bp tandem repeat was found to be present in all acrocentric short arms, PCR revealed that the precise PJ organisation is variable among the acrocentrics. Finally, it was shown that the rDNA sequence is not interrupted by other sequences and does not recombine with the DJ sequence. The rDNA sequence homogeneity among different NORs seems to be maintained by inter-chromosomal recombination most likely within the rDNA array itself (Sakai et al., 1995).

Probes derived from these presumptive junction sequences were used to screen single chromosome cosmid libraries, and a number of positive clones were sequenced. Searches of GenBank using the resulting sequences, coupled with BAC walking, ultimately resulted in identification of 15 BAC clones from the distal junction (DJ) (Fig. 4.3), five of which include rDNA, and 3 BAC clones from the proximal junction (PJ) (Fig. 4.4), one of which includes rDNA. These cover 379 kb of DJ sequence and 207 kb of PJ sequence that flank the rDNA repeats (Fig. 4.1) (Floutsakou et al., 2013). Preliminary evidence from the McStay lab showed that these putative junction sequences actually adjoin the rDNA. FISH on combed DNA molecules prepared from human nucleoli clearly showed the DJ region adjacent to the rDNA (Fig. 4.2) (Floutsakou et al., 2013). Bioinformatic screening performed by Saumya Agrawal and Austen Ganley (Massey University, New Zealand) also indicated that the DJ adjoins the rDNA (Floutsakou et al., 2013). Thus multiple sources of evidence demonstrate that the DJ immediately flanks the rDNA.
Figure 4.1 Mapping of the human rDNA flanking regions. This is a schematic representation of a human acrocentric chromosome (not to scale) highlighting the telomeric (light blue) and centromeric (pink) regions. The NOR region is expanded below to highlight the rDNA repeat array as well as the 207 kb of the Proximal Junction (PJ; purple) and the 379 kb of the Distal Junction (DJ; orange) flanking regions. The discontinuous lines in the rDNA array represent the approximately 40 repeats inside the array and the arrows represent the direction of the rDNA transcription.

Figure 4.2 DNA combing of the human rDNA and DJ flanking region. At the bottom there are three pictures of DNA combing of human nucleolar DNA hybridization that shows that the DJ (LA14138F-10 red) is physically linked to the rDNA (5.8 green). A schematic representation interpreting the hybridization results is shown at top of the picture (Floutsakou et al., 2013).
Figure 4.3 Positions of BACs and cosmids mapped on the DJ. This is a representation of all the cosmids and BACs (heavy black lines) sequences identified that are mapped onto a representative DJ sequence (green). The rDNA is indicated by dotted grey lines. The positions of the start, finish, and rDNA junction of each clone are indicated in the parentheses in bp. The names of the cosmids and BACs are shown on the top of each black line with the chromosome of origin indicated in parentheses. This representation is drawn to scale that is shown at the bottom (Floutsakou et al., 2013).
**Figure 4.4 Positions of BACs and cosmids mapped on the PJ.** This is a representation of all the cosmids and BACs (heavy black lines) sequences identified that are mapped onto a representative DJ sequence (green). The rDNA is indicated by dotted grey lines. The positions of the start, finish, and rDNA junction of each clone are indicated in the parentheses in bp. The names of the cosmids and BACs are shown on the top of each black line with the chromosome of origin indicated in parentheses. This representation is drawn to scale that is shown at the bottom (Floutsakou et al., 2013).
PCR, on DNA from a panel of mouse somatic cell hybrids, at five intervals across the DJ region and the single unique region in the PJ gave the expected product in all cases, showing that the DJ and PJ are conserved among all acrocentric chromosomes (Fig 4.5). Nevertheless, similar to the DJ metaphase FISH evidence was more difficult to obtain for the PJ region, particularly with hybridization-based approaches. Metaphase FISH showed the PJ to be centromere-proximal to the rDNA, as expected, but additional signals were observed distal to the rDNA (Fig. 4.6) and on other chromosomes (Floutsakou et al., 2013).

**Figure 4.5** DJ and PJ sequences are conserved among the acrocentric chromosomes. PCR performed at increasing distances (left) into the DJ from mouse somatic cell hybrids carrying a single human acrocentric chromosome and the reciprocal products (Xder21 and 21derX) of a chromosome 21 translocation that originates in the rDNA (indicated above). The results confirmed the DJ is located distally to the rDNA. The bottom panel is the same, but the primers used were to the single unique PJ region (Floutsakou et al., 2013).

**Figure 4.6** PJ flanking region hybridization observed on the DJ side of the chromosomes as well. The PJ sequences are conserved among the acrocentric chromosomes as shown in the picture. The human chromosomes were hybridized with the PJ BAC CR381535 (green probe) and acrocentric specific q-arm bacmid probes (red probe) (Floutsakou et al., 2013).
Finally, the known DJ and PJ sequences at the time I started my investigation are resumed in Fig. 4.7. Four BACs were used to build the DJ contig and two BACs for the PJ contig. These BACs are also the ones that were mostly used as probes for the DJ and PJ hybridizations in the experiments described in the following chapters. The most distal DJ BAC’s (AC011841) last ~40 kb were found to be CER satellite. The Centromeric Repeats (CER) satellite (Jurka et al., 2005) was originally identified as a repetitive sequence consisting of 8 copies of a 48 bp repeat mapped at the centromere of the q-arm of chromosome 22 (Metzdorf et al., 1988). This finding will be discussed in detail in the results section of this chapter.

The aims in this chapter are:

1. To confirm that DJ and PJ sequences are conserved among the acrocentrics in position to the rDNA.
2. To explain the hybridization pattern of the PJ BACs on the chromosomal metaphases.
3. To study the distribution of CER satellite. We wanted to see how it was shared among the acrocentrics and if it could be found somewhere else in the genome.
4. Extend the known DJ sequence with a newly identified BAC.
Figure 4.7 BACs used to construct DJ and PJ contigs. A) The DJ contig (green) was constructed by merging four BACs (black lines; clone name and chromosome of origin are indicated) from two different chromosomes and is 379,046 bp in length. B) The PJ contig (orange) was constructed by merging two BACs from chromosome 21 and is 207,338 bp in length. Shading lines represent regions of BAC overlap, % identity and overlap start and stop positions (in parentheses) are indicated. Dotted vertical lines isolate parts of the BACs used in the contig formation with the length of each fragment indicated immediately below. The rDNA is shown in grey and the scale is at the bottom (Floutsakou et al., 2013).
4.2 Results

4.2.1 Confirmation of the DJ and PJ sequences are conserved among the acrocentrics in position to the rDNA.

Firstly, I wanted to confirm that these putative junction sequences actually adjoin the rDNA. My first approach was to confirm the DJ’s position in comparison to the rDNA on acrocentric chromosomes. Therefore, I hybridized metaphase chromosome spreads from HT1080 cells with a DJ and an rDNA probe. The DNA used to make the fluorescently labelled probes were from the DJ associated CT476538 BAC (Fig. 4.7) and the 11.9 kb EcoRI fragment from the intergenic spacer (Fig. 4.10) to define the rDNA position boundaries (Fig.4.8).

![Image](image_url)

**Figure 4.8 Mapping of the DJ flanking region by FISH on metaphase spreads of HT1080 cells.** FISH was performed on HT1080 metaphase spreads with DJ bacmid CT476834 (red) and rDNA 11.9 kb EcoRI intergenic fragment (green) probes. Chromosomes are DAPI-stained (blue). The DJ localizes distal to the rDNA on all the acrocentric chromosomes. The enlarged picture shows an individual acrocentric chromosome showing that the DJ (red) is located on the distal side to the rDNA (green) near the telomere end of the chromosome.

The results showed that the DJ is indeed positioned distal to the rDNA on the short arms of the acrocentric chromosomes and is conserved among all the
acrocentrics. HT1080 cells are not karyotypically normal therefore they may contain more than 10 NORs. Notably, in some chromosomes the two signals seem more separated than in others. A possible interpretation is that these are active NORs appearing as a secondary constriction and where the signals are closer to each other could represent the condensed chromatin of silent NORs. The same method could not be used to confirm the PJ position since as mentioned earlier the PJ hybridizes on both the distal and proximal sides on the acrocentrics. This raised the question of whether this phenomenon occurs due to complex folding of the linear PJ DNA in condensed chromosomes (Fig. 4.9 B) or due to PJ related sequences existing also on the distal side (Fig. 4.9 A).

**Figure 4.9 PJ hybridization found on the DJ side possible explanations.** In panel A, the bottom image is a schematic representation of a human acrocentric chromosome (not to scale) highlighting the telomeric (light blue), centromeric (pink), rDNA (gray circle), PJ (purple) and DJ (orange) regions. The NOR region is expanded above to highlight the rDNA repeat array as well as the Proximal Junction (PJ; purple) and the Distal Junction (DJ; orange) flanking regions. The discontinuous lines in the rDNA array represent the approximately 40 repeats inside the array, the little blue boxes the rDNA repeats and the arrows represent the direction of the rDNA transcription. In the top scheme the PJ sequence (purple) has been repeated on the DJ sequence (orange) to represent the possible existence of PJ-like sequence preceding the DJ. In panel B, the middle scheme represents the rDNA looping out of the chromosomal structure (black discontinuous line) with the PJ sequence (purple) overlapping the DJ sequence (orange) to represent the possibility of the PJ’s complex chromatin folding during interphase. Bottom scheme represents the PJ’s complex folding during metaphase extending through the rDNA to the end of the DJ.
The first approach to solve this puzzle was to use FISH on somatic cell hybrid metaphases. The somatic cell hybrids used in this experiment were two mice monochromosomal cell lines each containing a human chromosome with the DMD translocation (t(X;21) (p21;p12) that were in fact used for the identification of the DMD gene (Worton et al., 1988). The GM09142 mouse somatic cell line contains the derivative human chromosome 21derX. In this case, chromosome 21 has the PJ part on its p-arm but is missing the DJ part and in its place has the translocated p-arm from chromosome X. The GM10063 somatic mouse cell line contains the derivative human chromosome Xder21. In this case, chromosome X has the DJ part from the acrocentric chromosome 21 p-arm translocated on its p-arm. Therefore it is missing the PJ part of the acrocentric chromosome. The DJ associated CT476538 (Chr21) BAC and the PJ CR381535 (Chr21) BAC were used to make fluorescently labelled green and red probes respectively. Also, the 11.9 kb EcoRI intergenic fragment was used as before to make the rDNA probe (Fig. 4.10).

Figure 4.10 Locations of regions on the acrocentric chromosomes used as probes. This is a schematic representation of a human acrocentric chromosome (not to scale) highlighting the telomeric (light blue) and centromeric (pink) regions. The NOR region is expanded above to highlight the rDNA repeat array as well as the Proximal Junction (PJ; purple) and Distal Junction (DJ; orange). The discontinuous lines in the rDNA array represent the approximately 40 repeats inside the array and the arrows represent the direction of the rDNA transcription. The arrows indicate the probes used from the different regions of the chromosome.
Metaphase chromosome spreads were prepared from these somatic cell hybrids and were hybridized with both mixtures of the probes, a DJ green/rDNA red probe and a PJ red/rDNA green probe. From the results it was clear that in the GM09142 cell hybrid, containing the derivative 21 chromosome carrying the PJ, there was only a PJ signal. Interestingly, in the GM10063 cell hybrid, containing the derivative X chromosome carrying the DJ, there were two signals one from the DJ and one from the PJ. This result proves that the reason PJ hybridisation is found on the DJ side of the acrocentric chromosomes is not due to complex folding of the linear DNA in condensed chromosomes but due to PJ related sequences being repeated on the distal side. In contrast, DJ sequences are only found distal to the rDNA on the acrocentric chromosomes and we observe no detectable hybridization on non-acrocentric chromosomes (Fig. 4.11).

**Figure 4.11 FISH analysis of chromosomes in metaphase from the GM somatic cell hybrids.** The hybridization of the 21derX chromosome (containing only the PJ side) shows the presence of the PJ in spectrum red and the rDNA sequence in spectrum green/red. The absence of the DJ is clear (top pictures). The hybridization of the Xder21 chromosome (containing only the DJ side) shows the presence of the PJ in spectrum red, the rDNA sequence in spectrum green/red and the DJ in green (bottom pictures). PJ related sequences seem to be repeated on the DJ side.
To confirm the hybrids, PCR was performed using a primer pair for the DJ side on position 292 (sequences can be found in the Materials and Methods chapter) near the end of the known DJ sequence against the GM09142 and GM10063 somatic cell hybrids. The results showed that the DJ is uniquely only found to the telomeric side of the NORs on the acrocentric chromosomes (Fig. 4.12).

![Figure 4.12 Validation of DJ content of GM monochromosomal cell hybrids.](image)

The DJ292 primer pair was used in a PCR on GM and Placental DNA, to verify the DJ/PJ content of the two GM cell hybrids. The 8% polyacrylamide gel validated that the GM10063 cell line (DJ), contained the DJ which is not present in the GM09142 cell line (PJ).

### 4.2.2 Explanation for the hybridization pattern of the PJ BACs on the chromosomal metaphases.

The next approach in order to map the PJ position on the acrocentric chromosomes was to use a technique called molecular combing (Caburet et al., 2005). This method allows the visualization of multiple aligned single DNA molecules that are attached uniformly and irreversibly to silanated glass coverslips (Genomic Vision). This technology is based on the fact that one or both extremities of the DNA molecules in a solution can be attached spontaneously onto these specially-treated coverslips when the latter are dipped into the solution. This attachment is irreversible, so when the coverslips are withdrawn from the solution at a steady pace using the molecular combing apparatus (Genomic Vision) the DNA fibres are stretched and aligned uniformly in parallel to each other over the coverslip's surface (Fig. 4.13).
Figure 4.13 Molecular DNA combing. This is a schematic representation of the molecular combing technique. The left figure represents the coiled DNA in a buffer solution and the silanized coverslip lifted out of the solution. The DNA molecules, whose one open end has attached to the surface of the coverslip, are now being stretched as single DNA fibres due to the two opposite direction forces upon them (adhesion force from the machine and the meniscus force). The figure on the right represents the final combed coverslip. The fibres here are of the same size. However, if the DNA fibres break from tension the result would be the same but with different fibre sizes. The DNA fibres’ starting points are different because of the different points of their attachment to the coverslip.

During analysis using optical microscopy, the physical length distance measured is the same as the DNA molecule. As mentioned earlier in Chapter 1, rDNA repeats can be found both in a canonical and non-canonical orientation on a single combed DNA molecule (Fig. 4.14). Approximately 1/3 of the human rDNA repeats are palindromic and are proposed to be transcriptionally silent (Caburet et al., 2005).

Figure 4.14 DNA combing of the human rDNA. This is a picture of DNA combing of (human) RPE-I nucleolar DNA hybridization with the 7.1 kb EcoRI fragment from the 28S rDNA (Bio-red) and the 5.8 kb EcoRI fragment from the 18S rDNA (Dig-green). On the top panel is a schematic representation interpreting the hybridization results is shown at top of the picture underneath.
Using a PJ probe the results showed that it was very rare to see rDNA immediately adjacent to the PJ (Fig. 4.15) in contrast to DJ that was always seen physically linked to the rDNA (Fig. 4.2). Therefore molecular combing did not provide a straightforward answer for the complex hybridization pattern observed with PJ probes. However, the results seem to point towards extensive segmental duplication of PJ-like sequences on non-acrocentric chromosomes.

![Figure 4.15 DNA combing of the human rDNA and PJ flanking region.](image)

**Figure 4.15 DNA combing of the human rDNA and PJ flanking region.** At the bottom is a picture of DNA combing of HeLa nucleolar DNA hybridization that shows that the PJ (LA1525H3 Biotin labelled-red) is physically linked to the rDNA (5.8 kb Digoxigenin labelled-green). A schematic representation interpreting the hybridization results is shown at top of the picture.

### 4.2.3 CER satellite.

The last ~40 kb of sequence of the most distal DJ BAC (AC011841) of the DJ contig are CER satellite. I decided to investigate how the CER was shared among the acrocentrics and if it could be found in other loci throughout the genome. My experiments show that CER could be centromeric as originally described but the major signals are found elsewhere. FISH on normal human chromosomal metaphases using a 5’FITC labelled CER probe consisting of a 48 bp consensus sequence oligonucleotide (5’-TTCCAGAACACTGCTRCKRG GGTCTGTTTCTCCTACATAGGA-3’ (R standing for either A or G and K for either G or T) and acrocentric specific probes mapped the CER blocks on the p-arms distal to the rDNA side on all the acrocentric chromosomes, with additional pericentromeric blocks on chr14 and chr22 (Fig.4.16).
Figure 4.16 48bp CER repeats are found distal to the NOR on all five acrocentric chromosomes. Human metaphase chromosomes were probed with acrocentric specific q-arm bacmid probes (red) and a 48 bp CER satellite probe (green). Note that peri-centromeric CER blocks are present on chromosomes 14 and 22.

