2011

LOSS OF BLOOM SYNDROME PROTEIN CAUSES DESTABILIZATION OF GENOMIC ARCHITECTURE AND IS COMPLEMENTED BY ECTOPIC EXPRESSION OF Escherichia coli RecG IN HUMAN CELLS

Michael Wayne Killen
University of Kentucky, michael.killen@gmail.com

Recommended Citation
Killen, Michael Wayne, "LOSS OF BLOOM SYNDROME PROTEIN CAUSES DESTABILIZATION OF GENOMIC ARCHITECTURE AND IS COMPLEMENTED BY ECTOPIC EXPRESSION OF Escherichia coli RecG IN HUMAN CELLS" (2011). University of Kentucky Doctoral Dissertations. 201.
https://uknowledge.uky.edu/gradschool_diss/201

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsvaky.edu.
ABSTRACT OF DISSERTATION

Michael Wayne Killen

The Graduate School
University of Kentucky
2011
LOSS OF BLOOM SYNDROME PROTEIN CAUSES DESTABILIZATION OF GENOMIC ARCHITECTURE AND IS COMPLEMENTED BY ECTOPIC EXPRESSION OF *Escherichia coli* RecG IN HUMAN CELLS

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Medicine, Department of Microbiology, Immunology, and Molecular Genetics at the University of Kentucky

By
Michael Wayne Killen
Lexington, Kentucky

Director: Dr. Andrew J. Pierce, Assistant Professor, Department of Microbiology, Immunology, and Molecular Genetics
Lexington, Kentucky
2011

Copyright© Michael Wayne Killen 2011
ABSTRACT OF DISSERTATION

LOSS OF BLOOM SYNDROME PROTEIN CAUSES DESTABILIZATION OF GENOMIC ARCHITECTURE AND IS COMPLEMENTED BY ECTOPIC EXPRESSION OF Escherichia coli RecG IN HUMAN CELLS

Genomic instability driven by non-allelic homologous recombination (NAHR) provides a realistic mechanism that could account for the numerous chromosomal abnormalities that are hallmarks of cancer. We recently demonstrated that this type of instability could be assayed by analyzing the copy number variation of the human ribosomal RNA gene clusters (rDNA). Further, we found that gene cluster instability (GCI) was present in greater than 50% of the human cancer samples that were tested. Here, data is presented that confirms this phenomenon in the human GAGE gene cluster of those cancer patients. This adds credence to the hypothesis that NAHR could be a driving force for carcinogenesis. This data is followed by experimental results that demonstrate the same gene cluster instability in cultured cells that are deficient for the human BLM protein. Bloom’s Syndrome (BS) results from a genetic mutation that results in the abolition of BLM protein, one of human RecQ helicase. Studies of Bloom’s Syndrome have reported a 10-fold increase in sister chromatid exchanges during mitosis which has primarily been attributed to dysregulated homologous recombination. BS also has a strong predisposition to a broad spectrum of malignancies. Biochemical studies have determined that the BLM protein works in conjunction with TOPOIIIα and RM11/RM12 to function as a Holliday Junction dissolvase that suppress inadvertent crossover formation in mitotic cells. Because of the similarities in their biochemical activities it was suggested that another DNA helicase found in E. coli, the RecG DNA translocase, is the functional analog of BLM. RecG shares no sequence homology with BLM but it can complement both the sister chromatid exchange elevation and the gene-cluster instability phenotype caused by BLM deficiency. This indicates that the physiological function of BLM that is responsible for these phenotypes rests somewhere in the shared biochemical activities of these two proteins. These data taken together give new insights into the physiological mechanism of BLM protein and the use of Bloom’s Syndrome as a model for carcinogenesis.

KEYWORDS: Bloom’s Syndrome, BLM, homologous recombination, genomic instability, cancer

Michael Wayne Killen

April 21, 2011
LOSS OF BLOOM SYNDROME PROTEIN CAUSES DESTABILIZATION OF GENOMIC ARCHITECTURE AND IS COMPLEMENTED BY ECTOPIC EXPRESSION OF *Escherichia coli* RecG IN HUMAN CELLS

By

Michael Wayne Killen

Andrew J. Pierce, Ph.D.________________
Director of Dissertation

Beth A. Garvy_______________________
Director of Graduate Studies

April 21, 2011
Date
RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name_________  Date_________

_________________________________________________________________________

_________________________________________________________________________

_________________________________________________________________________

_________________________________________________________________________

_________________________________________________________________________

_________________________________________________________________________
LOSS OF BLOOM SYNDROME PROTEIN CAUSES DESTABILIZATION OF GENOMIC ARCHITECTURE AND IS COMPLEMENTED BY ECTOPIC EXPRESSION OF *Escherichia coli* RecG IN HUMAN CELLS

---

DISSEMINATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Medicine at the University of Kentucky

By:
Michael Wayne Killen
Lexington, Kentucky

Director: Dr. Andrew J. Pierce, Assistant Professor, Department of Microbiology, Immunology, and Molecular Genetics Lexington, Kentucky 2011

Copyright© Michael Wayne Killen 2011
ACKNOWLEDGMENTS

I would like to thank my mentor and Dissertation Chair, Dr. Andrew J. Pierce, who has provided guidance and support over the last 5 years. Without his diligence, high expectations, and confidence in me this body of work would not exist. My time in his lab was one of the most edifying experiences of my life and I am a better scientist and person because of him.

I would also like to thank the members of my committee Dr. Charlie Lutz, Dr. Guo-min Li and Dr. Brett Spear my Co-Chair, for their guidance and feedback and encouraging me along the way. They are all busy individuals and I appreciate the time they made for me. I would also like to thank Dr. Udeni Balasuriya, my outside examiner for his thoughtful comments.

Lastly, if not for my friends and family none of this would have been possible. I want to thank my parents Jim and Karen Killen who taught me everything is possible and who taught me to believe in myself because they always would. It is because of them that I am who I am today. Most importantly, I would like to thank my wife, Sara. She is my heart. Without her unyielding faith in me and wonderful smile I would not have made it the first day.
## TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ........................................................................................................ iii
**LIST OF TABLES** ................................................................................................................ vi
**LIST OF FIGURES** ............................................................................................................... vii
**LIST OF ABBREVIATIONS** ............................................................................................... viii

**CHAPTER 1: BACKGROUND AND INTRODUCTION**

1.1: GENOMIC INSTABILITY AND CARCINOGENESIS
   1.1.1: CARCINOGENESIS ........................................................................................ 1
   1.1.2: SOURCES OF GENOMIC INSTABILITY ...................................................... 2

1.2: DNA REPAIR
   1.2.1: FAITHFUL MAINTENANCE OF THE GENOME .......................................... 3
   1.2.2: DOUBLE STRAND BREAK REPAIR ............................................................ 4
   1.2.3: REPAIR AT THE REPLICATION FORKS .................................................... 5

1.3: HOMOLOGOUS DIRECTED REPAIR
   1.3.1: HOMOLOGOUS RECOMBINATION ............................................................ 7
   1.3.2: GENE CONVERSION (GC) VS. NON-ALLELIC HOMOLOGOUS RECOMBINATION (NAHR) ..................................................................................... 8

1.4: GENE CLUSTER INSTABILITY AND CANCER
   1.4.1: GENE CLUSTERS .......................................................................................... 12
   1.4.2: THE HUMAN RIBOSOMAL RNA GENE CLUSTER (RDNA) .................. 13
   1.4.3: THE HUMAN GAGE GENE CLUSTER ....................................................... 14

1.5: BLOOM’S SYNDROME
   1.5.1: BLOOM’S PATIENTS AND CANCER ......................................................... 15
   1.5.2: CLINICAL MANIFESTATIONS ................................................................... 16

1.6: BLOOM’S SYNDROM PROTEIN (BLM)
   1.6.1: BIOCHEMISTRY OF RECQ HELICASES ................................................... 19
   1.6.2: BIOCHEMISTRY OF BLM ............................................................................ 20

1.7: RECG HELICASE PROTEIN
   1.7.1: BIOCHEMISTRY OF RECG HELICASE ..................................................... 22
   1.7.2: BIOCHEMICAL SIMILARITY OF RECG AND BLM ............................... 25

1.8: REFERENCES ............................................................................................................. 26

**CHAPTER 2: GENE CLUSTER INSTABILITY ASSAY**

2.1: SYNOPSIS ................................................................................................................ 36
2.2: INTRODUCTION ...................................................................................................... 36
2.3: MATERIALS ............................................................................................................ 39
2.4: METHODS ............................................................................................................... 45
2.5: NOTES ...................................................................................................................... 57
2.6: FIGURES .................................................................................................................. 60
2.7: REFERENCES ......................................................................................................... 63
CHAPTER 3: CONFIGURATION AND REARRANGEMENT OF THE HUMAN GAGE GENE CLUSTERS
3.1: SYNOPSIS................................................................................................................ 64
3.2: INTRODUCTION ....................................................................................................... 64
3.3: MATERIALS/METHODS ....................................................................................... 66
3.4: RESULTS ................................................................................................................. 67
3.5: CONCLUSIONS ....................................................................................................... 70
3.6: FIGURES.................................................................................................................. 71
3.7: REFERENCES ......................................................................................................... 76

CHAPTER 4: LOSS OF BLOOM SYNDROME PROTEIN DESTABILIZES HUMAN GENE CLUSTER ARCHITECTURE
4.1: SYNOPSIS................................................................................................................ 79
4.2: INTRODUCTION ....................................................................................................... 80
4.3: RESULTS ............................................................................................................... 82
4.4: DISCUSSION ......................................................................................................... 88
4.5: METHODS ............................................................................................................. 91
4.6: FIGURES................................................................................................................ 94
4.7: REFERENCES ....................................................................................................... 103

CHAPTER 5: Escherichia coli RECG IS A FUNCTIONAL ANALOG OF BLM, THE HUMAN BLOOM SYNDROME PROTEIN
5.1: SYNOPSIS.............................................................................................................. 107
5.2: INTRODUCTION .................................................................................................. 108
5.3: RESULTS ............................................................................................................... 110
5.4: DISCUSSION ......................................................................................................... 112
5.5: METHODS ............................................................................................................. 114
5.6: FIGURES................................................................................................................ 117
5.7: REFERENCES ....................................................................................................... 125

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS ............................................. 129

REFERENCES .................................................................................................................... 137

VITA.................................................................................................................................... 151
LIST OF TABLES

Table 2.1. Adobe Photoshop and the following algorithm ................................. 57
Table 5.1: The effect of RecG expression on SCEs per Chromosome ................. 119

LIST OF FIGURES

Figure 1.1. Replication fork repair ................................................................. 7
Figure 1.2. Pathway of double-stranded break repair ........................................ 11
Figure 1.3. Sister chromatid of BLM deficient cells line ................................... 17
Figure 1.4. Individuals affected with BLM syndrome ....................................... 18
Figure 1.5. The RecQ family of DNA helicases ............................................... 20
Figure 1.6. Graphic depiction of BLM binding double Holliday junctions .......... 22
Figure 1.7. Processing of damaged replication forks by RecG ......................... 23
Figure 1.8. Domain structure of RecG from *Thermatoga maritima* ................ 24
Figure 2.1: Schematic of the restriction digestion procedure for liberating intact gene clusters from bulk genomic DNA ....................................................... 60
Figure 2.2: Schematic of expected GCI results ................................................ 60
Figure 2.3: GCI analysis of U2OS cells ............................................................. 61
Figure 2.4: GCI analysis of U2OS clonal cell lines .......................................... 62
Figure 3.1: Human GAGE gene clusters ......................................................... 71
Figure 3.2: GAGE clusters in families and identical twins ............................... 72
Figure 3.3: GAGE clusters in adult solid tumors ............................................. 73
Figure 3.4: GAGE clusters in pediatric leukemia ............................................. 74
Figure 3.5: GAGE clusters in Rothmund-Thomson kindred and probands ......... 75
Figure 4.1: Screening cell lines for gene cluster instability (GCI) ....................... 94
Figure 4.2: GCI clonal analysis ................................................................. 95
Figure 4.3: Bloom syndrome GCI ............................................................. 96
Figure 4.4: GCI in isogenic BLM and BLM-complemented human cells .......... 97
Figure 4.5: BLM shRNA expressing clonal lines ......................................... 98
Figure 4.6: Werner syndrome clonal lines ...................................................... 99
Figure 4.7: HeLa versus ATM-deficient clonal analysis ..................................................... 100
Figure 4.8: CSB-deficient and wild-type versus DNA-ligase IV knockout cell clonal analysis with BLM cDNA................................................................. 101
Figure 4.9: SCE analysis of BLM cells complemented with BLM cDNA......................... 102
Figure 5.1: RecG expression constructs................................................................. 117
Figure 5.2: RecG expression complements the BLM elevated SCE phenotype .......... 118
Figure 5.3: RecG expression reduces SCE in a French-Canadian BLM line ............... 120
Figure 5.4: RecG expression has no effect on SCE levels in wild-type cells.............. 121
Figure 5.5: Loss of RecG expression restores the elevated SCE phenotype .............. 122
Figure 5.6: RecG stabilizes the gene cluster instability of BLM cells......................... 123
Figure 5.7: RecG has no effect on gene cluster instability in wild-type cells .......... 124
LIST OF ABBREVIATIONS