4.2.4 Extension of the known DJ sequence with a new identified BAC.

BLAST search for new BACs in order to extend the so far known DJ and PJ sequences resulted in the identification of a new BAC (AL591856). This BAC was shown to contain CER sequences on one end, PJ-like sequences on its other end and unique sequences on its middle part (Fig. 4.17). Therefore, my first aim was to try and position this BAC on either the DJ or PJ side on the acrocentric chromosomes. A primer pair at position 111 kb on the BAC was used from a DNA sequence that was highly unique for BAC (AL591856). The PCR results using DNA from the monchromosomal cell lines A9-13, 14, 15, 21, 22 and the GM09142 and GM10063 on the acrylamide gel were weak but looked positive for chromosome 15 and the DJ (GM10063). They were clearly positive for chromosomes 14, 22 and the PJ (GM09142) (Fig. 4.18). This finding addressed both previous FISH findings, of the CER found on both ends of the acrocentric chromosomes 14 and 22 and the PJ found on the DJ side.
Figure 4.17 Genomic organisation of BAC AL591856. The panel is a representation of similar sequences to BAC AL591856 found in the genome from a BLAST search. The red line represents the BAC and length. From the matching similar sequences it was observed that this BAC contains CER satellite sequences on one end and PJ-like sequences on the other end. However, in the middle it seems to contain unique sequences.

Figure 4.18 BAC AL591856 validation of sequence content from PCR on monochromosomal cell hybrids. The 111 primer pair with relatively unique sequence from BAC AL591856 was used in a PCR on DNA from GM and A9 monochromosomal cell lines, to verify its sequence content. The 8% polyacrylamide gel results showed that this new BAC contains sequences similar to chromosomes 14, 22 and the PJ (GM09142). It also seems possible that it contains similar sequences to chromosome 15 and the DJ (GM10063). The 156 bp product size is indicated with an arrow.
The next approach was to use FISH on normal human chromosomal metaphases using as probes the BAC and the 11.9 kb rDNA fragment to identify the position of this BAC on the chromosomes. The results showed that it hybridizes on all the acrocentric chromosomes but also on a set of nonacrocentric long chromosome, possibly chromosome 1. It was also interesting to find that 6 out of 10 signals were on the DJ side of the acrocentric chromosomes and 4 out of these 10 (possibly 14 and 22) had a split signal on both the DJ and PJ sides, exactly like in the CER hybridization (Fig. 4.20). Results from FISH on human normal chromosomal metaphases using as probes the BAC and the acrocentric specific q-arm bacmid probes, verified that the split signal is on chromosomes 14 and 22; and not found on 13, 15 or 21 (Fig.4.21). Therefore, this new BAC is likely to be positioned after the CER satellite sequence on the DJ side and containing PJ-associated sequences (Fig. 4.19). This finding provides an explanation for the PJ hybridization signal observed distal to rDNA on acrocentric short arms.

![Figure 4.19 Schematic representation of the known DJ sequence to date.](image)

This is a schematic representation of the DJ sequence (pink) that has been extended with the new BAC AL591856. The known DJ sequence is physically connected to the rDNA repeats and we know that at its end, the last BAC AC011841 contains CER satellite. AL591856 also contains CER satellite on one end (blue), unique sequences in the middle (orange) and PJ similar sequences at the other end (green). We have identified its position on the DJ and believe that the CER it contains is part of the same CER block on AC011841. Therefore, we have now extended the known DJ sequence. Not to scale.
Figure 4.20 FISH of normal human chromosomes in metaphase with BAC AL591856 and rDNA probes. The hybridization of the chromosomes with the BAC AL591856 probe (green) and the 11.9 kb EcoRI rDNA fragment (red) showed that the BAC is located on the DJ side on all the acrocentric chromosomes and on the pericentromeric area of possibly chromosome 1 (yellow arrowheads). A split signal is observed on possibly chromosomes 14 and 22 (white arrowheads).
Figure 4.21 FISH of normal human chromosomes in metaphase with BAC AL591856 and acrocentric specific probes. The hybridization of the chromosomes with the BAC AL591856 probe (green) and the acrocentric specific q-arm bacmid probes (red) showed that the BAC is located on the DJ side on all the acrocentric chromosomes and on the pericentromeric area of possibly chromosome 1 (yellow arrowheads). The split signal was verified to be only on chromosomes 14 and 22 (white arrowheads).
4.3 Discussion

Before I discuss my results I would like to mention the bioinformatics analysis results performed by our collaborating Ganley lab that gave more insight on the PJ’s presence on the distal site (Floutsakou et al., 2013). The bioinformatic screen indicated that the PJ region has a high level of segmental duplication (see next page). Consequently additional sequence-based evidence was sought for the PJ adjoining the rDNA. Cosmids that had been identified from each acrocentric chromosome to contain both PJ and rDNA were sequenced, showing that the PJ is linked to at least 16 kb of rDNA (Fig. 4.22). The position of the PJ/rDNA junction in these cosmid sequences is slightly different to the BAC junction position; therefore a bioinformatic strategy was implemented to look for PJ-rDNA junctions from whole genome sequence data. Both the BAC and cosmid PJ junction positions were identified, and no evidence for additional PJ junctions was found. It is possible that the PJ is adjacent to a large piece of segmentally duplicated rDNA, particularly as the PJ contig has previously been identified as a chr21 short arm pericentromeric region (Lyle et al., 2007). However, PJ-linked rDNA external transcribed spacer (ETS)/18S sequence is 98.8% identical to that linked to the DJ, whereas 18S segmental duplicates in other parts of the genome have a much lower sequence identity and no segmentally-duplicated ETS sequence was found elsewhere. Together, these results suggest that the PJ also adjoins the rDNA, but the precise junction position varies among the acrocentric chromosomes.
Figure 4.22 Cosmids and BACs spanning the DJ-rDNA and PJ-rDNA junctions. The coloured lines represent the DJ (green) and PJ (orange) regions, the grey lines the rDNA and arrows indicate the junction points. Indicated cosmid names are followed by chromosome of origin in parentheses. The length of the flanking region and rDNA in each cosmid is given in kb with a scale shown at the bottom. The position of the rDNA junction point is indicated in bp and the region in which the junction occurs in parentheses. BACs CT476837 and CR392039 are included for a comparison between cosmid and BAC junction positions. PJ is found linked to at least 16 kb of rDNA but the precise junction position varies among the acrocentric chromosomes (Floutsakou et al., 2013).
As expected, the bioinformatic screening also showed that the DJ adjoins the rDNA. However, even though interchromosomal conservation of the DJ was expected, it was not anticipated for the PJ. To quantify the level of DJ and PJ inter-chromosomal homogenization, bacmid and cosmid sequence identities were determined from intra- and interchromosomal pairwise comparisons. Intrachromosomal sequence identities approach 100% for both DJ and PJ sequences, as expected. However, interchromosomal identity is slightly lower for the DJ, (average 99.1%), while PJ interchromosomal sequence identity is lower still, but nevertheless substantial (average 93.3%) (Fig. 4.23). PJ interchromosomal differences predominantly arise from interchromosomal polymorphisms in the rDNA junction position and in Alu elements. These results show that together with very efficient DJ homogenization, PJ homogenization is also occurring, pointing to unappreciated levels of recombination in the pericentromeric regions of acrocentric chromosome short arms.

![Figure 4.23](image)

**Figure 4.23 Conservation of DJ and PJ among the acrocentric chromosomes.** The figure is a plot of the average intrachromosomal and interchromosomal DJ and PJ sequence identities from pairwise comparisons of representative BAC and cosmid clones. Intrachromosomal sequence identities approach 100% for both DJ and PJ. Interchromosomal identity for DJ is ~99% and for PJ ~93% (Floutsakou et al., 2013).

In order to investigate whether they harbour genomic features of interest, consensus DJ and PJ contig sequences were generated from a minimum set of overlapping BACs. BACs from the same chromosome were used when possible. Repeat analyses revealed that both regions contain proportions of
retrotransposons similar to the whole human genome, although LINEs and LTRs are overrepresented and SINEs are underrepresented in the DJ, hallmarks of inactive regions of the genome (Dunham et al., 2012). Strikingly, a large inverted repeat is present in the centre of the DJ. Each arm of the inverted repeat is ~109 kb, and the two arms share an average sequence identity of 80% (Fig. 4.24). Both regions also contain blocks of satellite repeats, as mentioned earlier, a large (38.6 kb) block of 48 bp CER satellite repeats at the distal end of the DJ. Finally, two blocks of a novel 138 bp tandem repeat that we call ACRO138 are present in the inverted repeat of the DJ.

Segmental duplications (SD; duplications of pieces of DNA to elsewhere in the genome) are a prominent feature of the human genome and are commonly enriched at centromere boundaries (Jurka et al., 2005, Bailey and Eichler, 2006). Given their proximity to centromeres a SD (defined as regions >1 kb with >85% identity) analysis of the DJ and PJ was performed. SDs from both regions are found across the genome, but PJ SDs were more frequent, longer, and had greater sequence identity than DJ SDs. Furthermore, the majority of PJ SDs were found in centromeric/peri-centromeric regions of the genome, as had been previously observed (Lyle et al., 2007), while the majority of DJ segmental duplicates were found in euchromatic/telomeric regions (Fig. 4.25). It was found that while 7.3% of the DJ is segmentally duplicated, 92.4% of the PJ is segmentally duplicated. This suggests that far from being a recombinationally inert region of the genome, the PJ experiences frequent and ongoing recombination with other pericentromeric regions of the genome. This high level of recombination is consistent with recent reports of active mitotic recombination in centromeric regions (Chiolo et al., 2011; McFarlane and Humphrey, 2010), and may be responsible for PJ interchromosomal homogenization, with additional recombination events that initiate in the rDNA potentially explaining the greater level of DJ homogenization. The high level of SD also likely explains our inability to locate more centromere-proximal BAC clones to extend the PJ.
Figure 4.24 DJ inverted repeat arms similarity plot. The location, arrangement and length (in bp) of the large inverted repeat (white arrows) in the DJ contig (green) is shown at the top. The start and end position lengths of each repeat are indicated in bp (parentheses). At the bottom, a VISTA plot shows the level of sequence identity between the two inverted repeat arms to be ~79.5%. The gaps represent indels (Floutsakou et al., 2013).
This SD data comes only from the reference human genome that does not contain the acrocentric short-arms. We now have shown from the FISH data that the acrocentric short arms also contain SD and further evidence for this was given by the identification of the BAC AL591856.

![Image](image.png)

**Figure 4.25 Segmental duplication analyses for PJ and DJ.** The coloured lines represent the segmental duplications from PJ (orange) and DJ (green), indicating the location of the duplicate on the human chromosomes that are numbered and arranged around the flanking regions. Segmental duplicate length is indicated by colour as shown below. The positions of centromeres (yellow) and telomeres (blue) are also indicated. While the DJ seems to be fairly segmentally duplicated, the PJ seems to be highly segmentally duplicated (Floutsakou et al., 2013).

In this chapter I focused on refining the genomic context of the flanking to the rDNA sequences (DJ and PJ). My first approach was to establish the DJ and PJ positions on the acrocentric chromosomes. FISH was used on HT1080 chromosomal metaphases using a DJ/rDNA probe. The results confirmed the DJ position being unique to the distal side of the acrocentric short arms and that it is conserved among all the acrocentrics. To confirm the PJ position to the proximal side of the acrocentrics, FISH was used on chromosomal metaphases from the mouse monochromosomal cell hybrids (GM09142, GM10063). These cell hybrids contain one human acrocentric derivative chromosome from the DMD translocation (t(X;21) (p21;p12). Strikingly, the hybridization with a PJ/rDNA probe showed a PJ signal also on the distal side.
This result was convincing that this is not due to the way the linear DNA on the NOR is folded during metaphase. It is because sequences similar to the PJ sequence exist also on the distal side of the acrocentric chromosomes. In order to establish the PJ position on both the centromeric and telomeric ends of the acrocentric chromosomes, I used a technique called molecular combing. This technique is ideal for visualizing single strands of DNA molecules. Results of the hybridization with a PJ/rDNA probe very rarely showed the PJ immediately adjacent to the rDNA in contrast to what is always seen with the DJ/rDNA hybridization. This difficulty of identifying the PJ on a single DNA molecule level is due to the PJ being highly segmentally duplicated with the AL591856 BAC providing support on this view. This makes PJ hard to use as a probe since it is also found in loci outside of the NORs.

Knowing the existence of the CER satellite repeats at the end of the DJ sequence it was intriguing to look for any PJ-like sequences that were further distal. FISH experiments with a 48 bp repeat CER probe on normal human chromosomal metaphases showed that the CER is interestingly not centromeric as originally described. Even more interesting was the observation of a split CER signal only on chromosomes 14 and 22. The identification of a new BAC that seemed to extend the known rDNA adjacent sequences gave more insight on the previous puzzling observations. The AL591856 BAC contains unique DJ-like sequences in its middle part that are found nowhere else in the genome. Also, on one end it contains CER satellite repeats and on the other end PJ-like sequences. PCR results showed that AL591856 BAC has similar sequences to 14, 15 and 22 acrocentric chromosomes as well as the PJ but seemed to exist on both the DJ and PJ sides. The results from FISH on normal human chromosomal metaphases using an AL591856/rDNA probe showed that this sequence is conserved on all the acrocentrics. It was also found to localize to the pericentromeric area of possibly chromosome 1. Interestingly, 6/10 signals were on the DJ side of the acrocentric chromosomes and 4/10 had a split signal on both the DJ and PJ sides. This correlates with the distribution of the CER hybridization on metaphase FISH. Further experiments showed that indeed the split signal was on the same chromosomes as with the CER hybridization; chromosomes 14
and 22. Therefore, this new BAC was mapped after the CER satellite sequence at the end of the DJ known sequence at the time and seems to contain PJ-associated sequences. The new BAC’s location and genomic content gave a possible explanation for the PJ hybridization on the DJ site observed in previous experiments. What was still puzzling was that some BAC hybridization was also observed on the metacentric chromosome.

After the bioinformatics analysis it was clear that the PJ has a very high content of segmental duplications. This explains the hybridisation on non-acrocentric chromosomes but at the same time makes it very hard to interpret PJ FISH in this study. Finally, such a high degree of segmental duplications makes the PJ less likely to contain any regulatory elements influencing the activity of the NORs. Therefore, I concentrated my continuing research on characterising the DJ which has been proven to contain sequences highly unique to this location on the acrocentric chromosomes. One of the outcomes of this work is that we now have a probe (DJ) that allows us to visualize the positioning of individual NORs in interphase.
5. Positioning of the Distal (DJ) and Proximal (PJ) Junctions.
5.1 Background

The nucleolus is formed around individual active NORs of the acrocentric chromosomes. Active NORs stain weakly with DAPI because of their decondensed state (Heliot et al., 1997). Therefore, while rDNA transcription is taking place inside the nucleolus, the nucleolus is also weakly DAPI stained. As mentioned earlier, the nucleolus is surrounded by a heterochromatic shell separating it from the rest of the nucleus. This heterochromatin is presumably derived from the satellite DNA on the acrocentric chromosomes short arms (Nemeth and Langst, 2011).

The rDNA is transcribed by Pol I. Specific inhibition of pol-I transcription by low concentrations of actinomycin D (AMD) induces a rapid and remarkable reorganization of the nucleolus (Flickinger, 1968, Chen and Jiang, 2004). During this reorganization, two of the regions comprising the nucleolus (FC and DFC) segregate into distinct domains albeit close proximity domains that relocate at the periphery of the nucleolus. This involves the dissociation of the nascent rRNA transcripts and their associated processing machinery that results in the collapse of the rDNA repeats and into the formation of the nucleolar caps at the nucleolar periphery. These caps consist of rDNA, Pol-I transcription machinery and some processing factors. Interestingly a large heterochromatic domain is more clearly visible surrounding the segregated nucleolus (Hadjiolova et al., 1993, Shav-Tal et al., 2005, Hernandez-Verdun, 2006).

Having established the identity of DJ and PJ sequences, I then wanted to see where they are localized relative to rDNA during the formation of the nucleolus. Hence, the main aim of this chapter was to identify the DJ and PJ positions in relation to the nucleolus and whether their location is specific and important for the caps formation. I constructed synthetic arrays from the DJ sequence to test if it contains any intrinsic localization signals at different sites. By transfecting cells with individual or a combination of DJ BACs and cosmids I addressed their location and whether it has any impact on NOR
localization. Furthermore, with the use of low concentrations of AMD the borders of the nucleolus could easily be identified, as well as the DJ and PJ localization in relation to the nucleolar caps and the endogenous and exogenous (constructed arrays) DJ signals in the transfected cells.
5.2 Results

5.2.1 DJ and PJ positioning in relation to the nucleolus and the nucleolar caps.

The unique genomic features of the DJ, discussed in the previous chapter, suggest that it may contain functional elements. That prompted me to analyse the positioning of both DJ and PJ in interphase nuclei to better understand their relationship with the nucleolus. In order to identify the DJ and PJ positions in relation to the nucleolus and the nucleolar caps containing the nascent rDNA, 3D-FISH was used to hybridize HeLa cells in interphase. The CT476834 (DJ) and the CR381535 (PJ) BACs were used as probes (Fig. 4.7 and 4.10). The PJ seemed to also localize in other locations in the nucleus apart from the nucleolus (Fig. 5.1). As explained earlier, this is due to the high level of segmental duplications within its sequence. Hence, this makes it difficult to determine which FISH signals derive from the PJ versus from unlinked segmentally duplicated regions. Thus, I decided to focus my research on the DJ.