BER: Base excision repair
BIR: Break induced replication
BLM: Bloom's syndrome helicase
BLM: Bloom syndrome protein
BS: Bloom's syndrome
BS: Bloom syndrome patients
CIN: Chromosome instability
CML: Chronic Myeloid Leukemia
CNV: Copy number variation
dHJ: Double Holliday junction
DSB: Double strand break
DSBR: Double strand break repair
DSBR: Double strand break repair
GC: Gene conversion
GCI: Gene cluster instability
GCI: Gene cluster instability
HDR: Homology directed repair
HJ: Holliday junctions
HNPCC: Hereditary Non-polyposis Colorectal Cancer
HR: Homologous recombination
IFN-γ: Interferon gamma
IG: Immunoglobulin
IRF1: Interferon regulatory factor
IRIs: Intermolecular recombination intermediates
LCR: Low copy repeats
LMP: Low melting point
LOH: Loss of heterozygosity
MEM: Minimal essential medium
MIN: Microsatellite instability
MMR: Mismatch repair
NAHR: Non-allelic homologous recombination
NER: Nucleotide excision repair
NHEJ: Non-homologous end joining
NOR: Nucleolar organizing region
PFC: Pulse field certified
PFGE: Pulse field gel electrophoresis
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMSF</td>
<td>Phenylmethanesulfonylfluoride</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>RTS</td>
<td>Rothmund-Thomson syndromes</td>
</tr>
<tr>
<td>SCE</td>
<td>Sister chromatid exchange</td>
</tr>
<tr>
<td>SCEs</td>
<td>Sister chromatid exchanges</td>
</tr>
<tr>
<td>SEER</td>
<td>Surveillance Epidemiology and End Results</td>
</tr>
<tr>
<td>SSA</td>
<td>Single strand annealing</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single stranded DNA</td>
</tr>
<tr>
<td>STRs</td>
<td>Short tandem repeats</td>
</tr>
<tr>
<td>WS</td>
<td>Werner's syndrome</td>
</tr>
</tbody>
</table>
CHAPTER 1: BACKGROUND AND INTRODUCTION

1.1 GENOMIC INSTABILITY AND CARCINOGENESIS.

1.1.1 CARCINOGENESIS

Cancer arises from aberrant gene expression typically thought to develop in a lengthy multi-step process that leads to the accumulation of gene mutations at different loci resulting in a gain or loss of function and leads to uncontrolled cellular growth (Sherr 2004) (Jefford and Irminger-Finger 2006). Although, the accumulation of single mutations is sufficient for carcinogenesis, the single most common trait of all cancer cells is an altered karyotype and gross genomic rearrangements (Murga and Fernandez-Capetillo 2007; Calasanz and Cigudosa 2008). Over the past few years cytogenetic analysis of cultured tumor cells has revealed that most solid tumors have aneuploid karyotypes, aberrant nuclei with altered chromosome numbers and structure (Lengauer, Kinzler et al. 1998; Sieber, Heimann et al. 2003). Although it has been debated whether genomic instability is a cause or consequence of tumorgenesis, recently the role of chromosome instability was demonstrated in tumor progression (Nowak, Komarova et al. 2002; Barbouti, Stankiewicz et al. 2004; Bruder, Piotrowski et al. 2008). Furthermore, recent publications have suggested SOS like models in which gross genomic instability would play a key role in the mutator phenotype of cells in tumorgenesis (Komarova, Sadovisky et al. 2008). Considering the rarity of mutations in normal cells and the large numbers of mutations observed in human cancers, it has been proposed that the spontaneous mutation rate in normal cells is not sufficient to account for the number of mutations found in human cancers (Loeb, Springgate et al. 1974). It has also been suggested that the multiple mutations found in tumor cells would result from mutations in genes that guarantee the fidelity of DNA synthesis and repair (Loeb, Loeb et al. 2003). Mutations in genes that function in the maintenance of genomic stability would manifest by increases in mutation rates and could drive tumor progression. The fact that the majority of chromosomal abnormalities are not tumor specific supports this underlying mechanism of instability in carcinogenesis (Loeb 2001). Thus it is important to explore genes that provide for maintenance of genomic stability as a source of gross genomic changes and determine if they could be a driving force of carcinogenesis.
1.1.2 SOURCES OF GENOMIC INSTABILITY

The fact that the sequence and configuration of DNA in cancer cells is typically very different from that of normal cells provides some indication that carcinogenesis involves substantial errors in DNA replication, deficits in DNA repair, and/or alterations in chromosomal segregation (Loeb 2001; Komarova, Sadovsky et al. 2008). The molecular components that maintain genomic stability can be broadly classified into two groups: pathways that maintain the fidelity of DNA replication via repair and the pathways that control cell division. The DNA repair pathways are responsible for keeping genes and genomes intact as well as maintaining chromatin epigenetic modifications, whereas the pathways that control cell division help maintain mitotic stability and chromosome integrity via modulation of sister chromatid cohesion and chromosome migration (Jefford and Irminger-Finger 2006). Genomic instability can be broadly classified into two types: microsatellite instability (MIN) associated with a mutator phenotype resulting in destabilization of DNA sequence and chromosome instability (CIN) which involves continuous gross alterations in chromosome number and structure. Both types of instability can lead to egregious consequences and there are many examples of both with very distinct correlations with carcinogenesis discussed below.

It is believed that changes in nucleotide sequence are in some cases sufficient for carcinogenesis. It is well documented that single nucleotide mutations of genes such as K-Ras, p53, and BRCA1/2 directly correlate with substantial increase of cancer risk (Almoguera, Shibata et al. 1988; Rahman and Stratton 1998; Hainaut and Hollstein 2000). MIN which results in the destabilization of nucleotide sequences is the cause of hereditary non-polyposis colorectal cancer (HNPCC). Although this can be caused by a mutation of a single gene, like hMSH2, the loss of this single gene causes the malfunction of the mismatch repair system responsible for repairing mistakes made by the polymerase during DNA synthesis (Aaltonen, Peltomaki et al. 1993; Li 2008). The malfunction of the MMR system as a whole leads to genomic instability and increase mutation rates.
Despite its direct link to cancer, MIN has not been linked to some of the most common traits found in tumors such as changes in chromosome number, translocations, amplifications, and deletions. These characteristics fall under the realm of CIN (Jefford and Irminger-Finger 2006). As discussed previously, abnormal ploidy is the most notable characteristic of all cancer types. Common examples of translocations causing cancer are found in Burkett’s lymphoma and the Philadelphia chromosome in chronic myeloid leukemia (CML). In Burkett’s lymphoma the myc oncogene is translocated to the locus where an immunoglobulin (Ig) enhancer resides. This leads to a drastic increase of Myc expression in lymphocytes (Dean, Kent et al. 1983). In a similar fashion the Philadelphia chromosome is due to the translocation of the Bcr gene to a locus where it is fused with the Abl kinase. The resulting hybrid gene produces a protein that is a constitutively active tyrosine kinase receptor and leads to uncontrolled cell proliferation (Jeffs, Benjes et al. 1998). Amplifications can increase the copy number of proto-oncogenes whereas deletions can eliminate genetic material that activates proto-oncogenes or lead to a loss of tumor suppressor genes. However, a common mechanism for how these abnormalities arise has not been established. One mechanism that has been proposed is dysfunctional repair on DNA by Homologous Recombination (HR) (Bishop and Schiestl 2003; Reliene, Bishop et al. 2007). While HR is integral to maintaining genomic stability, it can result in gross genomic restructuring that changes millions of base pairs at a time. The mechanism that can cause this gross chromosome instability is the subject of this dissertation.

1.2 DNA REPAIR

1.2.1 FAITHFUL MAINTENANCE OF THE GENOME

The maintenance of species requires that information contained in DNA be accurately copied and transmitted. The fidelity with which normal human cells accurately replicate DNA every time they divide is exceptional, the overall mutation rate in somatic human cells can be measured and has been estimated to be $2.0 \times 10^{-7}$ mutations/gene/cell division (DeMars and Held 1972). With this in mind it can be estimated that every cell
will accumulate at least one mutant gene during the lifespan of an individual (Loeb, Springgate et al. 1974). Therefore a mechanism through which these errors might be corrected is critical for the propagation of all organisms. Aside from mutations acquired during replication the cell can encounter a plethora of endogenous and exogenous agents that require repair to maintain the integrity of the genome. Because each of these agents can cause very different lesions the cell has evolved a variety of mechanisms to repair DNA damage. These include nucleotide excision repair (NER), base excision repair (BER), translesion synthesis, mismatch repair (MMR), homologous recombination (HR), non-homologous end joining (NHEJ) and others. There are diseases associated with deficiencies in all of these pathways; however, as mentioned before the subject of this dissertation is homologous recombination and how the malfunction of this repair pathway might lead to gross alterations in genomic architecture and content.

1.2.2 DOUBLE STRAND BREAK REPAIR (DSBR)

One of the most severe lesions that can be introduced into the genome is a DNA double strand break (DSB). This damage can be induced by various agents: endogenous factors such as free radicals formed as byproducts of metabolic functions, or stalled replication forks, or exogenous factors such as ionizing radiation or mutagenic chemicals. DSBs are often represented as clean cuts made in both strands of the DNA by restriction enzymes. However, they are more likely much more complex on a molecular level. In this discussion a DSB represents any gap or break found in both strands of the DNA leaving the chromosome discontinuous.

The importance of the DSB repair pathway for preserving genomic integrity is underscored by its conservation throughout evolution (Sonoda, Hochegger et al. 2006). It has been shown that one unrepaired DSB can potentially be lethal to a cell (Frankenberg-Schwager and Frankenберg 1990). Genomic instability arising from the illegitimate repair of DNA DSBs is subject of intense investigation. The human cell utilizes five known pathways to repair DSBs. There is the homology independent pathway, non-homologous end joining (NHEJ). It is worth noting that there is some evidence that
microhomologies in the sequence flanking the DSB may play a role in this mechanism (Yu, Marshall et al. 2004). Then there is homology-directed repair (HDR) which includes single strand annealing (SSA) and homologous recombination (HR). HR can be further broken down into gene conversion (GC), which is HR without associated crossover, and non-allelic homologous recombination (NAHR), which includes associated crossover. Another possible result is break-induced replication (BIR) which results in non-reciprocal crossover and occurs when synthesis continues on the donor strand until the end of the chromosome or a centromere is reached. Although BIR has recently been reported to play a role in meiotic variation in humans it is not known if it plays a role in somatic cells (Smith, Llorente et al. 2007; Zhang, Khajavi et al. 2009; Koumbaris, Hatzisevastou-Loukidou et al. 2011). Most of these pathways cooperate to ensure cell survival. If one pathway fails the other can be activated (Couedel, Mills et al. 2004) (Mills, Ferguson et al. 2004). All of the pathways have genetic consequences. NHEJ can lead to loss of the sequence surrounding the break, although this tends to localize damage and has far fewer consequences than an unrepaired DSB. NHEJ is most common form of repair utilized in human cells (Richardson and Jasin 2000). But GC can be the most faithful mechanism of repair and is often referred to as error free since it fills in the missing sequence by copying it from homologous sequence elsewhere in the genome. While this opens the door for loss of heterozygosity (LOH) it seems trivial in light of the consequences of NAHR. NAHR is not thought to occur at a high frequency and homologous recombination in vertebrate DNA repair is typically error-free without crossover (Johnson and Jasin 2000; Richardson and Jasin 2000).