From 3D-FISH results using the DJ and the 11.9 kb EcoRI rDNA fragment, it seemed that the DJ signal was mostly associated with the perinucleolar heterochromatin but sometimes it was also found in the nucleolus (Fig. 5.2). H2B GFP-tagged HeLa cell interphases, where the nuclear heterochromatin is pre-stained green, were also hybridized using the DJ probe. Exactly as before, the DJ was found to localize in the perinucleolar heterochromatin (Fig 5.3).
Figure 5.1 PJ localization is more complex than DJ. 3D-FISH was performed on HeLa cells that were fixed, permeabilized and hybridized as outlined in chapter 2. The intensity of the signals from the two probes was manipulated to be comparable on extended focus. The PJ (CR381535-red) hybridisation pattern is more complex than that obtained with the DJ (CT476834-green) probe. Note the presence of isolated PJ signals coming from segmentally duplicated PJ sequences. The nucleus is DAPI stained.
Figure 5.2 DJ localizes to the nucleolar periphery. 3D-FISH was performed on HeLa cells that were fixed, permeabilized and hybridized as outlined in chapter 2. The intensity of the signals from the two probes is comparable on extended focus. The rDNA (11.9 kb-red) resides within the nucleolus and the DJ (CT47634-green) localizes in the perinucleolar membrane. The single focal plains (red arrows) highlight the DJ positions on the perinucleolar heterochromatin. On the left of each Z-stack is a vertical slice and on the top a picture of a horizontal slice; the DJ signal being the central point of slicing. The yellow arrows point out the position of a small DJ in the central picture looking like it is localized inside the nucleolus but in the single plain underneath it can be seen that is localized in the perinucleolus. The white arrows point out at the inactive NORs. The DJ and rDNA signals are not associated with the nucleolus and seem more condensed. The nucleus is DAPI stained.
Figure 5.3 Evidence of DJ localization at the nucleolar periphery. 3D-FISH was performed on H2B GFP-tagged HeLa cells that were fixed, permeabilized and hybridized as outlined in chapter 2. The DJ (CT47634-red) is seen to localize in the perinucleolar membrane. The nucleus is GFP-tagged (green).

Figure 5.4 DJ perinucleolar localization in HT1080 cells. 3D-FISH was performed on HT1080 cells that were fixed, permeabilized and hybridized as outlined in chapter 2. The extended focus images (left) and the single focal plain image on the right show that the DJ sequences lie in the perinucleolar heterochromatin of HT1080 cells. The nucleoli are visualized with α-UBF Ab (red) and the DJ with a CT476834 (green) probe. The nucleus is DAPI-stained (Floutsakou et al., 2013).
This DJ localization was also observed before in HT1080 cells using the DJ probe and α-UBF Ab staining. α-UBF Ab and rDNA probes have been shown to overlap (Floutsakou et al., 2013). It seems that no matter the visualization procedure the result is the same, in that the DJ seems to localize in the perinucleolar heterochromatin and to be associated with the nucleolus (Fig 5.4). This finding encouraged the investigation of these cells using AMD. The HT1080 cells were treated with low concentrations of AMD and with the use of immuno-FISH were hybridized in interphase using the DJ probe and stained with the α-UBF Ab as above.

Figure 5.5 DJ seems to act as a perinucleolar anchor for the rDNA repeats in HT1080 cells. 3D-FISH was performed on HT1080 AMD treated cells that were fixed, permeabilized and hybridized as outlined in chapter 2. In the top panel, two representative cells are shown, one with an enlargement on the right. Inhibition of rDNA transcription with AMD results in the formation of nucleolar CAPs juxtaposed with DJ sequences in perinucleolar heterochromatin. The nucleoli are visualized with α-UBF Ab (red) and the DJ with a CT476834 (green) probe. The nucleus is DAPI-stained. In the bottom panel, two cartoon models are shown of the transition between active and withdrawn rDNA upon AMD treatment. The rDNA (red) retreats from the nucleolus (black) to the DJ (green) that is embedded in perinucleolar heterochromatin (dark blue). Note the DJ (green) signals that are not associated with the nucleolus and reside in the nucleus. These are inactive NORs (Floutsakou et al., 2013).
Again, the results showed that indeed the DJ of active NORs is always localizing to the perinucleolar heterochromatin. The inactive NORs were observed not to associate with the perinucleolar heterochromatin nor UBF (Fig. 5.5). It was also revealed that nucleolar caps form immediately adjacent to the DJ sequences positioned in the perinucleolar heterochromatin. Although some larger nucleolar caps are bi-lobed and associated with two DJ signals, it appears that the majority of caps are derived from individual NORs. The same outcome was observed in HeLa AMD treated cells (Fig. 5.6). These results suggest that the DJ anchors the linked rDNA to perinucleolar heterochromatin, and that retreat of the rDNA to the DJ upon AMD treatment provides an explanation for the positioning of the nucleolar caps. Therefore the rDNA flanking regions may help mediate nucleolar organization and NOR regulation.

Figure 5.6 DJ seems to act as a perinucleolar anchor for the rDNA repeats in HeLa cells. The HeLa AMD treated cells were fixed, permeabilized and hybridized as outlined in chapter 2. The extended focus image (left) and the single focal plains on the right (red arrows) show that the rDNA (11.9 kb-red) upon AMD treatment seems to localize to positions immediately adjacent to the DJ (CT476834 -green). The DJs seem to anchor the rDNA as explained in Fig. 5.5. The nucleus is DAPI-stained.
Finally, 3D-FISH was performed on HT1080 (Fig. 5.7) and HeLa (data not shown) in normal and AMD treated cell interphases using the new AL591856 BAC and the DJ BAC CT476834 as probes. The rationale behind this experiment was to establish that the AL591856 BAC’s sequence is closely linked to the known DJ sequence. From the results it looks like both BAC’s hybridization spots are adjacently positioned which is consistent with our view that they are only separated by a block of CER satellite.

**Figure 5.7** The AL591856 BAC is positioned immediately adjacent to the DJ. 3D-FISH was performed on HT1080 normal and AMD treated cells that were fixed, permeabilized and hybridized as outlined in chapter 2. The top panel shows on the left the hybridization of normal HT1080 cells in interphase with AL591856 BAC (green) and CT476834 BAC (DJ-green). The nucleus is stained with DAPI. The AL591856 BAC is localized immediately adjacent to the DJ signals as shown on the right pictures without DAPI staining. The bottom panel shows the same hybridization in HT1080 AMD treated cells to highlight the nucleolar borders. Again, the two signals localize next to each other and on the perinucleolar heterochromatin, showing that the AL591856 BAC is positioned on the DJ side of the acrocentric chromosomes.
5.2.2 Ectopic DJ BAC arrays

Having established the DJ sequences are localized within the perinucleolar heterochromatin I sought to address the possibility that sequences within it are directly responsible for this localization. To investigate this I transfected HT1080 cells using a standard calcium phosphate method with 7μg each from the three BACs AL592188, CT476834 and AC011841 mixed. As these BACs overlap each other it was expected that homologous recombination during the transfection procedure would generate large intact DJ arrays. Note that AL592188 contains 1.5 rDNA repeats and AC011841 contains ~40 kb CER satellite at the end of the DJ (Fig. 5.8). Mapping of these BACs is described in Fig. 4.3. Also, 100ng of the pJRC41 plasmid vector containing a blasticidin resistance gene were used to aid the selection of the positive clones.

![Diagram](image)

Figure 5.8 Positions of BACs relative to the DJ contig used for construction of exogenous DJ BAC array. The three BACs, AL592188, CT476834 and AC011841, encompassing the DJ contig were transfected into HT1080 cells as described in chapter 2. The positions of the BACs on the DJ are outlined in the above schematic representation of the DJ sequence. The chromosomes that these BACs have been identified from are indicated (parentheses). The centre, direction and length of the inverted repeat are indicated (arrows). AL592188 contains 1.5 rDNA repeats (green line) and AC011841 contains ~40 kb CER satellite (red line).

3D-FISH performed on interphases of the positively selected blasticidin resistant clones was the first screening procedure in order to identify clones that had clearly distinguishable exogenous DJ signals i.e. many copies of the ectopic DJ BAC array. The next step in clone selection involved screening of the uniformity of the clones for the insertion of the BAC array. Finally, two
clones were selected for further study, that were exerting clearly distinguishable exogenous DJ signals in fields of cells, which were named DJBAC 2 and DJBAC 3 (data not shown). The reason for this selection is that the largest exogenous signals not only can easily be distinguished from the endogenous DJ signals but it also gives more confidence that the studied localization is of the exogenous DJ signals. The probes used were the CT476834 (DJ) and the 11.9 kb EcoRI human IGS rDNA fragment. In order to identify the chromosomal location of the insertions of the BAC array into these clones, I used FISH on cells in metaphase. The probes used were the DJ BAC CT476834 and the 11.9 kb EcoRI rDNA fragment. Both clones seemed to have very large insertions of the BAC array integrated into non-acrocentric chromosomes (Fig. 5.9).

![Image](image-url)

**Figure 5.9** DJ arrays have integrated into non-NOR bearing metacentric chromosomes. Metaphase chromosome spreads prepared as outlined in section 2 from clones 2 and 3. FISH using rDNA (green) and DJ BAC CT476834 (red) probes is demonstrating that in both cases the DJ arrays have integrated into non-NOR bearing metacentric chromosomes. Note the presence of small amounts of rDNA at these ectopic arrays (Floutsakou *et al.*, 2013).

To verify the BAC array content in each of the two clones, Real Time PCR was used on DNA from each clone, to assess the fold increase of each BAC present. The primer pairs used to measure the fold increase of the DJ
sequences present in both clones were chosen from positions spanning the entire DJ sequence (40, 93, 153, 196 and 340). Primer pairs 40 and 93, representing the length in bp of their positions on the DJ, are located on BAC AL592188. Primer pairs 153 and 196 are located on BAC CT476834 and primer pair 340 is located on BAC AC011841. The results showed that the clones had a high level of insertions of each BAC; clone DJBAC3 being the larger with a 15-20 overall fold increase, followed by clone DJBAC2 with a 5-10 overall fold increase (Table 5.1). These figures were normalised against an HT1080 wild type control DNA where the 10 endogenous NORs had been subtracted from the final values.

<table>
<thead>
<tr>
<th>Location of primer pairs</th>
<th>Clone 2</th>
<th>Clone 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.6 kb</td>
<td>5.0</td>
<td>3.5</td>
</tr>
<tr>
<td>93.9 kb</td>
<td>10.0</td>
<td>21.5</td>
</tr>
<tr>
<td>153.5 and 275.9 kb</td>
<td>5.0</td>
<td>13.5</td>
</tr>
<tr>
<td>196.4 and 229.2 kb</td>
<td>7.6</td>
<td>25.2</td>
</tr>
<tr>
<td>340.2 kb</td>
<td>8.7</td>
<td>12.6</td>
</tr>
</tbody>
</table>

Table 5.1. Fold increase of DJ sequences present in BAC array clones determined by qPCR (Floutsakou et al., 2013).

Quantitative PCR primer pairs positioned across the DJ contig reveal that the DJ content of the ectopic arrays reflects that of the input BAC mixture, suggesting that intact DJ arrays had been constructed. These BAC array insertions contain BAC AL592188 that is known to contain 1.5 rDNA repeats. However, it is not known exactly how many copies of this BAC were inserted into each insertion. In order to prove that the rDNA repeats associated with the integrated BAC arrays on non-acrocentric chromosomes are inactive, I used silver staining on cells in metaphase from clone 3 that contains the largest insertion. The results showed that it does not contain any active rDNA repeats (Fig. 5.10).
I analyzed the localization of these two clones (DJBAC2 and DJBAC3) using 3D-FISH on interphases of AMD treated cells for higher resolution (Fig. 5.11 A). The probes used were the CT476834 (DJ) and the 11.9 kb EcoRI human rDNA fragment. I observed that the exogenous DJ signals were mostly associated with the perinucleolar heterochromatin (Fig. 5.11 B). This means that non-acrocentric chromosomes started to behave as acrocentric chromosomes, as in associating with the nucleolus.
Figure 5.11 Ectopic DJ arrays target peri-nucleolar heterochromatin. 3D-FISH was performed on AMD treated cells from clones 2 and 3 that were fixed, permeabilized and hybridized as outlined in chapter 2. The probes used were the 11.9 kb EcoRI fragment (rDNA -red) and CT476834 BAC (DJ-green). The large green hybridization signals identified by arrowheads indicate the ectopic DJ array (A). Endogenous DJ signals are also visible. Positioning of ectopic DJ arrays was classified as nucleolar associated, partially associated or non-associated and are indicated by white, yellow and orange arrowheads, respectively, and are quantified below (B) (Floutsakou et al., 2013).

The next step was to narrow down the DNA sequence within the DJ responsible for this localization. Therefore, I used the AL592188, CT476834, AC011841 BACs and the LA14 138-F10 cosmid individually as well as mixed (AL592188 with CT476834 and CT476834 with AC011841) (Fig. 5.12). Mapping of the BACs and cosmid used can be found in Fig. 4.3.
Figure 5.12 Positions of BACs relative to the DJ contig used for construction of individual exogenous DJ BAC arrays. The three BACs, AL592188, CT476834, AC011841 and cosmid LA14138F-10 were used individually or in couples to transfect HT1080 cells as described in chapter 2. The positions of the BACs on the DJ are outlined in the above schematic representation of the DJ sequence. The chromosomes that these BACs have been identified from are indicated (parentheses). The centre, direction and length of the inverted repeat are indicated (arrows). AL592188 contains 1.5 rDNA repeats (green line) and AC011841 contains ~40 kb CER satellite DNA at its end. Therefore, I wanted to see whether the rDNA or the satellite DNA could be responsible for the specific DJ localization. Again, I used 3D-FISH on the clone cells in interphases to select the ones whose cells uniformly contained a large exogenous DJ signal which would be easy to distinguish from the endogenous DJ signals (Fig. 5.13 and Table 5.2). The probes used were the individual inserts (for BAC AL592188 clones the LA14 138-F10 BAC probe was used to avoid probing the rDNA twice) and the 11.9 kb EcoRI human rDNA fragment. The clones chosen for further study were: LA14 138-F10-5 and -9, AL592188-5 and -9, AC011841-2, -7, -8 and -9, AL592188/CT476834-1. No inserted BACs were identified on CT476834 and CT476834/AC011841 clones. It is important to mention that the HT1080 cells transfected with the CT476834 BAC started dying soon after transfection, as if overexpression of this sequence was toxic to the cells. The transfection of the cells with this BAC was repeated three times, until finally only one viable clone could be used for further study. From the cells in metaphase of this clone there was no evidence of an exogenous insertion.
Clone | Location of integrant in cells in interphase
--- | ---
LA14 138-F10-5 | mostly in or close to the nucleolar heterochromatin
LA14 138-F10-9 | mostly in the nucleolar heterochromatin
AL592188-5 and -9 | mostly in the nucleolar heterochromatin
AL592188/ CT476834-1 | mostly in or near the perinuclear heterochromatin
AC011841-2 | mostly in or close to the nuclear heterochromatin
AC011841-7, -8 and -9 | mostly in the nuclear heterochromatin

Table 5.2. Location of the BAC insertions of the clones from a random selection of 20 cells in interphase shown in Fig. 5.13:

**Figure 5.13** Clones of individual or coupled BAC insertions show variable positioning of the exogenous DJs. 3D-FISH was performed on AMD treated cells from clones LA14 138-F10-5 and -9, AL592188-5 and -9, AC011841-2, -7, -8 and -9, AL592188/CT476834-1 that were fixed, permeabilized and hybridized as outlined in chapter 2. The probes used were the 11.9 kb EcoRI fragment (rDNA -red) and the respective DJ BACs for each clone insertion (DJ-green). The large green hybridization signals identified by arrowheads indicate the ectopic DJ array. Endogenous DJ signals are also visible with the exception of clones LA14 138F-10-5 and -9 due to the size and intensity of the exogenous signal. Positioning of the ectopic arrays was classified as nucleolar associated, partially associated or nuclear associated. (Figures continue to next page)
The results show that the specific DJ position on the perinucleolar heterochromatin: a) is not due to the satellite DNA at the end of the AC011841 BAC since it looks like the DJ array is locating to a similar heterochromatic sequence in the nuclear heterochromatin consisting of satellite DNA, b) is not due to the rDNA since the LA14 138-F10 cosmid does not contain any rDNA repeats but localizes to the nucleolus and c) in combination with the results from the previous transfection with a mixture of BACs, it looks like it is a combination of sequences within the DJ that are important for the DJ localization. Any clones containing the CT476834 sequence were not viable apart from one clone, AL592188/CT476834-1, that its percentage composition of the individual of the two BACs making the array is not known. It is important to point out that the CT476834 BAC contains only a part of one of the inverted repeats with the other part within the AL592188 BAC. CT476834 also contains a part of the adjoined inverted repeat with the rest of the sequence located within the AC011841 BAC. After the uniform clones with the most distinguishable exogenous DJ signals were chosen, I identified the locations of the BACs insertions using FISH on cells in metaphase. The probes used were from DNA from the individual BACs and from the 11.9 kb EcoRI rDNA fragment. From the results I observed that all BACs had inserted into non-acrocentric chromosomes apart from clone LA14 138-F10-9 (Fig. 5.14 and Table 5.3).

<table>
<thead>
<tr>
<th>Clone</th>
<th>Location of integration site</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA14 138-F10-5</td>
<td>non-acrocentric chromosome (9 or 10)</td>
</tr>
<tr>
<td>LA14 138-F10-9</td>
<td>long arm of acrocentric chromosome</td>
</tr>
<tr>
<td>AL592188-5</td>
<td>long arm of non-acrocentric chromosome</td>
</tr>
<tr>
<td>AL592188-9</td>
<td>centromere of non-acrocentric chromosome</td>
</tr>
<tr>
<td>AL592188/CT476834-1</td>
<td>non-acrocentric chromosome (1) near the telomere</td>
</tr>
<tr>
<td>AC011841-2</td>
<td>long arm of non-acrocentric chromosome</td>
</tr>
<tr>
<td>AC011841-7, -8 and -9</td>
<td>telomere of long arm of non-acrocentric chromosome</td>
</tr>
</tbody>
</table>

Table 5.3. Location of the BAC integration sites as determined by FISH on metaphase shown in Fig. 5.14:
Figure 5.14 DJ arrays have integrated into non-NOR bearing metacentric chromosomes apart from clone LA14 138F-10-9. Metaphase chromosome spreads prepared as outlined in section 2 from clones LA14 138-F10-5 and -9, AL592188-5 and -9, AC011841-2, -7, -8 and -9, AL592188/CT476834-1. FISH was performed using 11.9 kb rDNA fragment (red) and the respective DJ BACs for each clone insertion (green) apart from clone AL592188 that LA14 138F-10 cosmid was used to avoid double hybridizing the rDNA. The results demonstrate that in all cases, apart from LA14 138F-10-9, the DJ arrays have integrated into non-NOR bearing metacentric chromosomes (white arrows). Clone LA14 138F-10-9 seems to have an insertion in the long arm of an acrocentric chromosome.

The most interesting clone was LA14 138-F10-5 since the cosmid was inserted into a non-acrocentric chromosome and seemed to be driving the chromosome to associate with the nucleolus during interphase. Furthermore, this clone is not supposed to contain any rDNA or satellite DNA. In order to prove that this clone did not contain any active rDNA repeats, I used silver staining on chromosomal metaphases. A clone with no observable insertion of the same cosmid was used as a control (Fig. 5.15).
Figure 5.15 Clone LA14 138F-10-5 contains unidentified duplicated active NOR entities. Silver-staining of chromosome spreads, prepared and silver-stained as outlined in section 2, from clone LA14 138F-10 (top panel) reveals that there is no rDNA present within its ectopic DJ array. However, active NOR entities (red arrows) were observed that did not seem to be chromosomes. Clone LA14 138F-10-6 that seemed to have no observable insertion was used as a negative control (bottom panel) and as expected there is no silver staining or NOR entities.

The results showed that there was no active rDNA. However, I noticed some duplicated active NOR entities that were not present in clone LA14 138-F10-6 that did not seem to carry an insertion. Maybe the integration on one site is having consequences on the rest of the NORs. Hybridization of cells in metaphase from the LA14 138-F10-5 clone with an anti-centromeric probe
resulted in no observation of centromere staining on these small entities (Fig. 5.16). Hence, these new NOR entities are not new chromosomes or translocated chromosomes but they look more like double minutes.

Figure 5.16 The unidentified entities of clone LA14 138F-10-5 do not have a centromere. Silver-staining of chromosome spreads, prepared and silver stained as outlined in section 2, from clone LA14 138F-10, were also stained with an α-centromere Ab. The results revealed that these active NOR entities do not have a centromere and therefore they are not chromosomes.
5.3 Discussion

After mapping the DJ and PJ sequences on the acrocentric chromosomes, I then wanted to identify their positions in the nucleus of the cells during interphase. PJ seemed to localise not only in/next to the nucleolus but to other positions within the nucleus. As mentioned earlier, the high level of PJ segmental duplications makes it difficult to determine which FISH signals derive from the PJ versus from unlinked segmentally duplicated regions. Therefore, it seems unlikely that the PJ sequences could play a significant role on the positioning of the acrocentic chromosomes to the nucleolus and I decided to focus my research on the DJ.

On the other hand, the DJ seemed to associate only in/next to the nucleolus. Interestingly, it was observed that the rDNA arrays regressed to the perinucleolar heterochromatin in AMD treated cells, rather than the DJ sequences moving towards the rDNA foci within the nucleolar interior. This finding suggests that there might be elements within the DJ array that could be responsible for this perinucleolar localization, rather than the DJ being simply linked to the rDNA. To investigate this I transfected HT1080 cells with a mixture of three BACs that cover the entire DJ contig. Two stable clones containing integrated arrays of this BAC mixture were selected for further analysis. FISH with DJ BAC CT476834 and rDNA probes revealed that ectopic DJ BAC arrays had integrated into metacentric (non-NOR bearing) chromosomes. 3D-FISH was performed on cells from these clones to determine whether the ectopic DJ arrays associate with nucleoli. In order to more clearly reveal the boundaries of the nucleolus, cells were treated with AMD. For two clones (DJBAC2 and DJBAC3) I observed a remarkable degree of association of the ectopic DJ arrays with the perinucleolar heterochromatin. This was very easy to assess since these two clones were also the ones with the largest DJ BAC array insertions and therefore showing more distinguishable exogenous DJ signals. Quantitative PCR using primer pairs positioned across the DJ contig revealed that the sequence content of the ectopic arrays reflects that of the input BAC mixture. Moreover, in the majority
of the cells the ectopic DJ array appears to spread through the perinucleolar heterochromatin, covering a significant fraction of the nucleolar surface. Although the arrays contain small amounts of rDNA derived from BAC AL592188, rDNA transcription does not appear to explain the localization, as silver-staining (which in our hands can detect activity at endogenous NORs with comparable rDNA content) shows no activity. Therefore, I conclude that sequences within the DJ contig specify association with perinucleolar heterochromatin even when positioned on metacentric chromosomes.

I then investigated the possibility that sequences affecting the positioning of the DJ may lay within only one BAC. For example, the AL592188 BAC contains 1.5 rDNA repeats at its start and the AC011841 BAC contains ~40 kb of satellite DNA at its end. A logical question that rises from these facts is whether the rDNA or the satellite DNA could be responsible for the specific DJ localization. Therefore, I decided to introduce individual BACs and combinations of two in human cells. 3D-FISH results showed that BAC CT476834 integration seems to be toxic to the cells since I could only acquire one clone for the combination AL592188/CT476834 and on the third transfection attempt with the individual CT476834 only one clone that seemed to not contain any inserts in the end. That toxicity can maybe be explained by the fact that CT476834 BAC contains most part of one of the inverted repeats with the rest of it within AL592188 BAC and some part of the adjoined repeat with the rest within the AC011841 BAC.

However, from a random selection of 20 cells in interphase from each clone it was found that clones containing the AC011841 BAC were more prone to localise to the nuclear heterochromatin and clones containing cosmid LA14 138-F10 and BAC AL592188 to the nucleolar periphery. However, clone AL592188/CT476834-1 was found to localise mostly in the nuclear heterochromatin. These results point out that the specific DJ position on the perinucleolar heterochromatin: a) is not due to the satellite DNA at the end of the AC011841 BAC since it looks like the DJ array is localizing to a similar heterochromatic sequence in the nuclear heterochromatin consisting of...
satellite DNA and b) is not due to the rDNA since the LA14 138-F10 cosmid does not contain any rDNA repeats.

FISH on chromosomes in metaphase from these clones showed that all the insertions were in non-acrocentric chromosomes apart from clone LA14 138-F10-9 that was on the long arm of an acrocentric chromosome. The most interesting clone proved to be clone LA14 138-F10-5 since the cosmid had integrated into a non-acrocentric chromosome, did not contain any active rDNA repeats but was driving the chromosome to associate with the nucleolus during interphase. However, I noticed some duplicated active NOR entities that were not present in other clones. It could be suggested that the integration on one site is having consequences on the rest of the NORs. Anti-centromeric probe hybridization showed that these NOR entities are not new chromosomes or translocated chromosomes since I identified no centromere. These entities actually look like double minutes. Double minutes, are small fragments of extrachromosomal DNA found in human and murine cancer cell lines. They contain no centromere yet they do replicate during mitosis and persist in tumour cell populations if they are small in size (Kaufman et al., 1979, Masters et al., 1982).

The exogenous DJ signal in clones LA14 138-F10-5, -9 and AL592188-5 and -9 was shown to be mostly localising to the perinucleolar heterochromatin. However, this localization was established at ~60% of the time in 20 randomly chosen cells. It is not such a high percentage to reliably conclude that it is due to a specific DNA sequence within the LA14 138-F10 cosmid or AL592188 BAC. Furthermore, AL592188 BAC contains 1.5 rDNA repeats which makes it more difficult to prove that the rDNA does not have an active role in the localization procedure. Finally, the integration with cosmid LA14 138-F10 resulted also in the formation of double minutes showing that the cell line became unstable and therefore unreliable. Hence, I concluded that none of the clones with individual or coupled cosmid and BAC integrants gave any clear localization message and it seemed more likely that it is a combination of sequences within the DJ that are important for the DJ localization. I decided to switch my search to chromatin profiling of the DJ contig. I hoped that
chromatin profiling would help me identify any sequences within the DJ that might influence the specific localisation of the DJ and might have an underlying functionality such as influencing the activity of whole NORs but that will be discussed in the next chapter.
6.1 Background

The ENCylopedia Of DNA Elements (ENCODE) project is a public research consortium whose goal is to identify the functional elements in the human genome. All data generated are released into public databases. One of the most significant findings of this project was that 80% of the genome contains elements linked to biochemical functions. The main aim of the ENCODE project was to determine the elements that regulate the 1% of the genome that encodes the 20,000 genes of the human genome. It was recently reported that in the space between genes lie enhancers (regulatory DNA elements), promoters (initiation of transcription), and numerous regions that encode RNA transcripts that are not translated into proteins and may have regulatory functions. Notably, it has now been found that many DNA sequences that were previously correlated with certain diseases actually lie in or very close to non-coding functional DNA elements (The ENCODE Project Consortium, 2012).

Results from ultra-deep sequencing of RNAs prepared from numerous cell lines and specific compartments within the cell have shown that approximately 75% of the genome is transcribed at some point in some cells. Additionally, it was found that genes are connected with overlapping transcripts that are synthesized from both strands (Djebali et al., 2012). DNase I hypersensitive assays detect open chromatin regions, characteristic of enhancers, that are frequently located distant from the promoters they regulate. Results from DNase I based assays have identified cell-specific patterns of DNase I hypersensitive sites that are in agreement with both computationally predicted and experimentally determined binding sites of transcription factors. However, only 1/5th of the enhancers found were paired to their probable target genes (Thurman et al., 2012, Neph et al., 2012). Therefore, it became clearer that a more complex network exists above the linear genes and transcripts on the chromosomes. This network consists of loops and twists that allow promoters and more distant elements to come to close proximity and transfer their regulatory information. More than 1,000 long-range signals have been
mapped in each cell type by using chromosome-conformation-capture methods (Sanyal et al., 2012). The model suggested by the ENCODE data is that DNA binding of a few high-affinity transcription factors displaces the nucleosomes creating DNA I hypersensitive sites for more transcription factors of lower-affinity to bind. Thus, transcription from non-coding DNA could act as a pool for the creation of functional molecules i.e. regulatory RNAs. Furthermore, it is possible that the binding of transcription factors can block DNA methylation (silencing gene expression) (Sproul et al., 2011). Notably, the ENCODE results have highlighted a significant cell-type specificity of the identified regulatory elements. This is an important finding since regulatory elements may be linked to traits or diseases that have previously been linked to DNA variants. The function of these non-coding areas is yet to be identified; however, it is very possible that since many are cell type-specific any functional changes could be of evolutionary advantage in only certain cell types (Carroll, 2008) or even take part in polygenic adaptation (Prabhakar et al., 2008). Importantly all ENCODE data sets are publically available and were mapped onto the current or previous versions of the reference human genome which do not include acrocentric short arms. This provided us with opportunities to study the chromatin and transcriptional landscape of the DJ.

It is tempting to suggest that anchoring of rDNA to the perinucleolar heterochromatin by the DJ, as discussed in chapter 4, may require DJ-specific functional elements. Hence, the aims of this chapter were to map the ChIP-seq, RNA-seq, DNase-seq, FAIRE-seq, mRNA and EST datasets onto our constructed custom genome that contained rDNA and the DJ (Floutsakou et al., 2013) and experimentally validate the computational predictions of the identified positive peaks. These bioinformatic analyses were performed by Thong Nguyen and Cathal Seoighe (National University of Ireland, Galway) and all validation experiments were performed by myself.
6.2 Results

6.2.1 Chromatin profiling of the DJ using nucleolar ChIP and nucleolar FAIRE.

Chromatin profiling can aid the identification of regulatory elements in previously uncharacterized regions of the genome. Hence, using the publicly available ENCODE histone modification and insulator binding protein CTCF ChIP-seq datasets (Ernst et al. 2011) we mapped them to the DJ (Floutsakou et al., 2013). ChIP-seq data for ten chromatin marks (CTCF, H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K9me3, H3K27me3, H3K9ac, H3K27ac and H4K20me1) (Fig. 6.1 A) and Input were obtained for seven different cell types (GM12878, H1-hESC, HMEC, HSMM, K562, NHEK and NHLF) from the ENCODE Broad Histone (Fig. 6.1 B). DNase-seq and FAIRE-seq data were obtained from the ENCODE UNC/Duke. These data were mapped to the human genome to which the DJ sequences had been added using the ChromHMM (chromatin Hidden Markov Model) software (Ernst and Kellis 2012). The results revealed a complex chromatin landscape that is largely conserved among these cell types (Floutsakou et al., 2013).

Binding sites for CTCF, which can promote chromatin looping (Phillips and Corces, 2009), are observed across the DJ and CTCF consensus sequences are found at each of these sites. Chromatin signatures characteristic of promoters (e.g. H3K4me3) were found across the DJ at regular ~45 kb intervals which strikingly is also the length of a human rDNA unit (Gonzalez and Sylvester, 1995). These chromatin marks were found to coincide with DNase I hypersensitive sites and FAIRE signals (Fig. 6.2). These open chromatin peaks were found to be centred at 138kb and 290kb and correspond to the ACRO138 repeat blocks identified in the repeat analysis. The same bioinformatics analysis was not performed for the PJ due its high segmentally duplicated nature as explained in chapter 4.
Figure 6.1 Chromatin landscape of DJ. (A) ChIP-seq signals of different chromatin features (right) showing the existence of regulatory elements of the DJ in a single cell line are shown below a schematic of the DJ, including inverted repeats. Asterisks indicate enrichment sites. Control signal is shown in gray (bottom). (B) Chromatin states derived from the HMM (Hidden Markov Model) software package analysis (normalized to tags per million mapped reads) for seven different cell types with each coloured bar representing a specific chromatin state (Floutsakou et al., 2013).
Figure 6.2 H3K4me3 peaks overlap with open chromatin domains identified by DNasel-His-seq and FAIRE-seq. Top track shows the enrichment signal of H3K4me3 that was normalized to tags per million mapped reads (also shown in Fig. 6.1 A). The two bottom tracks show enrichment signals of two open chromatin marks, DNasel-HS (DNase I hypersensitive sites) and FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements), calculated using the F-Seq software (Floutsakou et al., 2013).
I experimentally confirmed, as shown below, that the H3K4me3 (indicative of promoters) histone modification mark and FAIRE (indicative of open chromatin) signal peaks are present in the HT1080 cell line at the DJ positions suggested by the bioinformatics analysis. To confirm the existence of the H3K4me3 enriched peaks I performed nucelolar ChIP-PCR (O'Sullivan et al., 2002) at positions 4.4, 138 and 292 kb. Nucleolar chromatin immunoprecipitated with an H3K4me3 Ab was used as a template for PCR using primer pairs across the DJ array. The primer pairs against positive peaks were at positions 4.4, 138 and 292 kb. The same primers were also used on genomic DNA and input chromatin as positive controls. Rabbit IgG Immunoprecipitated DNA was used as a control for the Dynabeads. The H3K4me3 hypersensitive site peaks on the DJ were confirmed to coincide with the ENCODE data (Fig. 6.3).