1.2.3 REPAIR AT REPLICATION FORKS

For many years the primary role of homologous recombination had been assumed to be creating genetic diversity during meiosis. In recent years there has been a shift to include the indispensable role that HR plays in the stabilization and recovery of collapsed and arrested replication forks (Bernstein, Byerly et al. 1985; Michel, Grompone et al. 2004). As discussed above endogenous and exogenous agents can also lead to dysfunction of the replication machinery. The lesions caused by these agents can lead to
single strand breaks in DNA. When the replication fork encounters these breaks it can lead to a DSB and a “collapsed” fork (Kuzminov 2001). Additionally, these lesions can stall replication on the leading or lagging strand and lead to uncoupling of the DNA replication machinery eventually blocking the progress of the replication fork without generating a DSB. This is typically referred to as a stalled replication fork (Cox 1999). The consequences of stalling are that the template strands can start to reanneal and migrate backward; the newly synthesized strands are displaced and also anneal creating a Holliday junction. This type of junction is sometimes referred to as a “chicken foot structure” because of its topological appearance. In addition to lesions causing the replication fork to stall and generate Holliday junctions, it has been observed that the superhelical tension created by normal replication can itself stall replication and cause the replication fork to regress to the chicken foot structure (Postow, Ullsperger et al. 2001; Olavarrieta, Martinez-Robles et al. 2002). These topological structures are resolved through the HR repair pathway and the formation of these intermediates as well as the potential for generation of DSB provides an important role for recombination in the stabilize replication forks (Bernstein, Byerly et al. 1985; Kuzminov 1999). It also provides the potential for illegitimate resolution of these intermediate substrates that can result in mega-base scale genomic restructuring.
Figure 1.1. Replication fork repair. A lesion blocks leading strand synthesis leading to fork collapse. Lagging strand synthesis continues for a short period. The fork regresses, allowing the nascent strands to anneal, creating a ‘chicken foot’ structure (Holliday junction). The leading strand is then extended using the longer lagging strand as a template. The fork is reset by branch migration of the four-way junction, and now the leading strand has been extended beyond the lesion. Adapted from (Bachrati and Hickson 2003) used with permission.

1.3 HOMOLOGY DIRECTED REPAIR (HDR)

1.3.1 HOMOLOGOUS RECOMBINATION

The genome is remarkably stable, however, without plasticity or a means to introduce variation into the genome, evolution and natural selection would not happen (Jefford and Irminger-Finger 2006). The majority of this variation arises during meiosis as a product of recombination. It is this mechanism that gives organisms their most efficient means for enriching advantageous mutations. However it is also this mechanism that could
produce the genetic instability from which cancer arises, a consequence that allows for the selection of viable mutations that favor proliferation (Cahill, Kinzler et al. 1999). Homologous recombination provides a means of copying genetic sequence from a homologous template located elsewhere in the genome and allows a high degree of fidelity to be maintained in repairing DNA damage. Unfortunately, since the template and the repaired locus are intertwined during HR they must be enzymatically separated. It is this process that is thought to distinguish GC and NAHR and determine if the repair is faithful or if it results in the alteration of the genomic architecture. The abundance of repetitive sequence found in the genome also makes dysregulated homologous recombination potentially hazardous. An inappropriate template might allow for homologous recombination (HR) to rearrange thousands, even millions of base pairs leading to deletions, amplifications, and translocations. This can alter massive amounts of genetic material in a single event. It is this illegitimate means of homologous repair that is believed to have the most potential for driving carcinogenesis. Although the majority of DSB repairs proceed through NHEJ and GC, NAHR does rarely occur and several lines of evidence suggest that there is a direct link between NAHR and carcinogenesis (Johnson and Jasin 2000; Richardson and Jasin 2000; Bishop and Schiestl 2003). Yet even an isolated event could have calamitous consequences (Bishop and Schiestl 2003). This possibility of genome instability caused by a single event provides a leadway into the hypothesis that NAHR could change a multitude of gene expression patterns and be one of the most efficient mechanisms through which carcinogenesis might proceed.

1.3.2 GENE CONVERSION (GC) vs. NON-ALLELIC HOMOLOGOUS RECOMBINATION (NAHR)

During HR new DNA is synthesized at both ends of the DSB by copying both strands of the homologous template extending past the site of the broken DNA. This intertwining of the DNA creates two separate Holliday junctions (HJ) or a double Holliday junction (dHJ). The double Holliday junction presents a very interesting topological problem for the cell. Distinct molecular mechanisms exist to solve this
problem. The cell can “resolve” the double Holliday junction by enzymatic cleavage that makes two cuts in the junction and results in an exchange of genetic material between the two loci resulting in crossover exchange. Alternatively, the cell can “dissolve” the junction such that the resulting structure does not allow for crossover (Wu and Hickson 2003). It is this junction that determines if the HR proceeds with GC or NAHR. If the dHJ is dissolved the donor sequence is left unchanged, it is simply used as a template to repair the discontinuous strand. This is termed gene conversion (GC). In order for gene conversion to prevail it is necessary for the cell to resolve the topological problem in a fashion that distinguishes the donor strand from the recipient and prevents the illegitimate recombination typically attributed to a cross-over reaction. Although, this process is not well characterized in mammals, the biochemistry suggests that the gene products involved in “dissolving” double Holliday junctions are BLM, TOPOIIIα and RMI1/RMI2 (Wu and Hickson 2003; Ralf, Hickson et al. 2006; Mankouri and Hickson 2007). The biochemistry of this reaction is discussed extensively in section 1.4.

Another potential outcome of HR is non-allelic homologous recombination (NAHR). NAHR is distinguished from GC because it is resolved through a cross-over reaction between the two homologous sequences and it changes the chromosomal context of both sequences, where the sequences are not alleles on homologous chromosomes or sister chromatids. Here there is a physical exchange between the donor sequence and the repaired recipient and possibly the flanking sequence. NAHR can lead to the deletion, amplification, or the translocation of millions of base pairs. These large scale genomic alterations in meiosis provide many examples of genomic disorders that cause human disease. However, its prevalence to mitosis is unknown and its potential contribution to carcinogenesis is unestablished, but structural rearrangements of chromosomes due to repair with crossing-over have been observed (Ira, Malkova et al. 2003; Stults, Killen et al. 2008). Unlike GC, the reaction resulting in NAHR does not distinguish the donor strand from the recipient in order to prevent illegitimate recombination. Here the enzymatic resolution of the double Holliday junction is thought to proceed through GEN1 or the S1x1-S1x4 complex (Ip, Rass et al. 2008; Svendsen, Smogorzewska et al. 2009; Svendsen and Harper 2010). These nucleases are thought to symmetrically cleave HJs in
a manner that allows for separation of the strands without additional processing, thus acting as a Holliday junction “resolvase”. As such, the complex cleaves the links between two homologous chromosomes that form during homologous recombination. This allows the two linked chromosomes to resolve into two unconnected double-strand DNA molecules (Svendsen, Smogorzewska et al. 2009; Svendsen and Harper 2010).

It is unknown how and why the cell makes the “decision” between GC and NAHR to pursue one repair pathway over the other. However, the two pathways compete for the same repair intermediate and have their own genetic consequences (Fukushima, Takata et al. 2001; Pierce, Hu et al. 2001; Allen, Kurimasa et al. 2002). Homology-dependent pathways can be responsible for 30-40% of all repair of DSB (Liang, Han et al. 1998; Takata, Sasaki et al. 1998; Johnson and Jasin 2000; Pierce, Hu et al. 2001). The majority of homologous repair in mammals results in gene conversions without cross-over (Moynahan and Jasin 1997; Richardson, Moynahan et al. 1998). The preference for homologous sequence used in homologous repair is one thousand times higher for sequence located on sister chromatids than the homologous chromosome and ten thousand times higher than a heterologous chromosome (Johnson and Jasin 2001). One of the main factors in determining the involvement of HR is the temporal and spatial limitation placed on homologous sequences. For instance, a sister chromatid or homologous chromosome is typically only available in late S phase and G2 (Takata, Sasaki et al. 1998; Rothkamm, Kruger et al. 2003). When HR does occur in mitotic cells the majority of the time it results in gene conversion without cross-over but occasionally mitotic recombination does occur with associated cross-over and recent studies in our lab have indicated that it is the most prevalent chromosomal alterations in adult solid tumors (Ira, Malkova et al. 2003; Stults, Killen et al. 2008; Stults, Killen et al. 2009).
Recombination is initiated by nucleolytic processing of the DSB to generate 3’ single-stranded DNA overhangs that are rapidly coated with RPA. Rad51 recruitment displaces RPA leading to the formation of a helical nucleoprotein filament with which can search for homologous template and catalyze invasion of the ssDNA into the donor sequence to form a joint molecule. Repair proceeds to form crossover and non-crossover products. Adapted from (Aguilera and Boulton 2007).
1.4 GENE CLUSTER INSTABILITY AND CANCER

1.4.1 GENE CLUSTERS

A gene cluster is a set of two or more genes located in close proximity that encode for the same or very similar proteins. Gene clusters are created through the process of gene duplication (Inoue and Lupski 2002). The genes within the cluster then diverge or converge dependent on evolutionary pressure. For our purpose gene clusters will represent a set of duplicated genes that convergently evolved to maintain a high degree of similarity from one copy to the next. HR is thought to drive the process of gene duplication (Hastings, Lupski et al. 2009). HDR can provide a mechanism for gene duplication and divergence through NAHR and for convergence by GC. Human populations show extensive polymorphism in the numbers of copies of some genes (Consortium 2004; Hastings, Lupski et al. 2009). This is known as copy number variation (CNV). These repetitive sequences are estimated to account for up to 12% of the genome (Redon, Ishikawa et al. 2006). With such a substantial contribution to the content of the genome it seems evident that HR is a driving force behind meiotic variation on an evolutionary timescale. Such sequences are hotspots for meiotic non-allelic homologous recombination. When crossing-over occurs between homologous sequence in a non-homologous context, unequal crossing-over occurs and can results in duplication, deletion, translocation, or inversion of sequence (Shaw and Lupski 2004). Several heritable genetic syndromes are known to arise by this mechanism, including Charcot-Marie-Tooth syndrome, Prader-Willi syndrome and Angelman syndrome (Stankiewicz and Lupski 2002) (Stankiewicz and Lupski 2002). However, CNVs have also been reported in somatic derived mutations in monozygotic twins indicating the involvement of NAHR in genomic instability in the lifespan of an individual (Bruder, Piotrowski et al. 2008). It has also been shown that i(17q), a common structural abnormality in human neoplasias, likely arises by NAHR (Barbouti, Stankiewicz et al. 2004). By providing a mechanism for generating physical alterations in the genomic architecture, dysregulated homologous recombination between the many non-allelic repetitive sequences in the human genome can also have disastrous consequences for genomic stability (Consortium 2004; Lupski and Stankiewicz 2005). Since NAHR is
thought to provide a mechanism for CNV, analysis of these gene clusters might provide insight into the involvement of NAHR in carcinogenesis.

Our lab has developed an assay that evaluates gene cluster instability as a sentinel biomarker of dysregulated homologous recombination. The assay involves physical analysis of gene clusters in which the repeated genes have very high levels of sequence identity and are in very high local concentration, both factors in accelerating the rate at which they undergo recombination-mediated structural alteration. NAHR between similarly oriented repeats causes expansions and contractions of the number of repeats in the cluster. These length changes can be resolved through pulsed-field gel electrophoresis (PFGE) and Southern hybridization (Stults, Killen et al. 2008). Described in chapter 2 the Gene Cluster Instability (GCI) assay was developed for detecting spontaneous recombination-mediated genomic restructuring in human cells. We have successfully applied this technique to show that NAHR is commonly associated with human adult solid tumors (Stults, Killen et al. 2009).

1.4.2 THE HUMAN RIBOSOMAL RNA GENE CLUSTER (rDNA)

Large gene clusters possess attributes conducive to recombinational alteration. The repeated genes that make up clusters tend to have very high levels of sequence identity, sufficient length to be substrates for recombination, and occur in high local concentration with respect to each other. Each eukaryotic ribosome contains four RNA molecules which together do the catalytic work of protein synthesis. Multiple copies of the ribosomal RNA (rRNA) genes are required in order to generate sufficient rRNA molecules to accomplish all of the ongoing protein translation in a eukaryotic cell. They are transcribed by Pol I to form a single 45S transcript, which is subsequently spliced into the 18S, 28S, and 5.8S rRNAs (Bartova, Galiova et al. 2010). The gene clusters encoding the 45S precursor transcript (collectively the ‘rDNA’) are perhaps the largest clustered gene arrangement in the human genome (Henderson, Warburton et al. 1972). There are approximately 600 copies of the 43 kb gene repeats dispersed across the 5 human acrocentric chromosomes accounting for up to 0.5% of the genome (Gonzalez and
Sylvester 1995). The short arms of the acrocentric chromosomes which contain these genes are localized to the nucleolar organizing region (NOR) which isolates them so they can be transcribed and regulated separately from the rest of the genome (Henderson, Warburton et al. 1972).

Previously we used the GCI assay to characterize the genomic architecture of the human ribosomal RNA gene clusters, evaluating their heritability and stability in normal healthy individuals (Stults, Killen et al. 2008). We characterized the physical lengths of these rDNA gene clusters in human individuals and found essentially complete cluster length heterozygosity, both between clusters on different acrocentric chromosomes and also between clusters on individual parental homolog chromosome pairs (Stults, Killen et al. 2008). This cluster length heterogeneity is driven by strong meiotic recombination at a rate of over 10% per cluster per meiosis. We found that meiotic rearrangements of the clusters occur frequently and found evidence of mitotic NAHR with associated crossover suggesting that the gene cluster would be suitable as sentinel biomarkers for dysregulated recombination (Stults, Killen et al. 2008). We then used the human ribosomal RNA gene clusters to explore human genomic instability that arises from NAHR in human cancer samples (Stults, Killen et al. 2009). We found that over half of the adult solid tumors had detectable rDNA rearrangements relative to surrounding non-tumor tissue and normal peripheral blood. This indicates that rDNA restructuring is among the most common chromosomal alterations in adult solid tumors (Killen, Stults et al. 2009; Stults, Killen et al. 2009). This interrelatedness between this type of genomic instability and cancer is the basis for the work explored in the GAGE gene cluster in chapter 3.

1.4.3 THE HUMAN GAGE GENE CLUSTER

In chapter 3 I investigate genomic instability in the GAGE gene cluster in human cancer. GAGE is a member of the CT family of genes, defined as having expression only in cancer and in testis (Scanlan, Simpson et al. 2004). Like many of the protein products of the CT genes, GAGE is potentially immunoreactive. Since testis is an immunoprivileged site, this provides a potential protein based immunotherapeutic
strategies for cancer treatment (Machiels, van Baren et al. 2002). However, here the GAGE gene cluster simply serves as a substrate for GCI analysis. We used the GCI assay with blood samples from human families and clinical specimens from human cancers to determine the range of human variation of the GAGE cluster, and the stability of the cluster under normal human meiosis and in the genomic pathology of human cancers.

This series of experiments were performed to complement previous work that demonstrated mitotic instability in the rDNA gene clusters in human cancer (Stults, Killen et al. 2008; Stults, Killen et al. 2009). The rDNA clusters have high local concentration, high sequence similarity and multiple copies which make the rDNA clusters ideal substrates for HR. (Stults, Killen et al. 2008) However, as mentioned, the rDNA has some other characteristics that might contribute to its behavior that are not common to most genes. The rDNA is largely nucleolar and is transcribed by Pol I unlike the majority of genes (Russell and Zomerdijk 2006). The genomic structure of the GAGE gene cluster is similar in nature to the clustered rDNA. It has a high local concentration, high sequence similarity and multiple copies. However, the unit repeat of the gene that produces the GAGE transcript is only 9.6 kb and the tandemly repeated cluster is estimated to only have 15 gene copies on the X chromosome (Hofmann, Caballero et al. 2008). The unit repeats are oriented in a head-to-tail manner without any intervening sequences making them very similar to the rDNA genomic structure. However the GAGE gene cluster has significantly fewer unit repeats and is transcribed by Pol II which makes this locus more similar to gene clusters elsewhere in the genome (Warburton, Hasson et al. 2008). Thus the GAGE gene cluster is employed in this dissertation to model homologous sequence that more typical of the human genome.

1.5 BLOOM SYNDROME

1.5.1 BLOOM SYNDROME PATIENTS AND CANCER

Diseases that lead to genomic instability are typically marked by progeria and/or a high disposition to cancer. Several of these diseases result from mutations in genes that
encode DNA helicases in the RecQ family and include Werner’s (WS), Bloom’s (BS), and Rothmund-Thomson (RTS) syndromes. Bloom’s syndrome is a rare autosomal recessive disorder that displays a broad spectrum of cancer tissue types, representing a model of generalized cancer. However, the cancers tend to occur earlier in life and tend to be more aggressive (German and Passarge 1989; German 1997). Bloom syndrome displays the strongest known correlations between chromosomal instability and an increased risk of malignancy at an early age. Speculation that BLM-deficient cells escape apoptotic death by constitutively inducing a SOS-like response dependent on recombination has recently been raised (Amor-Gueret 2006). This hypothesis is consistent with the hyper-recombinatory phenotypes characteristic of BS cells and further illustrates the plausibility of a role for HR as a mechanism in carcinogenesis in the general population (Amor-Gueret 2006).

1.5.2 CLINICAL MANIFESTATION

Bloom’s syndrome (BS) was first described by dermatologist Dr. David Bloom in 1954 (Bloom 1954). Bloom syndrome is an autosomal recessive disorder characterized by growth deficiency; sun-sensitivity; telangiectasia; hypo- and hyper-pigmented skin; immunodeficiency associated with otitis media and pneumonia; predisposition to malignancy; and most importantly chromosomal instability (German, Bloom et al. 1984). BS comes from a genetic mutation that results in the total loss in production of the Bloom’s Syndrome Helicase (BLM) (Ellis and German 1996). BLM is a member of the RecQ family of helicases and is a 1417 amino acid protein that is produced from the BLM gene found on chromosome 15 (Ellis, Groden et al. 1995). BS cells display an increased frequency of crossing over between homologous chromosomes, chromosome breaks, quadraradial formation and high spontaneous mutation rates (Vijayalaxmi, Evans et al. 1983; Kyoizumi, Nakamura et al. 1989; Langlois, Bigbee et al. 1989). Consequently, BS cells have been classified as hyper-recombinatory and manifest with a 10-fold elevated frequency of sister chromatid exchange (SCE) (Vijayalaxmi, Evans et al. 1983). This can be seen in Figure 1.3 below. Clinical diagnosis of BS is now assayed by the molecular sequence of the BLM gene. However, until recent years this diagnosis was
derived through cytogenetic analysis of sister-chromatid exchanges (SCEs). This analysis was typically preformed on blood lymphocytes in short-term culture. Because of the 10-fold increase in SCEs, BLM cells are typically associated with increased homologous recombination. However, this exchange of genetic material between sister chromatids is generally regarded as genetically silent since the sister chromatids are genetically identical. SCE analysis is used as a companion assay with GCI in chapters 4 and 5.

Figure 1.3. Left: Sister chromatid analysis of wild-type cell line GM00637. Right: Sister chromatid analysis of cell line deficient for BLM protein GM08505. Notice the harlequin pattern distinguishing the high number of exchanges between sister chromatids in BLM cells.

The BS registry currently has 265 registered patients of which 150 which are still living. Of the 103 patients who are deceased the mean age of death is 26 years. The mean age of those living is 30 years (range from 1-53 years of age). The 265 persons with BS are from 222 families (siblingships). 167 (75.2%) of the affected families are/were non-Jewish. The remaining 55 (24.8%) are Jewish, which argues against the use of the term “Jewish genetic disease” which is typically used to describe BS. Of the 265 patients, 122 (46%) individuals have been affected by cancer. The average age of onset
for all cancers was 26 years (range from 1-49 years of age) (data adapted from BLM registry web site, http://www.med.cornell.edu/bsr/data_from_registry ). Based on data from the US National Cancer Institute's Surveillance Epidemiology and End Results (SEER) Database this is a 100-fold increase in cancer risk for Bloom’s Syndrome patient compared to normal individuals. The likelihood of being diagnosed with any type of cancer for both sexes and all races between the age of 20 and 30 is 0.45% compared to 46% at 26 years with BS as mentioned above (data adapted from http://seer.cancer.gov/csr/1975_2007/results_merged/topic_lifetime_risk.pdf, and is based on incidence and mortality data for the United States from 2005 through 2007). In figure 1.4 below affected individuals are depicted. The plate on the left shows an individual with a sun-sensitive response. The second plate on the right depicts an affected individual and their sibling demonstrating the aforementioned growth deficiency.

Figure 1.4. Left: Individual with butterfly rash resulting from sun-sensitive response. Right: Affected individual and sibling demonstrating proportional growth deficiency (images from http://weill.cornell.edu/bsr/).
1.6 BLOOM SYNDROME PROTEIN (BLM)

1.6.1 BIOCHEMISTRY OF RECQ HELICASES

The RecQ helicase was first identified in *E. coli* as a component of the RecF recombinational repair pathway. A null mutation at this locus confers a significant reduction in recombination frequency (~100-fold) and increases UV sensitivity (~20-fold) in a recBC/sbcBC background (Nakayama, Nakayama et al. 1984). The primary function of *E. coli* RecQ is to process linear double-strand DNA (dsDNA) substrates and provide a single-strand DNA (ssDNA) substrate for RecA protein. Thus, RecQ helicase can act as an initiator during the early steps of homologous recombination. *E. coli* RecQ can also disrupt nascent joint molecules in vitro (Harmon and Kowalczykowski 1998). RecQ helicase action, however, is not solely confined to an initiation role and displays a late role with its propensity to unwind the D-loops formed by the RecA protein. Because the RecQ helicase can also disrupt the homologous pairing products formed in these reactions it may have additional roles in DNA metabolism. However the RecQ protein is not a branch migration-specific helicase because, in contrast to the RuvAB and RecG helicases, it displays no greater helicase activity on a four-way junction DNA substrate than it does on other DNA substrates. In addition the RecQ helicase displays a broad DNA binding specificity at internal regions of duplex DNA, such as at ssDNA gaps which are believed to be mediated by RecF pathway proteins (Lloyd and Buckman 1985; Lanzov, Stepanova et al. 1991).

Five human homologs of RecQ have been identified: RECQL, BLM, WRN, RECQL4 and RECQL5. Of these five homologs, defects in BLM causes Bloom’s syndrome (BS), mutations in WRN leads to Werner syndrome (WS), and defects in RECQL4 cause one form of Rothmund-Thomson syndrome (RTS). No human diseases have been identified with defects in RECQL or RECQL5. All three human syndromes have an elevated cancer risk, with the risk associated with BS being the most pronounced. Alignment of the amino acid sequence of these proteins shows definite homology with bacterial RecQ. However, some of the human RecQs have specific and unique protein domains. For instance, the WRN protein contains an exonuclease domain. The biochemistry also
points to each member of the family having a unique function in maintaining the human genome, although some functions may overlap or substitute for some vital functions when another member is defective.

Figure 1.5. The RecQ family of DNA helicases. Selected members from various species are shown aligned by the conserved central helicase domain. Six regions of homology are defined. The central helicase domain contains seven helicase motifs is depicted in green and present in all cases. Also included are the RecQ C-terminal domain (red), the helicase and RNase D C-terminal (blue), conserved stretches of acidic amino acids (purple), and the nuclear and localization signals (black). The exonuclease (exo) domain of WRN is shown in yellow. Adapted from (Ouyang, Woo et al. 2008)

1.6.2 BIOCHEMISTRY OF BLM

It has been proposed that BLM helicase resolves aberrant paired structures at stalled replication forks and DSB, reducing the availability of substrate for HR (Cheok, Wu et al. 2005; Friedburg 2006). In the early stages of HR, BLM can promote resection of DNA ends to create the 3’ single–stranded DNA tails required for RAD51 to facilitate DNA strand invasion. However, the resulting D-loop can also be disrupted by BLM. In the latter stages of HR, the BLM complex or GEN1/Slx1-Slx4 complex complete the HR reaction by catalyzing Holliday junction dissolution or resolution, respectively (Wu and Hickson 2003; Ip, Rass et al. 2008; Svendsen, Smogorzewska et al. 2009; Svendsen and Harper 2010). The outcome of different HR pathways in terms of crossover or non-crossover recombinant products is dependent on the enzymatic processing. BLM has been shown to stimulate 5’-3’ end resection activity of EXO1 creating a 3’ hydroxyl
overhand necessary for Rad51 loading and single strand annealing activity (Gravel, Chapman et al. 2008; Mimitou and Symington 2008). Another enzymatic activity of BLM is its ability to displace D-loop structures or prevent their formation by removing Rad51 from the 3’ single stranded DNA tail (Bachrati, Borts et al. 2006; Bugreev, Yu et al. 2007). Displacement of the D-loop can both promote and inhibit recombination. Displacement of the strand before replication will prevent recombinational repair from proceeding. Displacement of the D-loop after sufficient synthesis will drive HR and lead to GC. Preventing D-loop formation in the first place would inhibit recombinational repair suggesting a possible role for BLM in the “decision” to pursue HR or NHEJ. Displacement of the D-loop after sufficient synthesis driving HR toward gene conversion might serve as an alternative role for BLM in preventing NAHR (van Brabant, Ye et al. 2000; Bachrati, Borts et al. 2006).