I next, developed a nucleolar FAIRE protocol that was performed on HT1080 nucleolar chromatin as described in the Materials and Methods to confirm the positions of the open chromatin peaks on the DJ array identified in silica. PCR on the nucleolar FAIRE chromatin was performed using primer pairs for positive peaks of open chromatin 4.4, 138 and 292 kb and the primer pairs for negative peaks 106, 196(/229) and 204 kb on the DJ array. The samples tested were genomic DNA, input chromatin, nucleolar FAIRE chromatin and 10.1 TE as negative control. The results confirmed the mapping of the ENCODE FAIRE data sets showing positive peaks of open chromatin on the DJ array at positions 4.4, 138 and 292 kb (Fig. 6.4).
Figure 6.3 Nucleolar H3K4me3 ChIP-PCR validates the presence of H3K4me3 in the DJ. The nucleolar ChIP was performed as described in chapter 2. The samples used for this experiment were genomic DNA, input chromatin, H3K4me3 ChIP and IgG control ChIP. TE was used as negative control. On the left there is a chromatin map of the DJ with the peaks suggesting H3K4me3 presence. The red lines indicate negative peaks and the green lines positive peaks. The positions of the positive peaks on the DJ are also indicated next to the PCR products.
Figure 6.4 Nucleolar FAIRE-PCR validates the presence of open chromatin in the DJ. The nucleolar FAIRE was performed as described in chapter 2. The samples used for this experiment were genomic DNA, input chromatin and FAIRE chromatin. TE was used as negative control. On the left there is a chromatin map of the DJ with the peaks suggesting open chromatin presence (red) and FAIRE presence (blue). The peaks from the two maps coincide. Red dotted lines indicate negative peaks and green positive peaks. The positions of the positive peaks on the DJ are also indicated next to the PCR products.
Figure 6.5 Nucleolar Pol II ChIP-PCR validates the presence of Pol II in the DJ. The nucleolar ChIP was performed as described in chapter 2. The samples used for this experiment were genomic DNA, input chromatin, Pol I ChIP, Pol II ChIP and IgG control ChIP. TE was used as negative control. A) On the left there is a chromatin map of the DJ with the peaks suggesting Pol II presence. The red lines indicate negative peaks and the green lines positive peaks. B) EST and 18S r DNA primers were also used against Pol I presence as a positive control for Pol I. However, Pol II seemed to give a positive chromatin mark as well. The positions of the positive peaks on the DJ are also indicated next to the PCR products.
ChIP was also performed on HT1080 nucleolar chromatin using a Pol II Ab to confirm the Pol II suggested peaks on the DJ array. Pol I (α-PAF49) Ab was used as a positive control for the rDNA repeats and a negative control mouse control Ab, PCR was performed using primer pairs across the DJ array against the Pol II chromatin marks on the DJ for positive positions 4.4, 138 and 292 and negative positions 106 and 204 kb as well as the Pol I positive ETS, 18S and 28S. The samples tested were genomic DNA, Pol I Ab, Pol II Ab and mouse control Ab immunoprecipitated chromatin, input control chromatin and 10.1 TE as negative control. The results confirmed the positive peaks of the Pol II transcription mark on the DJ array at positions 4.4, 138 and 292 kb to coincide with the ENCODE data. Interestingly, Pol II gave a positive chromatin mark for the ETS and 18S rDNA (Fig. 6.5).

The ACRO138 repeat blocks that were found to coincide with open chromatin domains in the bioinformatics analysis was the next step in this investigation. I wanted to see if these putative promoters are functional domains. Therefore, I decided to follow the approach I had used for the BAC arrays in chapter 5. These sequences were amplified by PCR from DJ BACs DNA and transfected using a standard calcium phosphate protocol into HT1080 cells to see how they would behave. The cells were transfected with 20μg of DNA from the sequences DJ93HS, DJ138HS and DJ290HS. The DNA used was from the amplified PCR product of 765 bp for DJ93HS and 2.9 kb for DJ138HS from the CT476834 BAC DNA using primers DJ93HSF/R and DJ138HSF/R and of 3.12 kb for DJ290HS from the AC011841 BAC DNA using primer pair DJ290HSF/R. 100ng of the blasticidin resistant pJRC41 plasmid were also added to aid clone selection. The three clones acquired (DJ 93, 138, 290-1, 290-2) were screened using FISH on chromosomes in metaphases using probes derived from the same DJ sequences and the 11.9 kb human rDNA EcoRI fragment. No exogenous hybridization signals were observed, therefore there was no evidence of the exogenous sequence inserted. 3D-DNA FISH on cells in interphases, using the same probes as before, did not show any exogenous signal, distinguishable enough from the endogenous signals. Hence, having no distinguishable arrays made it difficult to interpret the results.
6.2.2 Transcription profiling.

The bioinformatics analysis we performed also showed that, interestingly, the CTCF binding sites were positioned close to the DJ/rDNA boundary and frame DJ transcription units. Moreover, chromatin signatures that are known to be associated with actively transcribed gene bodies (e.g. H3K36me3) were observed extending leftward and rightward from positions 187kb and 238kb respectively (Floutsakou et al., 2013). The chromatin profiling results suggest that despite being embedded in the perinucleolar heterochromatin, DJ sequences seemed to be transcriptionally active. Therefore, to directly investigate the possible transcription in the DJ, ENCODE RNA-seq data, RNA pol II and TAF11 chip-seq data and mRNA sequences and ESTs were collected. PolyA tailed RNA-seq data were obtained for 11 different cell types (GM12878, H1-hESC, HCT-116, HeLa-S3, HepG2, HSMM, HUVEC, K562, MCF-7, NHEK, NHLF) from the ENCODE Caltech RNA-seq. GenBank mRNAs and ESTs data were downloaded from the UCSC genome browser. All these collected data were mapped to the human genome to which the DJ sequences had been added using ChromHMM software (Ernst and Kellis, 2012). As mentioned above, actively transcribed gene bodies were observed extending from positions 187 kb and 238 kb during chromatin profiling of the DJ supporting their transcriptional activity. Therefore, we looked at these two sequences in more detail. Analysis of RNA-seq data by Cufflinks (Trapnell et al., 2010) suggested that these putative transcripts produced from these putative promoters were spliced. Moreover, cDNA clones of these putative transcripts were found present in GenBank (accession numbers AK026938 and BX647680) further strengthening the possibility of them being spliced (Fig 6.6).

These putative transcripts were termed disnor187, disnor 238 and disnor138 (from the ACRO138 repeats within the open chromatin peak at 138 kb and their relative expression levels were estimated from RNA-Seq data to be low to medium. They were found to be located within the DJ inverted repeat arms and together with another putative promoter in the ACRO138 repeat block centred at 297 kb, form a symmetrical arrangement of transcriptional units that
mirror the inverted repeat structure. This symmetry is highlighted by the first exons of disnor187 and disnor238 sharing high sequence identity but this symmetry’s functional significance is not known.

To experimentally confirm the existence of these putative spliced polyadenylated transcripts I performed RT-PCR. RT-PCR was performed on cDNAs synthesized from total RNA extracted from the HT1080 cell line and nucleolar RNA extracted from the HeLa cell line. The cDNA synthesis reactions were primed using Random hexamers and Oligo-dT primers. For the RT-PCR of disnor238, the primer pair used was DJ238.5f/242.6r, where the forward primer was positioned at the beginning of exon 1 and the reverse within exon 3 verifying the transcript is spliced. Transcript disnor238’s existence was validated by observing a 241 and a 251 bp (a splice variant in exon 1 that is just 10 nucleotides longer) products from the cDNAs with reverse transcriptase and no product from the samples without transcriptase. The transcript was also found to be polyadenylated in the HeLa nucleolar and HT1080 total RNA cell lines since the positive product cDNAs were made using Oligo-dT primers. For the RT-PCR of disnor187 transcript, the primer pair used was DJ150f/155.7r including the beginning of exon 3 and the end of exon 4 proving it is spliced. The cDNAs from HT1080 total RNA and nucleolar HeLa RNA confirmed the transcript’s existence by a PCR product of 236 bp for the samples where reverse transcriptase was added. It was also shown to be polyadenylated since the positive product was from cDNAs made with Oligo-dT primers. For the RT-PCR of disnor138 transcript, the primer pair used was DJ137.4f/137.6r. The cDNAs from HT1080 total RNA and nucleolar HeLa RNA confirmed the transcript’s existence by a PCR product of 200 bp for the samples where reverse transcriptase was added. As with the other transcripts, it was also shown to be polyadenylated since the positive product was from cDNAs made with Oligo-dT primers (Fig. 6.7). All the transcript PCR products were cloned into the pGEM®-T Easy vector (Promega) and sequenced. All sequencing perfectly matched the predicted spliced transcript.

Finally, to prove the purity of my nucleolar RNAs I also performed RT-PCR using GAPDH primers on cDNAs from total HeLa RNA and nucleolar HeLa
RNA using random hexamers and oligo-dTs. As expected the only positive PCR product came from the total RNA cDNA with reverse transcriptase. Since GAPDH mRNA localizes to the cytoplasm it was expected not to be found in the nucleolar sample with reverse transcriptase (Fig. 6.8).
Figure 6.6 Transcription profiling of the DJ. The top four tracks show ChIP-seq signals of four chromatin marks (TAF1, RNA Pol II, H3K4me3, and H3K36me3); and the bottom tracks show the structures of all DJ transcripts assembled from RNA-seq, mRNA, and EST data. Transcripts originating from promoters at 138 kb, 187 kb and 238 kb in the DJ (boxed in green) were validated (Floutsakou et al., 2013).
ChIP-seq data reveal chromatin features consistent with transcription originating from putative promoters at 187 kb and 238 kb in the DJ. Top four tracks are a blow-up of selected chromatin features from Fig. 6.6. Bottom two tracks show RNA-seq and cDNA mapping results. Exons are indicated by blocks. These identify spliced transcripts (disnor187 and disnor238) similar to cDNA clones AK026938 and BX647690 (top panel). RT-PCR results using primers to detect disnor138, disnor187 and disnor238 transcripts in HT1080 cells. Random and oligoT-primed RT-PCR products of the expected sizes for spliced transcripts were produced validating they are spliced and polyadenylated (bottom panel) (Floutsakou et al., 2013).
Figure 6.8 Nucleolar localization of the disnor transcripts. PCR performed on cDNAs from HeLa nucleolar RNA with random primers and oligo dT show the presence of spliced and polyadenylated transcripts in the nucleoli (left). PCR performed with GAPDH primers on cDNA from total cellular and nucleolar RNA confirms the purity of nucleolar RNA (right) (Floutsakou et al., 2013).

6.2.3 3D-RNA FISH.

After validating the existence of these transcripts the next step was to validate their nucleolar localization as it was suggested from the RT-PCR experiments. 3D-RNA FISH was performed on cells in interphase and decided to focus on transcript disnor238’s localization since it is the largest transcript with more unique sequence and with the least splice variants. The first 1.8 kb of exon 3 from transcript disnor238 was used as a probe due to its unique sequence (Fig. 6.9) (DJ 242.5F/244.3R).

Figure 6.9 Disnor238 transcript probe location. The disnor238 probe was prepared from the amplified PCR product of the first 1.8 kb of disnor238 transcript’s exon 3. It is indicated on the top panel of figure 6.7 by a red arrow.
3D-RNA FISH was performed on cells in interphase from HT1080 and RPE-I cell lines. The probes used were from the disnor238 for the transcript and human 5' ETS to highlight the nucleolus. The results showed that in both cell lines the transcript was associated with the nucleolus. To more clearly visualize the nucleoli and distinguish the nucleus from the perinucleolar heterochromatin, I performed RNA FISH on AMD-treated cells in interphases from HT1080 and RPE-I cell lines using the same probes as before. Pre-rRNA is observed in the nucleolar interior, as expected. Note that Pol II is not inhibited at this low AMD concentration. Importantly, disnor238 RNA foci are observed embedded in perinucleolar chromatin, similar to that observed with DJ DNA FISH (Fig. 6.10).

![Image of RNA FISH results](image.png)

**Figure 6.10 DJ transcripts accumulate in perinucleolar foci.** 3D-RNA FISH was performed on HT1080 and RPE-I cells that were fixed, permeabilized and hybridized as outlined in chapter 2. Results reveal that disnor238 transcripts (disnor238-red) accumulate in distinct foci that surround the nucleolus, indicated by a pre-rRNA probe (5'ETS-green). Perinucleolar accumulation was observed in both normal and AMD-treated cells.
In order to confirm that the RNA results are correct and it is not the DNA probe being observed as the transcript, 3D-DNA FISH and 3D-RNA FISH were performed using the same probes on the same cell lines. The probes used were disnor238 (whole exon 3- 242.5F/244.3R 1.8 kb product and 244.3F/248.6R 4.3 kb product) and the 11.9 kb human rDNA EcoRI fragment from the IGS that is not transcribed and would not hybridize under RNA FISH conditions. The cell interphases used were from HT1080 cell line AMD treated and untreated. The 3D-RNA FISH showed no 11.9 hybridization. The 3D-DNA FISH results showed 11.9 kb rDNA hybridization and in fact the spots were more and brighter than from the disnor238 transcript. This experiment confirmed that the RNA FISH reliably detects the DJ transcript (Fig. 6.11).

![Figure 6.11 Validation of RNA FISH detection method](image)

**Figure 6.11 Validation of RNA FISH detection method.** 3D-DNA (left) and RNA FISH (right) were performed on HT1080 and RPE-I cells that were fixed, permeabilized and hybridized as outlined in chapter 2. RNA FISH results are a reliable detection method for transcripts since they detect disnor238 transcripts (disnor238-red) but there is no hybridization signals coming from the 11.9 kb human EcoRI fragment from the IGS (green) that is not transcribed.
6.2.4 Over-expression studies.

3D-RNA FISH was performed on DJBAC clone 2 AMD treated cells. The probes used were from the first 1.8 kb of exon 3 (DJ 242.5F/244.3R) for transcript disnor238 and human 5’ ETS to highlight the nucleolus. The results showed that the transcript was associated with the nucleolus and it was more intense than in the control cell lines. The pre-rRNA was observed in the nucleolar interior and the disnor238 RNA foci were observed embedded in perinucleolar chromatin as in the wild type HT1080 cells and similar to that observed with DJ DNA FISH (Fig. 6.12).

Figure 6.12 DJ transcripts exhibit stronger transcript signals in DJBAC clone 2. 3D-RNA FISH was performed on AMD treated DJBAC clone 2 cells that were fixed, permeabilized and hybridized as outlined in chapter 2. Results reveal that disnor238 transcripts (disnor238-red) accumulate in distinct foci that surround the nucleolus, indicated by a pre-rRNA probe (5’ETS-green) as in the wild type cells. Interestingly, the signal coming from the transcript was more intense than in the control cell lines.
In order to confirm that the RNA results are correct and it is not the DNA probe observed as the transcript, I used 3D-DNA FISH and 3D-RNA FISH using the same probes on the DJBAC clone 2 AMD treated cells. The probes used were disnor238 (whole exon 3- 242.5F/244.3R 1.8 kb product and 244.3F/248.6R 4.3 kb product) and the 11.9 kb rDNA EcoRI fragment that would not hybridize under RNA FISH conditions. The 3D-RNA FISH showed no 11.9 hybridization and the 3D-DNA FISH results showed 11.9 kb hybridization as normal. In fact the spots were more and brighter than the disnor238 transcript’s exactly like in the HT1080 wild type cell hybridization. These experiments confirmed that the RNA FISH results reliably detect the DJ transcript. The 11.9 kb sequences are not transcribed under RNA FISH conditions, as explained earlier, therefore it can only be visualized under DNA FISH conditions (Fig. 6.13).

Figure 6.13 Validation of RNA FISH detection method in DJBAC clone 2. 3D-DNA (left) and RNA FISH (right) were performed on AMD treated DJBAC clone 2 cells that were fixed, permeabilized and hybridized as outlined in chapter 2. RNA FISH results are a reliable detection method for the overexpressed transcripts since they detect disnor238 transcripts (disnor238-red) but there is no hybridization signals coming from the 11.9 kb human EcoRI fragment from the IGS (green) that is not transcribed.

I then decided to transfect HT1080 cells with a BAC (CU633906 from 131306 to 297192 of DJ contig, 165887 in length) that encompasses both the inverted
repeat and all three transcripts. Additionally, it contains no rDNA or satellite DNA (Fig. 6.14). This way I could also confirm that the transcript detected by RNA-FISH derived from this specific DJ sequence. I screened the clones using 3D-DNA FISH on cells in interphase and chose the clones with a clearly distinguishable exogenous signal for further analysis. The probes used were this BAC DNA and 11.9 kb rDNA EcoRI fragment. The clone that was chosen to be investigated further was clone NDJ 3. FISH was used on chromosomal metaphases from the cells clone exhibiting a distinguishable exogenous DJ signal, using the same probes as above. Clone NDJ 3 looked to have integrated within the NOR of an acrocentric chromosome which was identified to be chromosome 22. 3D-RNA FISH with the disnor238 and the human 5’ETS probe showed the disNOR transcript’s expression to have increased and still located on the perinucleolar heterochromatin but also now smaller domains of hybridization were visible located all over the nucleus (Fig. 6.15).

Figure 6.14 Schematic representation and position of the CU633906 BAC. This is a schematic representation of the DJ array including length in kb. The CU633906 BAC that was used for the overexpression experiments is outlined in blue. The centre, direction and the length of the inverted repeat are also indicated (red dotted arrows).
Figure 6.15 NDJ3 exhibits increased expression of disnor238 transcript. 3D-DNA (left) and 3D-RNA FISH (right) were performed on AMD NDJ3 clone cells that were fixed, permeabilized and hybridized as outlined in chapter 2. NDJ3 cells (bottom panel) containing an array of DJ BAC CU633906 on chr22 short arm (top panel) show increased levels of perinucleolar DJ localization (left-green). The rDNA probe was the 11.9 kb EcoRI fragment (red). Furthermore, 3D-RNA FISH showed disnor238 transcript (right-red) exhibiting an intense RNA hybridization signal localized to the nucleolar periphery, consistent with the results from the wild-type cells. The nucleolus is visualized with the pre-rRNA 5'ETS probe (right-green).

3D-RNA immuno-FISH on AMD treated cells from the same transfectant using the disnor238 (the first 1.8 kb of exon 3) probe and NOP52 Ab revealed an intense RNA hybridization signal localized to the nucleolar periphery, consistent with the results from the wild-type cells. Again, in order to be sure that the RNA results are correct and it is not the DNA probe observed as the transcript, I used 3D-DNA FISH and 3D-RNA FISH using the same probes on interphases of AMD treated cells from clone HT1080 NDJ 3. The probes used were disnor238 (the first 1.8 kb of exon 3) probe and the 11.9 kb rDNA EcoRI fragment probe that would not hybridize with under RNA FISH conditions. The
3D-RNA FISH showed no 11.9 hybridization. The 3D-DNA FISH results showed 11.9 hybridization as normal (Fig. 6.16).