The BLM protein, thought to decreases illegitimate recombination, has been implicated in resolution of Holiday junctions in DSB repair in a process called dissolution (Wu and Hickson 2003). BLM protein in conjunction with topoisomerase IIIα and RMI1/RMI2 can dismantle double holiday junctions without cleavage or ligation. This process has been named “dissolution” (Wu and Hickson 2003; Raynard, Bussen et al. 2006). BLM binds and migrates double Holliday junctions to promote convergent branch migration, creating a hemicatenane structure that is “dissolved” to allow the two linked chromosomes to resolve into two unconnected double-strand DNA molecules. TOPOIIIα is required during this step to relieve the superhelical tension created by the convergence of the two branches. The hemicatenane is then decatenated by the single-stranded DNA passing activity of TOPOIIIα, in a reaction strongly stimulated by the RMI1–RMI2 complex. The outcome of the dissolution process is the exclusive formation of non-crossover recombinant products. This is believed to be the mechanism by which BLM suppresses NAHR.
Figure 1.6. Graphic depiction of BLM binding double Holliday junctions to promote convergent branch migration, creating a hemicatenane. In concert with TOPOIIIα and RMI1–RMI2, BLM the Hemicatenane structure is decatenated to form only non-crossover products. Adapted from (Chu and Hickson 2009).

1.7 RECG HELICASE PROTEIN
1.7.1 BIOCHEMISTRY OF RECG HELICASE

The *Escherichia coli* RecG protein is a 76-kDa DNA-dependent ATPase that binds specifically to model Holliday intermediates and drives branch migration of those Holliday intermediates in genetic recombination and DNA repair (Sharples and Lloyd 1991; Lloyd and Sharples 1993; Whitby, Ryder et al. 1993). RecG can catalyze the formation of Holliday junctions at stalled replication forks by promoting fork regression (McGlynn and Lloyd 2002). RecG unwinds both the nascent leading and lagging strands which anneal to each other and allow the parental strands to also reanneal generating a four-stranded Holliday junction. This process is depicted below in figure 1.7.
Aside from its ability to bind and migrate Holliday structures and regress multi-stranded DNA structures as mentioned above RecG has also been shown to dismantle D-loops where a 3'-OH single stranded nucleofilament has invaded a homologous DNA duplex (Whitby, Vincent et al. 1994; Vincent, Mahdi et al. 1996; Fukuoh, Iwasaki et al. 1997; McGlynn, Al-Deib et al. 1997). The crystal structure of RecG has provided great insight into how the enzyme accomplishes these tasks. The crystal structure in a complex with a partial DNA replication is depicted below in figure 1.8. RecG has conserved helicase domains linked to a novel N-terminal 'wedge' domain. This domain provides the specificity for binding three-way branched DNA structures. It has been proposed that the helicase motor acts as a double-stranded DNA translocase, differing from other helicases analyzed in that they typically mediate strand separation via translocation on single-stranded DNA (Mahdi, Briggs et al. 2003). This allows RecG to pull the two parental strands of a replication fork structure through two separate channels flanking the wedge domain, neither wide enough to accommodate dsDNA (Singleton, Scaife et al. 2001). This has the effect of “stripping off” the nascent strands, allowing the parental strands to re-anneal, as suggested by earlier biochemical studies (McGlynn, Lloyd et al. 2001; McGlynn and Lloyd 2002). Thus RecG protein also differs from many helicases in that it
functions primarily as a monomer (McGlynn, Lloyd et al. 2001; McGlynn and Lloyd 2002).

![Image of domain structure of RecG from Thermatoga maritima](image)

**Figure 1.8.** Domain structure of RecG from *Thermatoga maritima* in a complex with a partial replication fork structure lacking a leading strand (72). The N-terminal section of RecG shows the wedge domain responsible for specific DNA binding and unwinding (Mahdi, Briggs et al. 2003) with permission.

Similar to BLM, RecG has several roles in controlling recombination. Recently an alternative function to RecG in DNA synthesis was uncovered. Replication in *E. coli* generates 3’ flaps as the two replication forks collide and which displaces the nascent leading strands and provides potential targets for PriA to restart new replication forks. Here, RecG is thought to prevent replication fork restart by disassembling the D-loop after strand invasion. Essentially RecG and ssDNA exonucleases act to limit PriA-mediated re-replication of the chromosome and the consequent generation of linear DNA branches that provoke recombination. Like BLM, RecG seems to act to both promote and prevent recombination dependent upon the scenario (Rudolph, Upton et al. 2009; Rudolph, Mahdi et al. 2010).
1.7.2 BIOCHEMICAL SIMILARITY OF RECG AND BLM

BLM is a human homolog of bacterial RecQ, this homology is apparent in figure 1.5. However, unlike BLM, bacterial RecQ protein is not a branch migration-specific helicase as it does not display greater helicase activity on a four-way junction DNA substrates than it does on other DNA substrates (Harmon and Kowalczykowski 1998; Sun, Karow et al. 1998; Wu and Maizels 2001). This activity is thought to be the most prevalent biochemical activities of BLM. According to overlapping biochemistry, BLM seems to be more functionally similar to RecG with which it shares no homology. Both BLM and RecG can bind to and regress multi-stranded DNA structures that model stalled replication forks (McGlynn, Lloyd et al. 2001; Robu, Inman et al. 2004; Machwe, Xiao et al. 2006; Ralf, Hickson et al. 2006). Both proteins also have the capacity to bind and branch migrate Holliday junctions (Karow, Constantinou et al. 2000; Plank, Wu et al. 2006) (Whitby, Vincent et al. 1994) (Grove, Harris et al. 2008). BLM and RecG have also been shown to dismantle D-loops where a 3' single strand DNA has invaded a homologous DNA duplex (Whitby, Vincent et al. 1994; McGlynn, Al-Deib et al. 1997; van Brabant, Ye et al. 2000; Briggs, Mahdi et al. 2004; Bachrati, Borts et al. 2006; Grove, Harris et al. 2008). Chapter 5 explores this relationship in order to assess the importance of these overlapping biochemical activities to determine if they are in fact responsible for the genomic instability phenotypes observed in BLM deficient cells (Killen, Stults et al. 2009). If RecG and BLM are functional analogs this likeness can be exploited to explore the mechanism by which BLM suppresses genomic instability.
1.8: REFERENCES


CHAPTER 2: GENE CLUSTER INSTABILITY ASSAY

2.1: SYNOPSIS

A newly developed method for quantitatively detecting genomic restructuring in cultured human cell lines as the result of recombination is presented: the “gene cluster instability” (GCI) assay. The assay is physiological in that it detects spontaneous restructuring without the need for exogenous recombination-initiating treatments such as DNA damage. As an assay for genotoxicity, the GCI assay is complementary to well-established sister chromatid exchange (SCE) methods. Analysis of the U-2 OS osteosarcoma cell line is presented as an illustration of the method.

2.2 INTRODUCTION

Homologous recombination (HR) is essential in the maintenance of the integrity of the human genome and is the only mechanism for error-free DNA repair of double strand breaks. This form of repair utilizes non-broken homologous sequence located elsewhere in the genome, such as on a sister chromatid, to effectively replace damaged sequence through the process of gene conversion. However, dysregulated homologous recombination between the many non-allelic repetitive sequences in the human genome (Consortium 2004) can also have disastrous consequences for genomic stability (Lupski and Stankiewicz 2005) by providing a mechanism for generating physical alterations in the genomic architecture, including chromosomal translocations, inversions and deletions. Depending on the relative orientations of the recombining sequences, the formation of dicentric and acentric chromosomes is also possible (Acilan, Potter et al. 2007) (McClintock 1939). These structural anomalies may contribute to cellular cancer phenotypes (Tonon, Wong et al. 2005). This type of dysregulated homologous recombination is referred to as non-allelic homologous recombination (NAHR) and involves the physical exchange of genetic material through crossover between two different chromosomal loci with a high degree of sequence identity. Like error-free HR, NAHR is sequence similarity dependent and becomes more efficient when sequence
similarity between the recombining sequences is greater than 98% (Elliott, Richardson et al. 1998).

We describe here the Gene Cluster Instability (GCI) assay we developed for detecting spontaneous recombination-mediated genomic restructuring in human cells. We have successfully applied this technique to elucidate the genetics that regulates the NAHR reaction (Killen, Stults et al. 2009) and to show that NAHR is commonly associated with human adult solid tumors (Stults, Killen et al. 2009). In principle the assay is also suitable for evaluating the genomic toxicity of drugs. The assay involves physical analysis of gene clusters, genomic loci in which the repeated genes have very high levels of sequence identity and are in very high local concentration, both factors in accelerating the rate at which they undergo recombination-mediated structural alteration. NAHR between similarly oriented repeats causes expansions and contractions of the number of repeats in the cluster. These length changes can be monitored by excising the gene clusters of interest from the genome enzymatically, resolving cluster lengths through pulsed-field electrophoresis and detecting the clusters by Southern hybridization (Stults, Killen et al. 2008).

The gene clusters we usually employ (and describe here) to monitor recombination are the clusters expressing the precursor transcript to the three largest of the four ribosomal RNA molecules. There are ten such gene clusters in the human genome found at both paternal and maternal 13p12, 14p12, 15p12, 21p12, and 22p12 chromosomal loci (Henderson, Warburton et al. 1972) each consisting of a tandemly repeated 43 kb genes (the “rDNA”) with a variety of relative orientations (Caburet, Conti et al. 2005) but commonly oriented such that transcription proceeds towards the centromere (Worton, Sutherland et al. 1988). The individual rDNA clusters range from 1 to over 140 repeat copies representing overall lengths ranging from 43 kb to over 6 Mb with very strong variability demonstrated between individual humans (Stults, Killen et al. 2008). We have found that assaying cluster length changes in the size range from 50 kb to 1 Mb combines good sensitivity for detecting recombination-mediated genomic structural alterations with relative technical ease.

A schematic of the procedure is shown in Figure 2.1. High molecular weight genomic DNA from cells of interest is isolated in solid phase agarose to prevent mechanical
shearing and subjected to restriction digestion, also in the solid phase. Restriction enzymes are selected that do not have a recognition site in the gene cluster unit repeat. Since the gene repeats are nearly identical to each other, such an enzyme will likely not cut anywhere within the entire cluster, whereas non-repetitive flanking genomic DNA of essentially random sequence will be subject to digestion. So long as the cluster length is large relative to the cutting frequency of the enzyme in random sequence DNA, the cluster will be liberated with a relatively negligible length tail of non-cluster DNA on each end. For the rDNA, we find EcoRV to be an excellent enzyme: no recognition site in the rDNA clusters, frequent cutting in random sequence DNA, digests DNA in solid-phase agarose efficiently, and is affordable. (see Note 1).

Separation of the clusters using pulsed-field electrophoresis and detection though Southern hybridization generates a characteristic electrophoretic karyotype of the gene clusters. For a clonal cell line with no recombinational instability, this pattern will be well-defined with one band for each cluster, each band with a radioactive hybridization signal in proportion to the number of repeats in the cluster (Figure 2.2: ‘Initial Pattern’). If there is no recombination-mediated gene cluster restructuring, this pattern will be faithfully transmitted to all daughter cells. Recombination can, however, change the lengths of these clusters as cells are cultured. Thus, recombinational instability is manifested as subpopulations of cells in a culture with a different electrophoretic karyotype. Experimentally, these sub-populations can be detected by reduced-intensity bands that differ in length from the initial pattern superimposed upon the initial pattern (Figure 2.2: ‘Low GCI’). In the case of extreme instability, such as when the Bloom syndrome protein is lost or inactivated, active restructuring generates a ladder-like pattern that essentially completely obscures the initial pattern (Figure 2.2: ‘High GCI’) (Killen, Stults et al. 2009). The ladder-like banding is diagnostic for recombination-mediated changes, since recombination requires alignment of the repeated sequences and can thereby only change cluster lengths by integer multiples of the unit repeat length.

GCI analysis results for the human osteosarcoma line U-2 OS are illustrative. Figure 2.3A shows the electrophoretic karyotype for a stock culture of U-2 OS cells along with the SV40-transformed wild-type fibroblast line GM00637 (Coriell) and the HeLa S3
cervical carcinoma line (ATCC). Since these three lines are derived from three different individuals, the initial pattern of bands is expected to be different between the three lines, as indeed it is; this attribute of the rDNA clusters is useful to ensure that cell lines do not become confused with each other. The U-2 OS culture shows a profusion of minor-intensity bands with no well-defined pattern, indicative of recombinational instability. In a non-clonal culture such as that shown for the U-2 OS cells in Figure 2.3A, however, the total accumulated instability is a factor of three independent parameters: i) the spontaneous per-cell-division recombination rate, ii) the number of cell divisions elapsed since clonality, and iii) the degree to which the culture has been subject to periodic genetic population bottlenecks due to repeated splitting of the culture. Usually it is the spontaneous recombination rate that is of interest. This can be determined directly by re-initiating a culture from single-cell derived subclones to clear out all sub-populations, followed by free expansion without limit until genomic DNA is prepared. DNA isolated from such single-cell-derived subclones of the U-2 OS parental population yields the data shown in Figure 2.3B. Unlike the mixed parental population, now the subclones reveal relatively recombination-stable initial patterns with few minor-intensity bands. The process can be reiterated by generating sub-subclonal lines from individual cells of the now well-defined subclonal populations. Figure 2.4 shows such analysis from sub-subclonal populations generated from the U-2 OS parental subclones A and B respectively. Recombinational instability is still present, as indicated by missing initial pattern bands and new minor-intensity bands, but of a similar rate to that seen with other wild-type immortalized cell lines (Killen, Stults et al. 2009). Clearly the majority of the instability seen in the non-clonal parental U-2 OS culture (Figure 2.3A) was due to extended time in culture and/or repeated high-dilution passaging.

2.3. Materials
2.3.1 Thawing cryopreserved cells
1. Latex or nitrile gloves.
2. 10-cm tissue culture plates (or tissue culture flasks for non-adherent cells).
3. Adherent or non-adherent human cells.
4. Humidified, CO₂ supplemented tissue culture incubator.
5. Laminar flow tissue culture biosafety hood with standard tissue culture setup, including serological pipettes, micropipettes, and vacuum aspiration apparatus.
6. Tissue culture medium appropriate for each cell type such as minimal essential medium (MEM) or RPMI 1640
7. 95% reagent grade ethanol.
8. Fetal bovine serum.
9. L-glutamine/Penicillin/Streptomycin 100x solution (10,000 U/mL penicillin; 10,000 μg/mL streptomycin; 200 mM L-gln).
10. Plasmocin 25 mg/mL (for mycoplasma prophylaxis).
11. Clinical centrifuge (e.g., Marathon model 3200, Fisher, Pittsburg, PA).
12. 15 mL conical centrifugation tubes.
13. Complete culture medium appropriate for tissue culture growth of cells: MEM or RPMI supplemented with 5% to 10% fetal bovine serum, a 1:100 dilution of the stock Pen/Strep/L-gln solution, and a 1:10,000 dilution of the stock Plasmocin 25 mg/mL solution.
14. 70% ethanol: 95% ethanol diluted to 70% with sterile distilled water.

### 2.3.2 Subculturing cells for GCI analysis

1. Laminar flow hood with standard tissue culture setup, including serological pipettes, micropipettes, and vacuum aspiration apparatus.
2. Trypsin/EDTA solution (0.05% Trypsin, 0.53 mM EDTA).
3. Tissue culture medium such as MEM or RPMI 1640, suitably supplemented with fetal bovine serum and antibiotics.
4. 10-cm tissue culture plates (or tissue culture flasks for non-adherent cells).

### 2.3.3 Deriving and expanding clonal lines

1. 20 μL Pipetman (Gilson or equivalent) and sterile pipette tips.
2. 96-well flat bottom shaped well plates for adherent cells or V-bottom shaped well plates for non-adherent cells.
3. 24-well tissue culture plates.
4. 6-well tissue culture plates.
5. 10 cm tissue culture plates.
6. Hemocytometer or flow cytometer (e.g., Partec Analysis System PAS, Partec, Münster, Germany) for cell counting.

2.3.4 Preparing subcultured cells for DNA extraction
1. Trypsin/EDTA solution (0.05% Trypsin, 0.53 mM EDTA).
2. Tissue culture medium such as MEM or RPMI 1640, suitably supplemented with fetal bovine serum and antibiotics.
3. Clinical centrifuge.
4. 15 mL conical centrifugation tubes.
5. Sterile phosphate-buffered saline (also known as PBS or DPBS).
6. Hemocytometer or flow cytometer for cell counting.

2.3.5 Isolation of solid-phase high molecular weight genomic DNA from human cells
1. Low-melting-point (LMP) agarose, analytical grade (cat. #V2111, Promega, Madison, WI).
2. DPBS (cat. #21-031-CV, Mediatech, Manassas, VA).
3. 1 mL syringes (cat. #309602, Becton, Dickinson, San Jose, CA).
5. 8 mL flat bottom tubes.
6. Cell digestion buffer with proteinase K
7. Environmental incubator shaker capable of maintaining 50°C (e.g., G24 environmental incubator shaker, New Brunswick Scientific, Edison, NJ).
8. 50 mL conical tubes.
9. TE solution.
10. TE/glycerol solution.
11. Saturated phenylmethanesulfonylfluoride (PMSF) in isopropanol. CAUTION: very toxic!
12. Room temperature shaker
13. Proteinase K powder. Store at -20°C.
14. Proteinase K solution: Dissolve in water to 15 mg/ml. Store in 160 µl aliquots at
-20°C. CAUTION: proteinase K fines are intensely irritating. Wear suitable respiratory protection when using powered proteinase K.

15. Cell digestion buffer with proteinase K: 500 mM EDTA pH 8.0, 1% sarcosyl. Store at room temperature. Add 160 ml proteinase K per 5 ml digestion buffer for a final proteinase K concentration of approximately 0.5 mg/ml immediately before use.

16. TE: 10 mM Tris pH 8.0, 1mM EDTA pH 8.0

17. TE/glycerol solution. (10 mM Tris, 1 mM EDTA, 50% glycerol(w/v), pH 8.0)

18. Saturated PMSF in isopropanol: Add isopropanol to PMSF crystals to make a saturated solution. Some PMSF crystals should remain undissolved in the bottom of the container. Store at room temperature.

2.3.6 Enzymatic digestion of high molecular weight DNA for PFGE

1. Standard single edge safety razor blades.

2. 1.5 mL eppendorf tubes.

3. NEB buffer 3 (cat. #B7003S, New England Biolabs, Beverly, MA) or other suitable buffer for the restriction enzyme of choice.

4. EcoRV restriction enzyme (ca.# R0195L, New England Biolabs, Beverly, MA)

5. Warm room capable of maintaining 37°C or environmental incubator shaker capable of maintaining 37°C.

6. Mini-Labroller with 1.5 ml eppendorf tube holder attachment or equivalent (Labnet International).

2.3.7 Preparation of agarose gel for PFGE

1. Pulsed field certified (PFC) grade agarose (cat. #162-0137, Bio-Rad Laboratories, Hercules, CA).

2. 125 mL clean glass bottle with screw-cap lid.

3. 500 mL pyrex beaker.

4. Hot plate.

5. Nanopure H₂O.

7. Bio-Rad universal comb holder and 15 or 20 well comb (cat. #170-3699; 170-3627; 170-4322, Bio-Rad Laboratories, Hercules, CA).
9. Parafilm M sealing film (cat. #PM-996, SPI supplies).
10. 0.5x TBE (45.5 mM Tris / 45.5 mM Borate / 2.0 mM EDTA): To 1000 mL nanopure H₂O add 6.6 g Boric acid, 12.94 g TRIS base, and 4.8 mL of a 500mM EDTA pH 8.0 solution. Bring volume up to 2400 mL with water. Make fresh before each use.

2.3.8 Loading digested DNA samples into the agarose gel for PFGE
1. 0.5x TBE.
2. 200 µL Pipetman (Gilson or similar) and tips.
3. Low-melting-point (LMP) agarose, analytical grade (cat. #V2111, Promega, Madison, WI).
4. DPBS (cat. #21-031-CV, Mediatech, Manassas, VA).
5. Standard single edge razor blade.
6. Environmental incubator shaker capable of maintaining 50°C.

2.3.9 Loading and running pulsed field electrophoresis gels
1. Bio-Rad CHEF MAPPER XA with cooling module system (Bio-Rad Laboratories).

2.3.10 Ethidium staining and preparing the gel for in-gel hybridization of a radiolabeled probe
1. Gel imaging documentation system
2. Flat stainless steel pan
3. Hybridization oven. (TECHNE)
4. Kimwipes (Kimberly-Clark).
5. Ethidium bromide: 1% solution in water stored at 4°C. CAUTION: Ethidium bromide is a known carcinogen. Wear suitable respiratory protection when manipulating ethidium bromide powder.
6. Ethidium bromide/glycerol solution. To 197 mL nanopure H₂O add 3 mL 50% glycerol solution (w/v) and 15 µL of the 1% stock solution of ethidium bromide.

2.3.11 PCR reaction: non-radiolabeled Southern blot probe template preparation and radiolabeled 45S rDNA Southern blot probe generation

1. Oligonucleotides: rDNA11-T: GGGCTCGAGATTTGGGACGTCAGCTTCTG and rDNA11-B: GGGTCTAGAGTCCC TTCCTCTGTGAG
2. Thermocycler (e.g., Mastercycler gradient, Eppendorf Scientific)
3. dGATC-TP nucleotide mix, combined and diluted to a final concentration of 10 mM each nucleotide (cat. #10297-018, invitrogen, Carlsbad, CA).
4. dGCT-TP nucleotide mix, combined and diluted to a final concentration of 40 µM each nucleotide (see Note 2).
5. dATP diluted to a final concentration of 20 µM.
6. α-³²P-dATP (50 uCi at 3000 Ci/mmol) (Perkin Elmer)
7. TAQ polymerase (New England Biolabs)
8. 10x TAQ buffer (New England Biolabs)
9. spin-50 mini-column (USA Scientific)
10. FlexiGene DNA Kit (Qiagen).
11. Illustra GFX PCR DNA and Gel Band Purification Kit (GE Biosciences).

2.3.12 Southern blot analysis using in-gel hybridization of the radiolabeled rDNA probe

1. Hybridization tube (TECHNE)
2. Nylon mesh (PGC Scientifics)
3. Hybridization oven (TECHNE)
4. Polyvinyl-chloride plastic wrap
5. Molecular Probes Phosphor Screen (GE Lifesciences)
6. PhosphorImager (e.g., Storm 860, Molecular Dynamics)
7. SYBR®safe stain (Invitrogen)
8. Denaturation solution (0.4N NaOH, 0.8M NaCl): 1.6 g of NaOH pellets and 4.67 g NaCl dissolved in water to 100 mL. Make fresh before use.
9. Neutralization solution (0.5M Tris pH 8.0, 0.8M NaCl): Add 4.67 g NaCl to 50 ml of 1M Tris pH 8.0 solution and add water to 100 mL final volume. Make fresh before use.

10. 20x SSC (3M NaCl and 300mM sodium citrate): Dissolve 175.4 g of NaCl and 88.2 g of sodium citrate dihydrate in 800 ml nanopure H₂O. Adjust volume to 1 liter.

11. Hybridization solution (2x SSC, 7% SDS, 0.5% Hammersten casein): Add 100 mL 20x SSC to 700 mL nanopure H₂O. Warm to 65°C. Dissolve 70 g SDS and 5 g Hammersten casein (USB corp. cat. #12840). Adjust volume to 1 liter. Store at room temperature. CAUTION: SDS power fines are very irritating. Wear suitable respiratory protection.

12. Wash solution 1 (2x SSC and 0.1% SDS): Add 10 mL of 20x SSC to 89 ml nanopure H₂O. Lastly add 1 mL of a 10% SDS solution and mix well. Make fresh before each use.

13. Wash solution 2 (0.5x SSC and 0.1% SDS): Add 2.5 mL of 20x SSC to 96.5 ml nanopure H₂O. Lastly add 1 mL of a 10% SDS solution and mix well. Make fresh before each use.

2.4 Methods.

2.4.1 Thawing cryopreserved cells

1. Prepare a tissue culture plate with pre-warmed culture medium at least an hour before thawing cells allowing for ample time for it to equilibrate to the correct pH and temperature in the tissue culture incubator at 37°C and with 5% CO₂.

2. Remove a vial of frozen cells from liquid nitrogen storage and wipe it down with 70% EtOH in a laminar flow tissue culture hood.

3. Open the vial slightly to allow the gas pressure inside the cryovial to equalize with the ambient atmosphere. Close the vial of cells and continue thawing in your gloved hand until there is a still-frozen pellet in thawed liquid that is mobile when the tube is shaken.

4. Shake the vial vigorously, then remove the lid from the vial and pour the liquid and semi-frozen pellet of cells into the medium of the prepared tissue culture dish. Swirl the cells to thaw the residual frozen pellet and to get an even distribution of cells across the plate. Place the plate containing the cells immediately back in the tissue culture incubator overnight.
5. The next day the medium should be replaced to get remove any residual cryopreservative from the freezing media. For adherent cells the medium should be aspirated off and replaced with fresh pre-warmed medium. For non-adherent cells the cell suspension should be removed to a 15 mL conical tube and centrifuged at 200 g for 5 minutes. The media can then be aspirated off and the resulting cell pellet be resuspended in pre-warmed culture medium and placed in a new tissue culture plate.