Figure 6.16 Overexpressed transcript disnor238 localizes to the nucleolar periphery in clone NDJ3. 3D-DNA (top panel), 3D-RNA FISH (middle panel) and 3D-immuno FISH (bottom panel) were performed on AMD NDJ3 clone cells that were fixed, permeabilized and hybridized as outlined in chapter 2. AMD-treated cells from clone NDJ3 containing the DJ BAC CU633906 array on chr22 short arm show increased levels of the ectopic DJ array (top panel-green) in 3D-FISH. 3D-RNA FISH (middle and bottom panels) using probes against the non-transcribed IGS of human rDNA 11.9 kb EcoRI fragment (green), disnor238 (red), and the nucleolus via anti Nop52 antibodies (green) show an intense transcript signal and validate its nucleolar localization.
6.3 Discussion

From the bioinformatics analysis of the DJ array’s chromatin profiling of the ENCODE data (ChIP-seq, DNase-seq and FAIRE-seq data) in several human cell lines it was found that the DJ contains regulatory elements (i.e. H3K4me3) and a very complex chromatin landscape. In Fig. 6.1 A there are peaks indicated with asterisks within the DJ at the same position (93 kb) for both H3K27ac and H3K27me3. It seems contradictory since H3K27ac is a mark of active promoters and H3K27me3 a mark of inactive genes. It is possible that this reflects the presence of both active and silent NORs within the cell line.

I experimentally confirmed, using nucleolar ChIP Immunoprecipitation and nucleolar FAIRE on nucleolar chromatin, the existence of H3K4me3 (indicative of promoters) Pol II (indicative of transcription) and FAIRE (indicative of open chromatin), positive peaks of open chromatin on the DJ array and the hypersensitive sites at 4.4, 138 kb and 292 kb in the HT1080 cell line that I also used later in my immuno-FISH experiments. The transfection with the DJ hypersensitive sites of the ACRO138 repeat blocks into different cell lines and the identification of any specific functions gave inconclusive results. Moreover, these results did not associate individual sequences with specific function and it is very likely that they are not operating in isolation but more likely in combination with each other or another DNA sequence or mechanism. Another hypothesis is that, as the chromatin profiling suggests, they are related to transcription.

The chromatin profiling results suggested that despite being embedded in perinucleolar heterochromatin, the DJ sequences are transcriptionally active. Further evidence was acquired from RNA-seq data, RNA pol II and TAFii1 chip-seq data, and mRNA sequences and ESTs mapped onto the DJ. The putative promoters at 187 kb and 238 kb, in particular, gave multiple lines of evidence supporting their transcriptional activity. Analysis of RNA-seq data indicated that the putative transcripts (disnor187 and disnor 238) produced from these putative promoters were spliced, with further evidence
strengthening this view being the presence of cDNA clones of these putative transcripts in GenBank. These two putative transcripts along with the ACRO138 repeats, at 138 kb collectively termed disnor138, all lie within the DJ inverted repeat arms. Together with another putative promoter in the ACRO138 repeat block centred at 297 kb, were found to form a symmetrical arrangement of transcriptional units that mirror the inverted repeat structure within the DJ array. Interestingly, this symmetry is also highlighted by the fact that the first exons of disnor187 and disnor238 share a high sequence identity. However, the functional significance of this symmetry is yet not known. The length of the transcripts suggests they may function as long non-coding RNAs (lncRNAs). Furthermore, lncRNAs are often found to function at their site of synthesis (Lee, 2012). I confirmed by RT-PCR all three putative transcripts’ existence at the hypersensitive sites and also that they are spliced. Furthermore, I also found that they are polyadenylated nascent transcripts and that they seem to localise to the nucleolus which is also their sight of synthesis. The purity of the nucleolar fraction was validated by performing RT-PCR using GAPDH primers on cDNAs from total HeLa RNA and nucleolar HeLa RNA using random hexamers and oligo-dTs as for the transcripts. GAPDH mRNA localizes to the cytoplasm and as expected did not show a positive PCR product for the nucleolar sample with reverse transcriptase.

In order to confirm the transcript’s nucleolar localization as it was suggested from the RT-PCR experiments, I performed 3D-RNA FISH on AMD-treated cells. Strikingly, the pre-rRNA was observed in the nucleolar interior as expected but disnor238 RNA foci were observed embedded in the perinucleolar heterochromatin similar to that observed with 3D-DNA FISH performed with the DJ array. I confirmed that the RNA results are correct and it is not the DNA probe observed as the transcript by using 3D-DNA FISH and 3D-RNA FISH on AMD treated and untreated cells in interphase using probes against disnor238 transcript and the 11.9 kb rDNA EcoRI fragment that would not hybridize under RNA FISH conditions since it is from the IGS ie not transcribed.
I decided to use overexpression to confirm the source of these transcripts in cell clones with ectopic DJ arrays. 3D-RNA FISH on AMD treated cells in interphase from the DJBAC clone 2 showed more intense transcript hybridization than in the control cell lines and the transcript foci formed were observed to localize to the perinucleolar heterochromatin. To narrow down the overexpressed DNA sequence and to confirm that the transcripts detected by RNA-FISH derived from these specific DJ sequences, cells were transfected with BAC CU633906 that contained all three repeats but no rDNA or satellite DNA. Clone HT1080 NDJ 3 was chosen for further study since it was exhibiting a clearly distinguishable exogenous transcript that was associated with the nucleolus in the 3D-FISH screening. Using FISH on chromosomes in metaphase I found that the insert in clone NDJ 3 had integrated into the acrocentric chromosome 22. 3D immuno-FISH results on AMD treated cells using disnor238 transcript probe and NOP-52 Ab, revealed an intense RNA hybridization signal localizing to the nucleolar periphery, consistent with the results from the wild-type cells. Also, this clone seemed to exhibit more RNA from transcript disnor238 than in the wild type cells and is again always associated with the nucleoli. Once again, in order to confirm that the RNA results were correct, I used 3D-DNA FISH and 3D-RNA FISH with the same probes on interphases of AMD treated clone cells using the 11.9 kb that would not hybridize under RNA FISH conditions. The results confirmed the liability of the RNA FISH technique. These results combined suggest that disnor238 may function at or close its site of synthesis.

Proteins are encoded by the mRNA transcripts that are processed co-transcriptionally, through a complex RNA splicing mechanism, to give mature mRNAs (Wahl et al., 2009). After splicing, most mRNAs also have their 3’-ends polyadenylated (Licatalosi and Darnell, 2010). These two processes are catalytic for the maturation of the mRNA in order to be translated into a functional protein. However, recent studies have identified thousands of transcripts that even though do not seem to have any protein-coding capacity, are actually spliced and polyadenylated. These transcripts are called long non-coding RNAs (lncRNAs) and are involved in many cellular processes and particularly in developmental processes (Mattick, 2009). What is so interesting
about them is that even though they do not code for proteins, they contain
important genetic information that makes them have important regulatory roles
in transcription modulation. The most well known example of a IncRNA is the
X-inactive specific transcript (Xist), a IncRNA that is expressed by the inactive
X chromosome (Brockdorff et al., 1992) and is involved in transcriptional
silencing of the whole X chromosome by a process often described as
“coating” (Chow and Heard, 2009). Recently, another IncRNA was found
upstream of the Xist termed Jpx. Interestingly, Jpx was found to positively
regulate the Xist RNA expression from the inactive X chromosome, like the
eRNAs that will be descibed further down (Tian et al., 2010). Another finding
that is interesting to mention is that during genomic imprinting at the 2- and 4-
cell stage in mouse blastomeres, the silencing of the paternal X-inactivated
intergenic repetitive elements does not require Xist, the elements lie in the
perinucleolar heterochromatin, while active traditional coding genes (genic)
loci alongside the maternal active X chromosome, reside on the Cot-1 positive
regions of the nucleus and are silenced several divisions later (Namekawa et
al., 2010). The connection between the nucleolus and this imprinting silencing
phenomenon is yet unknown.

Recent studies have identified a class of IncRNAs that have a different role to
that of repression of transcription as it was described in the past. This class of
IncRNAs are involved in gene-specific activation (Orom et al., 2010) and are
also distal to protein-coding genes (De Santa et al., 2010). The disnor138
transcript, is likely to fall into a recently described class of enhancer-
associated RNAs (eRNAs) (Ren, 2010). eRNAs are described as a class of
IncRNAs that consist the region from which IncRNAs are transcribed
bidirectionally, are shorter in size than most IncRNAs and are suspected to
play an active role in the regulation of nearby genes (Kim et al., 2010, De
Santa et al., 2010). On the other hand, the length of the disnor187 and
disnor238 transcripts as well as their chromatin profiling characteristics
suggest they function as long intervening ncRNAs (lincRNAs). LincRNAs are
a large class of IncRNAs that are identified based on their chromatin histone
marks that are usually associated with transcription (Guttman et al., 2009,
Khalil et al., 2009). For example, they have an increased H3K4me3, that is
usually indicative of promoter function found at the start site of an actively transcribed gene as well as an increased H3K36me3 that is often indicative of the gene body of the transcribed gene. These histone marks as well as chromatin marks of Pol II transcription at the same position of H3K4me3 have been identified as mentioned earlier for both disnor187 and disnor238. Importantly, those two transcripts have also been shown to be spliced and all transcripts have been shown to be polyadenylated. This further shows that these lncRNAs are processed into mature lncRNAs. LncRNAs often function at their site of synthesis (Lee, 2012, Mercer et al., 2009) and in my experiments I observed disnor238 localizing to the perinucleolar heterochromatin. However, since the nascent and mature transcripts cannot be formally distinguished from each other in my experiments, the simplest interpretation of my results is that the transcripts' expression is a unique signature of active NORs.
7. Final Discussion
Final conclusions and future research.

In the short-arms of each of the five acrocentric human chromosomes lie nucleolar organizers, regions that direct the assembly and function of the nucleolus (Roussel et al., 1996). While much is known about the regulation of individual rDNA repeats within an NOR little is known about the regulation of complete NORs. Active NORs form nucleoli that can fuse forming mature nucleoli comprising multiple active NORs. Silent NORs are dissociated from nucleoli and have a condensed chromatin state (McStay and Grummt, 2008). The mechanisms determining the activity status of NORs, a fundamental question in nucleolar biology, have remained mysterious.

Work from the Grummt lab has focused on the NoRC complex as a mechanism for individual rDNA repeat silencing (Strohner et al., 2001). They describe TIP5 in mice and BAZ2A in humans as being one of the key factors in this silencing complex. It has also been suggested that this complex may be implicated in the inactivation of whole NORs. In chapter 3 of this thesis I provide evidence that BAZ2A, is neither a nucleolar protein nor involved in NOR regulation as previously claimed. This was achieved through the development of a highly specific antibody that could be used in western blotting, immuno-staining and 3D-immunoFISH. BAZ2A was shown to be a nuclear protein that is excluded from the nucleolus. Furthermore, depletion of this protein did not seem to have any observable effect on nucleolar structure, did not affect the levels or distribution of the Pol I machinery in the nucleus and did not re-activate silent NORs in the cells. Further doubt on a role for BAZ2A in NOR regulation has come from the recent finding that both human BAZ2A and its Drosophila counterpart are involved in the regulation of transcription by RNA polymerase II outside of nucleoli (Emelyanov et al., 2012).

In the remaining chapters of the thesis I focused my attention on the contribution that non-rDNA sequences play in NOR structure and regulation. It was revealed that NORs are more than a simple rDNA array. Sequences
located distally and proximally to the ribosomal gene arrays were found to be conserved among all the acrocentric chromosomes and to share a complex genomic architecture. Although these regions were believed to be heterochromatic, repeat-rich deserts, it was shown that their composition is similar to euchromatic regions of the genome. However, they have distinct genomic characteristics. Sequences proximal to the rDNA are almost entirely segmentally duplicated, like those surrounding centromeres (She et al., 2004). Therefore, the PJ sequence is highly unlikely to contain any nucleolus-specific functional elements. Instead, a possible suggestion is that they form a buffer zone, protecting the rDNA array from the rest of the chromosome or vice versa. The segmental duplications and the unexpectedly high level of interchromosomal identity, suggests that the PJ undergoes frequent interchromosomal recombination. These recombination events that homogenize the acrocentric short arms may be responsible for Robertsonian translocations (translocations between acrocentric short arms that appear to derive from pericentromeric regions) (Therman et al., 1989) and are associated with genetic disorders. The PJ sequence could be the perfect candidate to further investigate this observation.

In contrast, the DJ region is full with unique sequences and was mapped exclusively to the distal region to the NORs as shown from 3D-FISH results on monochromosomal cell lines containing either the DJ or the PJ. However, the same results implied that the PJ apart from proximally also maps distally to the rDNA. This phenomenon was found to be possibly due to PJ-like sequences existing at the end of the DJ such as the newly identified sequence (BAC AL591856) that was mapped at the end of the then known DJ sequence. Interestingly, the DJ displays evidence of functionality. One example of this manifestation comes from results from 3D-FISH and immuno-FISH experiments suggesting that the DJ is acting as a “control panel” for the localization and/or regulation of the entire NOR. By determining the transcriptional status of the linked rDNA array and if active DJ anchors the rDNA to the perinucleolar heterochromatin and if inactive the NOR loses this localization and does not participate in nucleolar formation. Consequently, DJ probes can now be used to count the number of NORs per nucleolus since
the number of nucleoli is less than the number of active NORs in most human cell lines.

The introduction of ectopic DJ arrays (3 BACs encompassing the whole DJ contig) in the cells showed that if integrated into nonacrocentric chromosomes they can redirect them to the nucleolus approximately 95% of the time. Since it was shown that the integrants do not contain any active rDNA repeats, the view of the DJ sequence containing the localization information is further strengthened. I observed that in these DJ array clones the endogenous rDNA arrays are not efficiently withdrawn to the nucleolar periphery upon AMD treatment. This suggests that anchoring of rDNA has been titrated out by the DJ array. Other work in the McStay lab (Van Sluis and McStay unpublished) has provided good evidence that efficient repair of double strand breaks in rDNA by the homologous recombination repair pathway is dependent on efficient nucleolar segregation. Thus, it could be predicted that disruption of NOR anchoring may be associated with genetic instability within the rDNA array or even more widespread on the acrocentric short arms.

Analyses in silico and wet lab approaches revealed a complex transcription landscape in the DJ that is characterized by zones of promoter activity that have the same spacing as rDNA units, as well as the presence of a large inverted repeat. Strikingly, characterization of transcripts from these promoters suggests they are IncRNAs that localize to the NORs where they are synthesized as it was shown from 3D-RNA FISH cells in interpahse. DJ clones with inserted ectopic BAC arrays were shown to exhibit transcript disnor238 overexpression and highlight the perinucleolar localization of these polyadenylated transcripts. Therefore, they may function in cis to regulate local chromatin structure, such as has been found for other IncRNAs (Lee, 2012, Mercer et al., 2009).

The sequences described here begin to fill one of the major remaining gaps in the human genome, the short arms of the acrocentric chromosomes. Their identification is an important step towards the complete understanding of
nucleolar biology and also provides a scaffold from which the completion of the short arms can build on. The high level of segmental duplication observed in these sequences and particularly of the PJ, suggests that some previously identified nucleolus-associated chromatin domains (NADs) may actually be segmental duplicates (Nemeth et al. 2010, van Koningsbruggen et al. 2010). Hence, our sequences will permit the development of a more refined picture of NADs. Furthermore, as mentioned earlier the DJ sequence will now allow the use of hybridization-based approaches to determine whether human nucleoli contain multiple NORs which has remained difficult to prove.

This work provides the tools to determine the molecular mechanisms that underlie the well-known heterogeneity in nucleolar morphology observed between various cancers that serve as diagnostic indicators (Derenzini et al., 2009), suggesting the existence of mechanisms other than direct upregulation of rDNA transcription to have a role in the development of malignancy. Since acrocentric short arms are implicated in both nucleolar form and function, we hypothesize that genetic alterations on these chromosome arms contribute to tumorigenesis and other human diseases. Hence, the DJ and PJ sequences lay the foundations for addressing the roles that genetic and epigenetic changes in them play in human disease and also provide a plethora of new tools for studying nucleolar biology.