2.4.2 Subculturing cells
1. Once the cells are growing and semi-confluent in their tissue culture dish, remove the cells from the incubator and place them in a laminar flow tissue culture hood. For non-adherent cell lines, skip to step 6.
2. Aspirate the medium completely from the plate.
3. Add an appropriate amount of trypsin/EDTA solution to the plate making sure that the entire bottom surface of the plate is evenly covered. For a 10-cm dish 2 mL is typically sufficient.
4. Allow cells to sit for several minutes while gently rocking the plate to maintain an even distribution of the solution and monitor the cells as they start to detach from the plate. The amount of time it takes for the cells to detach will vary widely depending on the cell type, however, this is easily determined by careful observation.
5. Once most of the cells have detached add 2x the volume of culture medium containing at least 5% FBS. The FBS in the culture medium inactivates the trypsin and prevents it from damaging the cells through prolonged exposure. If the trypsin is not inactivated, prolonged exposure can cause the cells to lyse.
6. Pipette the cells up and down vigorously in order to break up any clumps and to ensure all cells have detached from the plate and are homogenously suspended.
7. Once you have a single cell suspension distribute the suspension among new culture dishes with pre-warmed medium in a ratio that is ideal for the cell line of interest. This can typically range from a 1:3 dilution for slower growing cells to a 1:10 dilution for faster growing cells such as HeLa.

2.4.3 Deriving and expanding clonal lines
When the GCI assay is used to determine ongoing genomic instability in cultured cell lines it requires the development of clonal lines derived from a given cell line.

1. Start by preparing a single cell suspension in the same fashion as you would above as if subculturing the cells. The single cell suspension is then used to prepare a limiting dilution series and derive colonies grown from a single cell.

2a. For adherent cells set up a dilution series and plate the cells in 10-cm plates at 1:10; 1:50; 1:250; 1:500; 1:1000; 1:2000; and 1:4000 dilutions.

or

2b. For non-adherent cells determine the cell density with a hemocytometer or flow cytometer, then dilute cells suitably with medium and aliquot cells into separate 96-well V-bottom plates at dilutions of 100 cells/well; 10 cells/well; 5 cells/well; 1 cell/well; and 1 cell/5 wells. Place cells into all 96 wells for each of the plates used (see Note 3). Over the next 8 to 14 days colonies from begin to form either on plates for adherent cells or in wells for non-adherent cells.

3. For adherent cells only: once colonies are sufficiently 2-3mm in diameter they can be picked from the plate with a 20 µL pipette and moved to a single well in a 96-well tissue culture plate. For fragile cells, partial trypsinization by treatment with trypsin/EDTA solution diluted 10:1 with DPBS will help ensure cell integrity in the colony transfer process. Henceforth, treat adherent and non-adherent cultures similarly.

4. For all cells: clones should be expanded from 96-well plates to 24-well plates, 6-well plates and finally 10 cm dishes successively. Cells should be allowed to grow to confluency before each expansion step. Once confluent they are subcultured as described above using smaller volumes of trypsin/EDTA solution for adherent cells: 50 µl per each well of a 96-well plate, 200 µl per well for 24-well plates, 500 µl per well for 6 well plates. Each step represents approximately a 1:4 expansion of the cells, which in our hands works well for almost any line chosen.

2.4.4 Preparing cultured cells for high-molecular weight solid-phase DNA isolation

1. Melt certified nuclease-free low-melting-point (LMP) agarose at 1.2% weight/volume in DPBS (phosphate-buffered saline without Ca+2 or Mg+2, tissue culture grade) at
70°C, place the melted gel solution in a waterbath or oven at 42°C and allow the temperature to equilibrate.

2. With adherent cells, treat with trypsin/EDTA to detach cells followed by addition of a 2x volume of culture medium and pipeting up and down in a serological pipette to achieve a single cell suspension. Likewise non-adherent cells can simply be pipetted up and down to break up clumps and create a single cell suspension. From this point cells are stored on ice at all times except during centrifugation which can be performed at room temperature.

3. Remove the suspension to a 15-mL conical vial and centrifuge at 200 g for 5 minutes.

4. Resuspend the cells in 10 mL DPBS to rinse.

5. Determine the concentration of cells with either a hemocytometer or flow cytometer.

6. Centrifuge cells at 200 g for 5 minutes and aspirate the DPBS rinse solution.

7. Resuspend cells in DPBS to a final concentration of 3x10^7 cells / ml, taking into account the non-zero volume of the pelleted cells themselves. Store cells on ice temporarily if necessary.

8. Add two volumes of melted 1.2% LMP agarose solution to the 3x10^7 cells / ml cell suspension for a final concentration of 1x10^7 cell/ml in 0.8% LMP agarose and mix thoroughly by vortexing.

9. Draw the cell/gel suspension into a 1 ml syringe and immediately place the syringe on ice and cover with ice to solidify the agarose before the cells have a chance to settle.

10. After the agarose solution has solidified cut the end off the syringe with a single edged razor blade and extrude the DNA/agarose “worm” carefully into an 8 ml flat-bottomed tube.

11. Add 5 ml of digestion buffer with proteinase K, invert gently several times to mix and incubate overnight at 50°C with gentle agitation (see Note 4).

12. After suitable digestion has cleared the appearance of the DNA/agaorse, remove the DNA/agarose “worm” to a fresh 50 ml tube, add 40 ml TE and agitate gently at room temperature for 30 minutes. Do this step twice.

13. Decant the TE and transfer the “worm” to a new 8 ml flat-bottomed tube.

14. Add 6 ml TE and 6 µl saturated PMSF in isopropanol. Mix well and gently agitate at room temperature for 60 minutes.
15. Remove the TE/PMSF solution and add 6 mL of fresh TE without PMSF to rinse with gentle agitation at room temperature for 30 minutes. Do this step twice.
16. Pour off the TE and add back chilled TE-glycerol and gently agitate at 4°C for 30 minutes.
17. Remove the TE/glycerol and add back 6 ml fresh TE/glycerol. Gently agitate overnight at room temperature.
18. Cut off the tip of a fresh 1 mL syringe with a clean razor blade and transfer the TE/glycerol equilibrated agarose “worm” into this syringe. Seal the syringe with Parafilm and store indefinitely at -20°C.

2.4.5 Enzymatic Digestion of high molecular weight DNA for PFGE
1. Agarose/DNA slices of ~15 μL volume are cut from the high molecular weight DNA/agarose “worms” prepared and stored in 1 ml syringes with a standard single edge razor blade. Each slice should be approximately 0.8 to 0.9 mm thick. It is helpful to practice your slicing technique with an agarose “worm” that does not contain DNA, measuring slice volumes on an analytical balance until uniform thickness slices can be consistently achieved. Aim for a consistent slice weight of approximately 15 mg.
2. Place the DNA-containing agarose slice inside the lid of an inverted 1.5 mL eppendorf tube. The lid of an inverted 1.5 mL eppendorf tube provides a perfect sized container for the slice with a flat bottom and walls that also holds 200 μL of buffer.
3. Remove the glycerol/TE from the slice by washing the slice in the lid of the 1.5 ml eppendorf tube 3 times with 200 μL of 1x NEB buffer 3 by pipetting gently up and down so the slice is agitated in the buffer solution and then pipetting off the wash solution. Take care not to damage the slice with the pipette tip during the washes.
4. Remove the last wash and replace with 200 μL of NEB buffer 3 containing 20 units of EcoRV restriction enzyme.
5. While keeping the eppendorf tube inverted, gently but firmly close the eppendorf tube lid. Mount the inverted eppendorf tube in the Mini-Labroller at 30° off vertical so that when the Labroller is turned on the tubes will be gently agitated while maintaining their inverted orientation. Place the Labroller with the inverted eppendorf tubes in a warmroom at 37°C overnight, and turn on the Labroller.
6. The next day, remove the Labroller apparatus with eppendorf tubes from the warm room. Remove the inverted eppendorf tubes from the Labroller, while maintaining their inverted orientation.

7. Rinse the agarose slices containing the now digested genomic DNA with 0.5x TBE by pipetting gently up and down and then pipetting off the rinse, taking care not to disrupt the delicate agarose slices with the pipette. Do this rinse twice. Additionally, slice off and likewise rinse a thin sample of S. cerevisiae chromosome molecular weight markers.

8. The slices are now ready to load into the pulsed-field gel. Leave the slices in the final rinse solution in sealed inverted eppendorf tubes until ready to load the samples.

2.4.6 Preparation of an agarose gel for PFGE

1. Make 2.4 L of 0.5x TBE.

2. Place 1 gram of PFC grade agarose into a 125 ml glass bottle (eg. Kimax) along with 100 ml of 0.5x TBE. Gently swirl to ensure the agarose is hydrated and non-clumpy. Try not to let agarose clumps adhere to the walls of the bottle above the liquid. Cap the bottle tightly.

3. Place 250 ml water in a 500 ml glass beaker and bring to a gentle boil on a hot plate. Melt the agarose in 0.5x TBE by placing the sealed bottle containing the agarose/0.5x TBE mixture into the boiling water bath. Heat with periodic gentle swirling until all the agarose has been completely homogenously melted. The melted agarose solution should be clear, colorless and featureless when swirled.

4. Allow the melted agarose to cool only slightly (less than 5 minutes) on the benchtop prior to pouring into the gel casting set up. This should be ample time to set up the casting apparatus. We have found the pouring the gel while it is hot produces optimal resolution of the gene cluster bands. The gel casting system should be set up with the universal comb holder and 15 or 20 well comb and placed flat on a Bio-Rad (or similar) leveling table.

5. Once the casting apparatus is prepared pour the melted agarose solution slowly into the casting tray trying to avoid forming any bubbles. Eliminate any bubbles that may arise while the agarose solidifies. Pay special attention to any bubbles that might form along the wells of the comb and the edges of the casting tray. Allow the gel to solidify at
room temperature on the benchtop for one hour. Then, place the solidified gel, still in the casting tray, in a refrigerator at 4°C for an additional hour. The gel will then be ready for loading the prepared DNA samples (see Note 5).

2.4.7 Loading digested DNA samples into the agarose gel for PFGE
1. Place a mixture of 0.8% LMP agarose suspended in 0.5x TBE into a tightly capped 50 ml plastic conical tube and heat in a boiling water bath until the agarose mixture is uniformly melted. Remove from the bath and place at room temperature (see Note 6).
2. Remove the solidified pulsed-field gel from the refrigerator and carefully remove the comb from the wells.
3. Remove the 0.5x TBE final buffer rinse from the first agarose slice in the lid of an inverted 1.5 ml eppendorf tube. Gently flick the lid of the eppendorf tube onto a clean razor blade so the gel slice lands flat on the razor blade. Remove excess liquid from the gel slice with a pipette tip or the corner of a Kimwipe (see Note 7).
4. Push the slice from the razor blade into the well of the gel with a clean pipette tip. Make sure that the agarose slice flatly contacts the front of the well facing the direction in which the gel will be run. Allowing the slice to stick to the back of the well instead will compromise the resolution and final location of resulting bands.
5. Gently push the slice down into the well until the bottom of the slice rests on the bottom of the well, then fill the remaining space in the well carefully with melted 0.8% LMP agarose.
6. Repeat this step for each sample to be loaded, including the S. cerevisiae chromosome molecular weight markers.
7. Once the samples are all loaded the gel should be allowed to sit at room temperature for 10 minutes to allow the sealing agarose to begin to set (see Note 8).
8. Next the casting stand should be removed taking care not to dislodge the gel from the black running platform.
9. Wipe away any accumulated gel waste on the bottom of the running platform.
10. The gel and platform should then be placed at 4°C for 30 minutes. Again this step is essential for optimal resolution.
2.4.8 Loading and running the Bio-Rad CHEF MAPPER apparatus for PFGE

We have found that the optimal range for GCI analysis of the human rDNA gene clusters is from 50 kb up to 1 Mb. The instructions and the algorithm provided here are optimized for this size range. If this methodology is used to look at gene clusters other than the rDNA or in of size ranges of the rDNA clusters the conditions will require optimization for the desired size range. Consult Birren and Lai (Birren 1993) for more details.