Future work should include further characterization of the DJ sequences and in particular their transcripts. One way to study the function of these transcripts and their epigenetic mechanisms could be the use of antisense RNA (asRNA). When asRNA is introduced into the cells, it can bind to any translated mRNAs complementary to its sequence and obstruct their translation (Weiss et al., 1999). Another interesting approach would be the use of ever improving genome editing technologies such as CRISPR (clustered regularly interspaced short palindromic repeats) to delete or over express any genes that may code for these transcripts (Marraffini and Sontheimer, 2010). RNA pull down assay, that can selectively extract a Protein-RNA complex from a sample, could also be used to identify what factors are associated with the DJ transcripts (Khanam et al., 2006).
Furthermore, alterations in the DJ sequence and/or their function could be investigated in cancer cells; for example whether disnors are found to be upregulated. Finally, the role that efficient DJ function plays in genome stability should be investigated. It could be directly tested whether disrupted DJ function gives rise to a reduced ability to repair double strand breaks in rDNA resulting in enhanced genome instability of the rDNA array for example.
REFERENCES


French, S. L., Y. N. Osheim, F. Cioci, M. Nomura and A. L. Beyer. "In Exponentially Growing Saccharomyces Cerevisiae Cells, Rrna Synthesis Is Determined by the


van Koningsbruggen, S., M. Gierlinski, P. Schofield, D. Martin, G. J. Barton, Y. Arlyurek, J. T. den Dunnen and A. I. Lamond. "High-Resolution Whole-


APPENDIX
The shared genomic architecture of human nucleolar organizer regions

Ioanna Floutsakou,1,4 Saumya Agrawal,2,4 Thong T. Nguyen,3,4 Cathal Seoighe,3 Austen R.D. Ganley,2 and Brian McStay1,5

1Centre for Chromosome Biology, School of Natural Sciences, National University of Ireland, Galway, Galway, Ireland; 2Institute of Natural and Mathematical Sciences, Massey University, Auckland 0632, New Zealand; 3School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Galway, Ireland

The short arms of the five acrocentric human chromosomes harbor sequences that direct the assembly and function of the nucleolus, one of the key functional domains of the nucleus, yet they are absent from the current human genome assembly. Here we describe the genomic architecture of these human nucleolar organizers. Sequences distal and proximal to ribosomal gene arrays are conserved among the acrocentric chromosomes, suggesting they are sites of frequent recombination. Although previously believed to be heterochromatic, characterization of these two flanking regions reveals that they share a complex genomic architecture similar to other euchromatic regions of the genome, but they have distinct genomic characteristics. Proximal sequences are almost entirely segmentally duplicated, similar to the regions bordering centromeres. In contrast, the distal sequence is predominantly unique to the acrocentric short arms and is dominated by a very large inverted repeat. We show that the distal element is localized to the periphery of the nucleolus, where it appears to anchor the ribosomal gene repeats. This, combined with its complex chromatin structure and transcriptional activity, suggests that this region is involved in nucleolar organization. Our results provide a platform for investigating the role of NORs in nucleolar formation and function, and open the door for determining the role of these regions in the well-known empirical association of nucleoli with pathology.

[Supplemental material is available for this article.]

A detailed description of how the genome is organized within the nuclei of human cells to facilitate proper cellular functions is one of the major unsolved problems in biology. While genome-wide technologies are beginning to have an impact (Lieberman-Aiden et al. 2009; The ENCODE Project Consortium 2012), the picture is incomplete since critical regions of the human genome remain unidentified. Prominent among these missing regions is the chromosomal context around which nucleoli form, termed the nucleolar organizer region (NOR) (McClintock 1934). The nucleolus is the largest functional domain within the nucleus and is the site of ribosome biogenesis. It has a distinct structure and houses ribosomal RNA gene (rDNA) transcription, preribosomal RNA (pre-rRNA) processing, and preribosome assembly (Olson 2011). The rDNA repeats encode the major rRNA species and are organized into large head-to-tail tandem arrays located at the NORs (McStay and Grummt 2008). Extensive binding by the nucleolar DNA binding protein, UBF (encoded by UBF), across the rDNA is responsible for their distinctive appearance on metaphase chromosomes, in which they form secondary constrictions (Mais et al. 2005; McStay and Grummt 2008). Despite ribosome biogenesis being central to cellular biology, many aspects of nucleolar formation, organization, and function remain to be elucidated, and the genomic architecture of NORs has not been described for any vertebrate to date.

In humans, the approximately 300 rDNA repeats are distributed among five NORs on the short arms of the acrocentric chromosomes (HSA13-15, HSA21, and HSA22) (Henderson et al. 1972; Schmickel 1973; Stults et al. 2008). In most human cells, a majority of NORs are active and coalesce to form between one and three nucleoli (Savino et al. 2001). While discrete, nucleoli are not encapsulated and instead seem to be self-organizing, highly dynamic structures (Dundr et al. 2002; Andersen et al. 2005; Sirri et al. 2008) that are spatially isolated from the rest of the nucleoplasm by a shell of heterochromatin (Nemeth and Langst 2011). However, the apparent heterochromatic nature of the regions flanking the rDNA has made them a low priority for genomic analysis (International Human Genome Sequencing Consortium 2004), and thus, the five acrocentric chromosome short arms are missing from the current human genome assembly.

The enormous demand for ribosomes by actively growing cells puts nucleoli at the forefront of cellular growth regulation. Links between nucleoli and growth pathologies date back over a century to observations of abnormal nucleoli in tumor cells (Pianese 1896). Recently, molecular studies have begun to clarify this relationship, with evidence supporting roles for tumor suppressor genes and oncogenes in rDNA transcriptional regulation (Budde and Grummt 1999; Hannan et al. 2000; Grummt 2003; Grandori et al. 2005), and a direct role for increased rDNA transcription in the development of malignancy (Bywater et al. 2012). Surprisingly, recent studies have also revealed that the nucleolus plays roles in many other biological processes, ranging from aging and cell cycle progression to X-chromosome inactivation and viral replication (Visintin et al. 1999; Boisvert et al. 2007; Zhang et al. 2007; Ganley et al. 2009). The central role of the nucleolus in growth regulation, coupled with the potential for the regions adjacent to the rDNA to contribute to NOR.
function, prompted us to investigate the genomic context in which the rDNA resides. Here we present the identification and characterization of >550 kb of sequence flanking both sides of the rDNA array.

Results

Identification of rDNA flanking regions

To obtain the sequences flanking the rDNA arrays, we made use of preexisting sequences adjacent to the rDNA on the proximal (centromeric; 493 bp) (Sakai et al. 1995) and distal (telomeric; 8.3 kb) sides of the rDNA (Gonzalez and Sylvester 1997). Probes derived from these sequences were used to screen single-chromosome cosmid libraries, and several positive clones were sequenced. Searches of GenBank using the resulting sequences, coupled with BAC walking, ultimately resulted in identification of 15 BAC clones from the distal junction (PJ), five of which include some rDNA sequence (Supplemental Fig. 1). Similar searches identified three BAC clones from the proximal junction (DJ), one of which includes rDNA (Supplemental Fig. 2). These cover 379 kb of DJ sequence and 207 kb of PJ sequence flanking the rDNA (Fig. 1A).

We sought evidence that these putative junction sequences adjoin the rDNA. Hybridization of DJ BAC clones to metaphase chromosome spreads places these regions distal to the rDNA on the acrocentric chromosome short arms (Fig. 1B). Further, FISH on combed DNA molecules (Bensimon et al. 1994) clearly shows the acrocentric region (Fig. 1C). Finally, a bioinformatic screen confirmed that the DJ adjoins the rDNA (Supplemental Methods).

A schematic human acrocentric chromosome showing telomeric (blue) and centromeric (yellow) regions, and the NOR (black line), expanded below into rDNA, PJ (orange), and DJ (green) regions. Not to scale. (B) DJ and PJ localize distally and proximally to rDNA, respectively, on all acrocentric chromosomes. FISH was performed on normal human metaphase spreads with DJ BAC (red) and rDNA (green) probes (left panels), and PJ BAC (green) and rDNA (red) probes (right panel). Chromosomes are DAPI-stained. (C) DNA combing of HeLa cell nucleolar DNA shows DJ (red) is physically linked to 18S rDNA (green). Three representative images are shown below the hybridization scheme.

Interchromosomal conservation of rDNA flanking regions

Previous reports suggested that the sequences distal to the rDNA are conserved across all acrocentric chromosomes (Worton et al. 1988; Gonzalez and Sylvester 1997). Our DJ metaphase FISH results are consistent with this, showing hybridization signals for all the acrocentric chromosomes (Fig. 1B). The FISH also suggests that the PJ is conserved across the acrocentric chromosomes (Fig. 1B). To confirm this and to further validate the integrity of the flanking region sequences, we screened genomic DNA from a panel of mouse somatic cell hybrids, each containing a single human acrocentric chromosome for flanking region sequence. PCR at five intervals across the DJ each gave the expected product for all five acrocentric chromosomes (Fig. 2A). We could not screen across the PJ because only one unique region is present (see below). Nevertheless, this region is also shared among all five acrocentric chromosomes (Fig. 2A). Furthermore, screening genomic DNA from lines containing a chr21 translocation originating within the rDNA confirmed the orientation of the DJ and PJ relative to the rDNA (Fig. 2A) and that the unique PJ region is located exclusively on the proximal side. These results suggest that both the DJ and the PJ are conserved across all five acrocentric chromosomes.

The FISH also suggests that the PJ is conserved across all acrocentric chromosomes (Fig. 1B). To confirm this and to further validate the integrity of the flanking region sequences, we screened genomic DNA from a panel of mouse somatic cell hybrids, each containing a single human acrocentric chromosome for flanking region sequence. PCR at five intervals across the DJ each gave the expected product for all five acrocentric chromosomes (Fig. 2A). We could not screen across the PJ because only one unique region is present (see below). Nevertheless, this region is also shared among all five acrocentric chromosomes (Fig. 2A). Furthermore, screening genomic DNA from lines containing a chr21 translocation originating within the rDNA confirmed the orientation of the DJ and PJ relative to the rDNA (Fig. 2A) and that the unique PJ region is located exclusively on the proximal side. These results suggest that both the DJ and the PJ are conserved across all five acrocentric chromosomes.

To quantify how similar the DJ and PJ regions are between the different acrocentric chromosomes, we determined...
Genomic architecture of human nucleolar organizers

The human genome as a whole (Supplemental Fig. 8), except for a lack of known DNA transposons in the DJ. Both regions also contain blocks of satellite repeats, most notably a large (38.6 kb) block of 48-bp satellite repeats at the distal end of the DJ (Fig. 3A). These repeats, initially classed as chr22 pericentromeric repeats (Metzdorf et al. 1988), are now referred to as CER satellites (Jurka et al. 2005). CER blocks are found distal to the rDNA on all acrocentric chromosomes, with additional pericentromeric blocks on chr14 and chr22 (Supplemental Fig. 9). Finally, we looked for novel repeats. Two blocks of a novel 138-bp tandem repeat that we call ACRO138 are present within the DJ (Fig. 3A; Supplemental Fig. 10).

Most strikingly, we discovered a large inverted repeat that dominates the DJ (Fig. 3A). Each arm of the inverted repeat is ~109 kb, and the two arms share an average sequence identity of 80%. Alignment of the two arms reveals that the underlying sequence identity is higher than this but is interrupted by a number of indels (Supplemental Fig. 11), implying either that the inverted repeat is young or that there are mechanisms to maintain sequence identity between the arms.

Although these rDNA flanking regions were thought to be heterochromatic (International Human Genome Sequencing Consortium 2004), we wondered whether they contain any genes. To address this, we designed a gene predication pipeline that integrates ORF prediction, mRNA, EST, and protein data to identify potential gene-coding regions. Eight putative DJ genes and four PJ genes were predicted. These gene models all had support from multiple data sources, but the majority are single exon (Supplemental Fig. 12; Supplemental Tables 1, 2). Experimental validation will be required to determine which of these putative genes are real.

Segmental duplications (duplications of DNA to elsewhere in the genome) are a prominent feature of the human genome and are commonly enriched at centromere boundaries (She et al. 2004; Bailey and Eichler 2006). Given the proximity of the DJ and PJ to centromeres we undertook a segmental duplication analysis. A number of segmental duplications (>1 kb in length and with >85% sequence identity) were found from both regions (Fig. 3B). Interestingly, the DJ and PJ show different segmental duplication patterns. PJ segmental duplicates are more frequent and longer and have greater sequence identity than DJ segmental duplicates (Fig. 3B; Supplemental Tables 3, 4). Furthermore, the majority of PJ segmental duplicates are found in centromeric/pericentromeric regions of the genome as previously observed (Piccini et al. 2001; Lyle et al. 2007), while the majority of DJ segmental duplicates are found in euchromatic/telomeric regions (Fig. 3B). Most strikingly, we found that the level of segmental duplicated DNA is vastly different, with 7.3% of the DJ being segmentally duplicated versus 92.4% for the PJ (Table 1). These results demonstrate that these two rDNA flanking regions have different genomic characteristics in humans: The segmental duplication profile of the PJ resembles pericentromeric regions, while that of the DJ resembles euchromatic regions. The high level of segmental duplication likely explains the problems encountered using hybridization-based approaches with the PJ. Finally, this high level of segmental duplication prevented us from extending the PJ further, as we could not unambiguously assign additional sequences to the PJ region.

**Perinucleolar positioning of the DJ**

To better understand the relationship between these flanking regions and nucleolar architecture, we analyzed DJ and PJ positioning in interphase nuclei of HT1080 cells. 3D immuno-FISH revealed that, despite their close proximity to rDNA, DJ sequences...
are excluded from the nucleolar interior, instead appearing as discrete foci embedded in the perinucleolar heterochromatin (Fig. 4A). The majority of DJ foci associate with perinucleolar heterochromatin (Supplemental Videos 1, 2). The high level of PJ segmental duplication makes it difficult to determine which FISH signals derive from the PJ versus from unlinked segmentally duplicated regions (Supplemental Fig. 13); therefore, we focused on the DJ for the remainder of this study.

The rDNA is transcribed by RNA polymerase I (pol-I), and inhibition of pol-I transcription by low concentrations of actinomycin D (AMD) induces rapid and remarkable nucleolar reorganization (Hadjiolov 1985). This involves dissociation of nascent rRNA transcripts, resulting in collapse of the rDNA repeats into nucleolar caps that form at the nucleolar periphery and consist of rDNA, pol-I transcription machinery, and processing factors (Supplemental Fig. 14; Prieto and McStay 2007; Sirri et al. 2008). Given the localization of the DJ, we wondered whether it has any relationship to these nucleolar caps. Strikingly, 3D immuno-FISH on AMD-treated cells revealed that nucleolar caps form immediately adjacent to DJ sequences positioned in perinucleolar heterochromatin (Fig. 4B). Although some larger nucleolar caps are bilobed and associated with two DJ signals, it appears that the majority of caps are derived from individual NORs. These results suggest that the DJ anchors the linked rDNA to perinucleolar heterochromatin and that the retreat of the rDNA to the DJ upon AMD treatment provides an explanation for the positioning of nucleolar caps (Fig. 4B).

**DJ sequences can target perinucleolar heterochromatin**

rDNA arrays regress to perinucleolar heterochromatin in AMD-treated cells, rather than DJ sequences moving toward rDNA foci within the nucleolar interior (Fig. 4). This suggests that elements within the DJ may be responsible for its perinucleolar localization.

**Table 1. Segmental duplication comparison between the DJ and PJ**

<table>
<thead>
<tr>
<th>Segmental duplication feature</th>
<th>DJ</th>
<th>PJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of segmentally duplicated regions</td>
<td>31</td>
<td>98</td>
</tr>
<tr>
<td>Average segmental duplicate length (kb)</td>
<td>2.3</td>
<td>11.8</td>
</tr>
<tr>
<td>Average percent identity between segmental duplicates</td>
<td>88.2%</td>
<td>93.1%</td>
</tr>
</tbody>
</table>
rather than the DJ being simply linked to the rDNA. To investigate this, we transfected HT1080 cells with a mixture of three BACs that cover the DJ contig. Two stable clones containing large integrated arrays of this BAC mixture were selected for further analysis. FISH with DJ BAC CT476834 and rDNA probes revealed that ectopic DJ BAC arrays had integrated into metacentric (non-NOR bearing) chromosomes (Supplemental Fig. 15A). Quantitative PCR using primer pairs positioned across the DJ contig revealed that the sequence content of the ectopic arrays reflects that of the input BAC mixture (Supplemental Table 5). 3D FISH was performed on cells from these clones to determine whether the ectopic DJ arrays associate with nucleoli. In order to more clearly reveal the boundaries of the nucleolus, cells were treated with AMD (Fig. 5). For both clones, we observe a remarkable degree of association of the ectopic DJ arrays with perinucleolar heterochromatin. Moreover, in the majority of cells, the ectopic DJ array appears to spread through the perinucleolar heterochromatin, covering a significant fraction of the nucleolar surface. Transcription from the small amounts of rDNA derived from BAC AL592188 does not appear to explain this localization, as silver staining (which can detect activity at NORs with comparable rDNA content; A Grob and B McStay, in prep.) shows no activity. Additionally, transcriptionally active ectopic rDNA arrays (neo-NORs) that do not contain any DJ sequences are usually not associated with endogenous nucleoli (A Grob and B McStay, in prep.). Therefore, we conclude that sequences within the DJ contig specify association with perinucleolar heterochromatin even when positioned on metacentric chromosomes. (Supplemental Fig. 15B).

Chromatin profiling of the DJ

Our results suggest that the DJ region plays a role in nucleolar organization. Specifically, we hypothesize that DJ sequences provide an anchor point within perinucleolar heterochromatin for the linked rDNA array that is normally present in the nucleolar interior. If so, the DJ may have a chromatin organization that facilitates this role. To profile its chromatin organization, we mapped available histone modification and insulator binding protein CTCF ChIP-seq data sets from the ENCODE Project (Ernst et al. 2011) to the DJ. Discrete patterns of enrichment were observed at specific points across the DJ (Fig. 6A). Integration of these chromatin data sets using ChromHMM (Ernst and Kellis 2012) with multiple cell types revealed a complex chromatin landscape that is largely conserved among cell types (Fig. 6B; Supplemental Fig. 16). Strikingly, chromatin signatures characteristic of promoters are found at regular ~45-kb intervals across the DJ, interspersed among marks associated with heterochromatin. The periodicity (~45 kb) of these putative promoters is interesting, as it closely mirrors the size of the rDNA unit. Chromatin marks indicative of promoters (e.g., H3K4me3) coincide with DNase hypersensitive sites and FAIRE signals (Supplemental Fig. 17), and we experimentally confirmed that the H3K4me3 and FAIRE peaks are present in the HT1080 cell line used in our immuno-FISH experiments (Fig. 6C). The open chromatin peaks centered at 138 kb and 290 kb.
correspond to the ACRO138 repeat blocks identified in the repeat analysis (Fig. 3A). Moreover, chromatin signatures associated with actively transcribed gene bodies (e.g., H3K36me3 and H3K20me1) (Ernst et al. 2011) are observed extending leftward and rightward from the promoters at 187 kb and 238 kb, respectively, (Fig. 6A). CTCF, a multivalent DNA binding protein involved in many cellular processes (Phillips and Corces 2009), has recently been shown to be involved in the transcriptional regulation of ribosomal genes (van de Nobelen et al. 2010) and human nucleolar organization (Hernandez-Hernandez et al. 2012). Multiple CTCF binding peaks correspond to the ACRO138 repeat blocks identified in the repeat analysis (Fig. 3A). Moreover, chromatin signatures associated with actively transcribed gene bodies (e.g., H3K36me3 and H3K20me1) (Ernst et al. 2011) are observed extending leftward and rightward from the promoters at 187 kb and 238 kb, respectively, (Fig. 6A). CTCF, a multivalent DNA binding protein involved in many cellular processes (Phillips and Corces 2009), has recently been shown to be involved in the transcriptional regulation of ribosomal genes (van de Nobelen et al. 2010) and human nucleolar organization (Hernandez-Hernandez et al. 2012). Multiple CTCF binding peaks were observed across the DJ, coinciding with CTCF consensus sequences (Supplemental Fig. 18). Interestingly, CTCF binding sites are positioned close to the DJ/rDNA boundary and frame the DJ transcription units described above. Together, these results reveal that the DJ has a complex and structured chromatin landscape.

**Figure 6.** Chromatin landscape of the DJ. (A) ChIP-seq signals of different chromatin features (right) in H1-hESC cells, normalized to tags per million mapped reads are shown below a schematic of the DJ, including inverted repeats. Asterisks indicate enrichment sites. (Bottom) Control signal is shown in gray. (B) Chromatin states derived from the multivariate HMM analysis for seven different cell types (right). Each colored bar represents a specific chromatin state, as annotated below left. (C) Nucleolar H3K4me3 ChIP-PCR and nucleolar FAIRE-PCR using HT1080 cells validate the presence of H3K4me3 and FAIRE in the DJ. DJ positions of the primers used are shown to the right, and red boxes correspond to peaks of H3K4me3 from A. Genomic DNA (gDNA), input and negative controls (ve and IgG) are shown.

Transcription profiling of the DJ

The chromatin profiling results suggest that despite being embedded in perinucleolar heterochromatin, DJ sequences are tran-
specific functional elements. The high level of interchromosomal meres (She et al. 2004), and thus are unlikely to contain nucleolus-segmentally duplicated regions, like those surrounding centro-
and have a complex sequence feature composition.
the rDNA are conserved across all five acrocentric chromosomes
terization of these regions. We reveal that the sequences flanking
this study, we have identified this.
yr later, we can begin to appreciate their genomic architecture. In
expression levels (Supplemental Fig. 20). The largest open reading frames present in disnor187 and disnor238 are 120 and 144 amino acids, respectively. Therefore their size and limited coding capacity suggest that they may function as long noncoding RNAs (lncRNAs). We also confirmed that the ACR0138 repeats within the open chromatin peak at 138 kb are actively transcribed (Supplemental Fig. 21). These three transcripts all lie within the DJ inverted repeat arms and, together with another putative promoter in the ACR0138 repeat block centered at 297 kb, form a symmetrical arrangement of transcriptional units that mirror the inverted repeat structure. Therefore the DJ, rather than being a passive block of heterochromatin, shows a specific pattern of localization, a distinct genomic and chromatin organization, and transcriptional activity.

Discussion
NORs were originally defined in 1934 as chromosomal regions that organize formation of the nucleolus (McClintock 1934). Now ~80 yr later, we can begin to appreciate their genomic architecture. In this study, we have identified >550 kb of sequence from the regions flanking the rDNA array and have performed an in-depth characterization of these regions. We reveal that the sequences flanking the rDNA are conserved across all five acrocentric chromosomes and have a complex sequence feature composition.

Sequences proximal to the rDNA almost entirely consist of segmentally duplicated regions, like those surrounding centromeres (She et al. 2004), and thus are unlikely to contain nucleolus-specific functional elements. The high level of interchromosomal sequence conservation of the PJ, coupled with its high level of segmental duplication, strongly suggests that, far from being a recombinationally inert region of the genome, the PJ experiences frequent and ongoing recombination. This recombination appears to occur predominantly with other peri-/centromeric regions of the genome, implying colocalization of these regions, and may be responsible for Robertsonian translocations that appear to derive from the pericentromeric regions of acrocentric short arms (Therman et al. 1989) and are associated with genetic disorders. Our identification of the sequence of the PJ provides a means to investigate this.

In contrast, the DJ region is replete with unique sequences and displays evidence of functionality. We show that the DJ is local-ized to perinucleolar heterochromatin, where it appears to anchor the rDNA array to this region, and sequences located within the DJ are likely to be important for this localization. We propose that the DJ acts as a “control panel” for the entire NOR, where it can determine the transcriptional status of the linked rDNA array. In this way, the DJ may be involved in regulating the coalescence of nucleoli around individual NORs. Our model is that active NORs are localized to perinucleolar heterochromatin, where they form nucleoli, while inactive NORs lose this localization and form silent arrays that do not participate in nucleoli.

The sequences we describe begin to close one of the major remaining gaps in the human genome, the short arms of the acrocentric chromosomes, and their identification is an important step toward a complete understanding of nucleolar biology. The level of segmental duplication we observe, particularly of the PJ, suggests that some previously identified nucleolus-associated chromatin domains (NADs) may actually be segmental duplicates (Nemeth et al. 2010; van Koningsbruggen et al. 2010), and our sequences will allow a more refined picture of NADs to be developed. Additionally, the DJ sequence will allow researchers to use hybridization-based approaches to determine whether human nucleoli contain multiple NORs (Supplemental Videos 1, 2), something that has remained difficult to prove. The degree of heterogeneity in nucleolar morphology observed between cancers (Derenzini et al. 2009) suggests that mechanisms other than direct up-regulation of rDNA transcription have a role in the development of malignancy. As acrocentric short arms underpin both nucleolar form and function, we hypothesize that genetic alterations on these chromosome arms contribute to tumorigenesis and other human diseases. Therefore the sequences that we...
report here lay the foundations for addressing the roles that genetic and epigenetic changes in the DJ and PJ play in human disease, as well as providing a wealth of new tools for studying nucleolar biology.

Methods

Genomic cosmid and BAC clones

Acrocentric chromosome cosmid libraries LA13 NC01, LA14 NC01, LA15 NC01, LL21 NC02, and LL22 NC03 were obtained from the UK HGMP resource center. To obtain cosmids spanning the DJ, libraries were screened with a 638-bp PCR product generated using DJ1f/DJ1r primers (Worton et al. 1988). To obtain cosmids spanning the PJ, libraries were screened with a 220-bp PCR product generated using PJf/PJr primers (Sakai et al. 1995). Clone names for DJ and PJ cosmids identified and used in this study are shown in Supplemental Figure 4. We identified BAC clones in GenBank representing the DJ and PJ regions flanking the rDNA using BLAST. Cosmids LA14 138F10 and N 29M24 were used as the initial query sequences to search for BACs representing the DJ and PJ, respectively (Supplemental Methods). BAC clones were obtained from BACPAC Resources.

DNA sequencing and assembly

DNA sequencing of cosmid clones was performed through a combination of standard Sanger and next-generation sequencing (NGS). The sequence of the insert in cosmid LA14 138F10 and most of the insert in N 29M24 were determined by Sanger sequencing. The inserts of the remaining cosmids were end sequenced using Sanger sequencing and then subjected to NGS. Indexed libraries were prepared from individual cosmids using a Nextera DNA sample prep kit and Nextera barcodes (Epicentre NGS). NGS was performed on an Illumina Genome Analyzer IIx, using 54-bp singleton processing (Ambry Genetics). Sequences of cosmids were assembled using ABySS v1.2.7 (for parameters, see Supplemental Methods) (Simpson et al. 2009). Velvet v.1.01 (Zerbino and Birney 2008) was used to refine the ABySS assemblies.

Cell lines

HeLa cells and HT1080 human fibrosarcoma cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS). RPE-1 cells were maintained as above but supplemented with 5% FBS, 2 mM L-glutamine, and 0.348% sodium bicarbonate. Mouse A9 cells containing individual human acrocentric chromosomes were previously described (Sullivan et al. 2001), and those containing X/21 reciprocal translocation products (GM09142 and GM10063) were obtained from Coriell Cell Repositories. To generate cells that contain ectopic DJ arrays, HT1080 cells were cotransfected using a standard calcium phosphate protocol with BACs AL592188, CT476834, and AC011841 together with a blasticidin selection marker in a 200:1 w/w ratio. Stable transfectants were maintained as above but supplemented with 5 μg/mL blasticidin.

FISH and 3D immuno-FISH and RNA FISH

Probes for FISH experiments were labeled using spectrum red or green dUTP (Abbott Molecular). For chromosome mapping experiments, slides of human normal male metaphase chromosome spreads (Applied Genetics) were denatured in 70% formamide/2× SSC for 5 min at 73°C. Slides were then dehydrated through a 70%/100% ethanol series, washed, and air dried. Denatured probe (50 ng/slide) combined with human Cot-I DNA (10 μg/slide) in 20 μL/slide Hybrisol VII (Qbiogene) was then added to the slides and allowed to hybridize for 24–48 h at 37°C in a humidified chamber. For CER satellite detection, hybridizations were performed with a 5’ FITC-labeled oligo and herring sperm DNA. Post-hybridization washes were 0.4× SSC/0.3% NP-40 for 2 min at 74°C followed by 2× SSC/0.1% NP-40 at ambient temperature for 1 min. Slides were air dried and mounted in Vectorshield, including DAPI (Vector Laboratories). For 3D immuno-FISH experiments, cells were fixed, denatured, probed, and antibody stained as described previously (Mais et al. 2005; Prieto and McStay 2007). Z-stacks of fluorescent images were captured using a Photometric Coolsnap HQ camera and Velocity 5 imaging software (Improvision) with a 63 Plan Apochromat Zeiss objective mounted on a Zeiss Axioplan2 imaging microscope. In some cases, extended focus projections of deconvolved Z-stacks (iterative restoration) are presented; in other cases, individual focal planes are shown. Movies (Supplemental Videos 1, 2) were prepared from 3D images constructed from deconvolved Z-stacks using Velocity 6 (Improvision). 3D images were rotated to create a series of bookmarks. The movies are an animation of the transitions between these bookmarks.

Nucleolar DNA combing

Nucleoli, prepared from HeLa cells as previously described (Andersen et al. 2002), were resuspended at a concentration of 1×10⁸ to 2×10⁸ cell equivalents/mL in TE (10 mM Tris at pH 8.0, 1 mM EDTA). Resuspended nucleoli were mixed with an equal volume of 1% low melting point agarose in TE at 50°C. The mixture was pipetted into a plug mould (BioRad, 100 L/slot). Embedded nucleoli were deproteinized, and encapsulated high-molecular-weight nucleolar DNA was combed onto silanized coverslips as previously described (Caburet et al. 2005) using a Molecular Combing apparatus supplied by Genomic Vision Paris. Coverslips were then hybridized with biotin-labeled DJ cosmid LA14 138F10 and a digoxigenin-labeled 5.8-kb EcoRI restriction fragment that contains S′ ETS and 18S RNA from human rDNA. Hybridization and detection were performed as described previously (Caburet et al. 2005).

Bioinformatic analyses

The bioinformatics pipeline that was used to identify potential junction regions from whole-genome sequence data is described in Supplemental Methods, as is the method for determining intra-/interchromosomal sequence identity between DJ and PJ clones.

Construction and analysis of DJ and PJ contigs

To construct the DJ, four BACs were merged (CT476837, CT476834, CU633906, and AC011841). The overlapping regions between BACs CT476837 and CT476834 and BACs CT476834 and CU633906 are 100% identical, but the identity decreases to ~98% between CU633906 and AC011841 (Supplemental Fig. 7A). To form the PJ, two BACs (CR392039 and CR381535) with an identical overlapping region were merged to obtain a single contig (Supplemental Fig. 7B). The PJ is identical to that previously published (Lyle et al. 2007) and is deposited in GenBank under accession no. NT113958. The repeat content analysis method and the gene identification pipeline that utilizes gene prediction, mRNA, EST, and protein sequence data to identify potential DJ and PJ genes are both presented in Supplemental Methods. Segmental duplicates in the DJ and PJ contigs were detected using a modified BLAST-based detection scheme called the “whole genome assembly comparison” (Bailey et al. 2001). The human genome assembly (hg19) was
broken into 400-kb pieces. Repeats in this fragmented human genome and in the DJ/PJ contigs were masked using RepeatMasker (Smit et al. 2010). DJ and PJ contigs were then matched to the fragmented masked genome using BLAST, with a cutoff of ≥85% identity >1 kb. Next, the repeats were reinserted into these matched sequences, and global alignments were created. All steps were performed using a series of Perl scripts (J. Bailey, University of Massachusetts Medical School). Low identity ends of the sequences were identified from the alignments and trimmed. Where the ends of two human genome fragments match a single region or where a fragment is interrupted by repeats, these fragments were merged together. The merged sequences were then aligned again to recalculate the identity.

ENCODEx data
ChIP-seq data for 10 chromatin marks (CTCF, H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K9me3, H3K27me3, H3K9ac, H3K27ac, and H4K20me1) and input were obtained for seven different cell types (GM12878, H1-hESC, HMEC, HSMM, KS62, NHEK, and NHLF) from ENCODE Broad Histone (Ernst et al. 2011). DNase-seq and FAIRE-seq data were obtained from ENCODE UNC/Duke (Song et al. 2011). Polya tailed RNA-seq data were obtained for 11 different cell types (GM12878, H1-hESC, HCT-116, HaLa-S3, HepG2, HSMM, HUVEC, KS62, MCF-7, NHEK, NHLF) from ENCODE Caltech RNA-seq. GenBank mRNAs and ESTs data were downloaded from the UCSC Genome Browser (Fujita et al. 2011) on January 1, 2012. These data were mapped to the human genome to which the DJ sequences had been added (Supplemental Methods).

Nucleolar ChIP and FAIRE
ChIP was performed on nucleolar chromatin isolated from HT1080 cells with H4K4me3 antibodies (Millipore, catalog no. O-4745) as described previously (Mals et al. 2005). We adapted a FAIRE protocol (Giresi et al. 2007) for cross-linked nucleolar chromatin. One hundred microliters of HT1080 nucleolar chromatin was extracted using an equal volume of phenol/chloroform. DNA was recovered from the aqueous phase by ethanol precipitation and resuspended in 100 µL of TE buffer. PCR was performed using 2 µL of recovered DNA. DNA recovered from input nucleolar chromatin served as a control for ChIP and FAIRE experiments.

Transcriptome profiling
Paired-end RNA-seq data from the 11 cell types was mapped to the human genome with DJ sequences added using TopHat (v1.2.0) (Trapnell et al. 2009) with mostly default parameters (+ 50 -a 8). We then merged the output alignments from all replicates using SAMtools (Li et al. 2009). Paired-end RNA-seq data from all replicates was merged using Cuffmerge (Trapnell et al. 2010) to obtain the final transcriptome. We also used this final transcriptome to estimate DJ transcript abundance. We used BLAT (Kent 2002) to map the mRNA and EST data to the DJ using the parameters “-fine –q=rna -minid=95 -maxIntron=70000”, and “-minid=97 -maxIntron=70000”, respectively.

Data access
The DJ contig nucleotide sequence and feature list are available in Supplemental Data 2 and 3. Assembled DJ and PJ cosmId sequences are available from GenBank (http://www.ncbi.nlm.nih.gov/genbank) under accession numbers KC876024–KC876030. Raw NGS data have been submitted to the Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/Traces/sra/) under accession number SRP024282.

Acknowledgments
B.M. acknowledges Science Foundation Ireland (PI grant 07/IN.1/B924) for funding work in his laboratory. A.R.D.G. acknowledges the Auckland Medical Research Foundation, the Marsden Fund, the Royal Society of NZ, and the Bio-Protection Research Centre for funding. C.S. thanks Science Foundation Ireland (07/SP/M12111b) for funding. T.T.N. received a PhD fellowship from the Irish Research Council for Science, Engineering and Technology. We also thank Samantha Raggett, Jane Wright, and Mark Anderson for help in library screening and DNA sequencing during the early stages of this work, and Kevin Sullivan and Carol Duffy for comments on the manuscript.

References