1. Once the gel is loaded and has been chilled at 4°C, place the remaining 0.5x TBE buffer into the electrophoresis cell of the Bio-Rad CHEF MAPPER XA system.
2. The buffer pump should be turned on to circulate buffer and eliminate air bubbles from the system before the gel is place into the cell.
3. Once this is done turn the pump off and place the gel into the cell making sure the retention bracket is in place to prevent the gel from moving and that the wells are oriented at the top of the gel relative to the direction the DNA is going to travel.
4. Close the lid and make sure the safety interlock is engaged and turn the buffer pump back on.
5. Engage the cooling module of the Bio-Rad CHEF MAPPER XA system and allow it to cool the buffer and the gel to 14°C.
6. Enter a two state program such that the electrical field vector included angle is 120° and the electrical field strength for each vector is 6V/cm. Set the run time for 24 hours and the switch time from 3 seconds to 90 seconds with a ramp factor of 0.357. Once you are sure the algorithm is entered correctly and the chiller has cooled the buffer and gel to 14°C press “START” on the MAPPER. Look for bubbles from the electrodes to be sure the gel is running correctly.

2.4.9 Ethidium staining and drying the gel in preparation for in-gel hybridization with a radio-labeled probe

The next day, after the PFGE program has completed running, the gel can be stained with ethidium bromide and visualized to determine the outcome. The gel also needs to be prepared for in-gel hybridization of the radiolabeled probe. The in-gel hybridization involves drying the gel so that it is thin enough to permit the probe to easily diffuse into
and out of the gel yet not so thin as to shatter. This is accomplished by equilibrating the
gel to a final concentration of 0.5% glycerol and then drying the gel at 65°C (see Note 9).
1. After the PFGE program has finished its run carefully remove the gel and place it in
200 ml of ethidium bromide/glycerol solution. Including the 100 ml volume of the gel,
the final concentration of glycerol will be 0.5%. Incubate at room temperature for 30
minutes with gentle agitation.
2. After incubation the gel can be placed on the imaging system so that the ethidium
stained gel can be documented. Here it is important to minimize the amount of UV
exposure as it can degrade the DNA especially in the presence of EtBr.
3. After the ethidium stained gel has been documented the surfaces of the gel are dried
with a Kimwipe. Place the gel upside-down on a flat surface such that the flat side of the
gel is on top. Fold up two Kimwipes together, then partially wet one edge of the folded
Kimwipes with 0.5% glycerol solution and use this pre-wet edge of the Kimwipe
“sandwich” to wick liquid from the top surface of the gel. Using the partially wet edge of
the Kimwipes will prevent the Kimwipes from adhering to the gel surface. Similarly
wick liquid away from all sides of the gel. Now invert the gel so the flat wiped-dry face
points down and place the gel in a metal pan. Wick any remaining liquid from the top
surface of the gel using Kimwipes (see Note 10).
4. Place the gel and pan in the hybridization over at 65°C until the gel appears
homogenously dry and flat. The gel will dry to a thickness similar to a sheet of paper.
The gel can be left overnight to dry but we have found that removal of the gel
immediately after it finishing drying results in better hybridization of the radiolabeled
probe and thereby better results.
5. Once the gel is dry it can be processed immediately or stored covered with PVC wrap
(eg. Saran Wrap) for up to 2 weeks in a dark dry area.

2.4.10 Preparing the template DNA for PCR radiolabeling
Template DNA is prepared by PCR-amplifying a region of the rDNA gene from human
genomic DNA using primers 5′-GGGCTCGAGATTTGGGACGTCAGCTTCTG and
5′-GGGTCTAGAGTGCTCCC TTCCTCTGTGAG, to yield a 532-bp fragment. The
PCR product can be subsequently digested with XhoI and XbaI and subcloned into
pBluescript II SK− or other suitable cloning vector for long term propagation. It is also possible to simply use the PCR product as a template in the subsequent radiolabeling PCR reaction directly without subcloning, although this is not recommended.

1. Isolate human genomic DNA from any human cell line using a Qiagen FlexiGene DNA kit. Dissolve the DNA to a final concentration of 1 ml/ml in water.

2. Set up the following PCR reaction:

<table>
<thead>
<tr>
<th>µl</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>10x TAQ buffer with MgCl₂</td>
</tr>
<tr>
<td>1.0</td>
<td>genomic DNA at 1 µg/µl</td>
</tr>
<tr>
<td>0.4</td>
<td>dGATC-TP nucleotide mix at 10 mM each nucleotide</td>
</tr>
<tr>
<td>2.0</td>
<td>rDNA11-T primer at 1 µM concentration</td>
</tr>
<tr>
<td>2.0</td>
<td>rDNA11-B primer at 1 µM concentration</td>
</tr>
<tr>
<td>0.4</td>
<td>TAQ DNA polymerase (2U total)</td>
</tr>
<tr>
<td>12.2</td>
<td>water</td>
</tr>
<tr>
<td>20</td>
<td>FINAL VOLUME</td>
</tr>
</tbody>
</table>

3. Run the following PCR program with lid temperature set to 105°C:

<table>
<thead>
<tr>
<th>#</th>
<th>Instructions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>94°C for 3 minutes</td>
<td>initial denaturation</td>
</tr>
<tr>
<td>2.</td>
<td>94°C for 30 seconds</td>
<td>denature</td>
</tr>
<tr>
<td>3.</td>
<td>45°C for 30 seconds</td>
<td>anneal</td>
</tr>
<tr>
<td>4.</td>
<td>72°C for 1 minute, plus 2 seconds per cycle</td>
<td>extend</td>
</tr>
<tr>
<td>5.</td>
<td>goto step 2, repeat 29 times</td>
<td>30 cycles total</td>
</tr>
<tr>
<td>6.</td>
<td>72°C for 7 minutes</td>
<td>polish</td>
</tr>
<tr>
<td>7.</td>
<td>hold at 4°C</td>
<td>store indefinitely</td>
</tr>
</tbody>
</table>

4. Run the complete PCR reaction on a 1% agarose gel with suitable size makers.

5. Stain with ethidium bromide, visualize with UV light and excise the 523bp band with a clean razor blade.

6. Purify the template DNA from the gel slice using the GFX Kit.

7. Quantify the recovered DNA using Hoechst 33258 fluorimetry or similar methodology. Store at -20°C.

8. Optional, but recommended: Subclone the PCR product into a convenient cloning vector and have sequenced to ensure the rDNA probe is correct.

2.4.11 Preparing the radiolabeled 45S rDNA probe
The radio labeled probe is prepared by PCR using a$^{32}$P radiolabeled dATP. This method yields a probe that has very high specific activity (see Note 11).

1. If using plasmid as a template, dilute an aliquot of concentrated plasmid stock solution to 50 pg/ml. If using isolated PCR product as a template, dilute an aliquot of the PCR product to 10 pg/ml (see Note 12).

2. Set up the following PCR reaction:

<table>
<thead>
<tr>
<th>Volume (µl)</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>10x TAQ buffer with MgCl$_2$</td>
</tr>
<tr>
<td>2.0</td>
<td>rDNA containing plasmid at 50 pg/µl or PCR product at 10 pg/µl</td>
</tr>
<tr>
<td>2.0</td>
<td>rDNA11-T primer at 1 µM concentration</td>
</tr>
<tr>
<td>2.0</td>
<td>rDNA11-B primer at 1 µM concentration</td>
</tr>
<tr>
<td>5.0</td>
<td>α$^{32}$P dATP (50 µCi @ 3000 Ci/mmol)</td>
</tr>
<tr>
<td>2.0</td>
<td>dGTC-TP at 40 µM concentration each</td>
</tr>
<tr>
<td>1.0</td>
<td>dATP at 20 µM concentration – not radioactive</td>
</tr>
<tr>
<td>0.4</td>
<td>TAQ DNA polymerase (2U total)</td>
</tr>
<tr>
<td>3.6</td>
<td>water</td>
</tr>
<tr>
<td>20</td>
<td><strong>FINAL VOLUME OF THE REACTION</strong></td>
</tr>
</tbody>
</table>

3. Run the following PCR program with lid temperature set to 105°C:

<table>
<thead>
<tr>
<th>#</th>
<th>Instructions</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>94°C for 3 minutes</td>
<td>initial denaturation</td>
</tr>
<tr>
<td>2.</td>
<td>94°C for 30 seconds</td>
<td>denature</td>
</tr>
<tr>
<td>3.</td>
<td>45°C for 30 seconds</td>
<td>anneal</td>
</tr>
<tr>
<td>4.</td>
<td>72°C for 1 minute, plus 2 seconds per cycle</td>
<td>extend</td>
</tr>
<tr>
<td>5.</td>
<td>goto step 2, repeat 39 times</td>
<td>40 cycles total</td>
</tr>
<tr>
<td>6.</td>
<td>72°C for 7 minutes</td>
<td>polish</td>
</tr>
<tr>
<td>7.</td>
<td>hold at 4°C</td>
<td>store indefinitely</td>
</tr>
</tbody>
</table>

4. After the PCR reaction the probe is separated from the unincorporated nucleotides and primers with a USA scientific spin-50 mini-column. The radiolabeled probe is in the flow-through non-bound liquid phase (see Note 13).

5. Boil the probe for 2 minutes in a boiling water bath and then snap cool on water-saturated ice with periodic agitation for 2 minutes.
6. The probe should be kept on ice until it is used for hybridization. The probe should be used immediately rather than stored long-term to avoid radiolytic degradation.

2.4.12 In-gel hybridization of the radiolabeled rDNA probe and Southern analysis

While the radiolabeled probe is being prepared the dried gel can be processed to prepare it for hybridization.

1. Place the dried gel into H₂O for 10 minutes with gentle agitation and allow it to rehydrate.
2. After that the gel should be easily removable from the surface of the pan and can be placed in a deeper stainless steel pan or glass dish for the subsequent washes.
3. Wash the gel with 100ml of H₂O for 5 minutes. Do this step twice.
4. Remove the water and replace it with 100ml of denaturation solution, incubate for 30 minutes with gentle agitation.
5. Remove the denaturation solution and replace it with 100 ml of neutralization solution, incubate for 30 minutes with gentle agitation.
6. Remove the gel from the neutralization solution and transfer it to a hybridization roller tube placing nylon mesh between the gel and the glass surface. The mesh allows the gel to get an even exposure to hybridization solution on both sides.
7. Pre-hybridize the gel for 3 hours in 35 ml of hybridization solution at 65°C in the hybridization oven.
8. After 3 hours replace the hybridization solution with 25 ml of fresh pre-heated hybridization (65°C) solution.
9. Add the radiolabeled probe directly into the hybridization solution and close the end caps on the bottle. Place the hybridization tube back in the hybridization oven at 65°C and allow the gel to hybridize for at least 12 hours (see Note 14).
10. Following hybridization, remove the cap from the hybridization tube and discard the radioactive solution appropriately. Add 50 ml of wash solution 1. Place the tube back in the hybridization oven and incubate for 30 minutes at 65°C.
11. Discard the wash solution 1 from the hybridization tube and add 50 ml of fresh wash solution 1 for and additional 60 minutes at 65°C.
12. Discard the wash solution 1 and rinse the gel twice with wash solution 2. Each wash uses 50 ml solution 2 and a two hour incubation at 65°C.
13. Remove the gel from the hybridization tube and rinse briefly in 2x SSC.
14. Place the washed gel in all-purpose polyvinyl-chloride plastic wrap (Saran Wrap or equivalent) so that it does not dry out, being careful to blot up excess liquid with a paper towel or Kimwipe.
15. Expose the gel on a PhosphorImager screen overnight and detect the following day on a Molecular Dynamics Storm PhosphorImager or similar (see Note 15).
16. After data is acquired the gel can be stained in a 1:10,000 dilution of SYBR®safe dye. This fluorescently stained gel can be scanned on the Molecular Probes Storm PhosphorImager in fluorescent mode to give an accurate representation of the location of the size markers on the dehydrated gel.
17. Process the gel image using suitable image processing software.

Table 2.1. Adobe Photoshop and the following algorithm:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Convert the image to 8-bit grayscale and save in TIFF format.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Crop the image suitably.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Adjust the levels of the grayscale to maximize the difference between the white background and the darkest band present.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Blur using the following horizontal band-enhancing custom filter with a scale factor of 105 and zero offset to leave overall intensities unchanged:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 5 5 5 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 9 9 9 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 5 5 5 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Unsharp mask: 200% with 5.0 pixel radius and zero threshold.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Unsharp mask again: 50% with 3.0 pixel radius and zero threshold.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.5 Notes
1. The radiolabeled probe sequence and restriction enzymes described herein are specific for the human rDNA sequence. Gene cluster analysis in other other species and/or with non-rDNA clusters will require a different suitable choice of Southern hybridization probe and restriction enzymes.
2. There is no dATP in this solution. Low concentration dATP will be added separately.
3. The V-bottoms on the 96-well plates tend to concentrate the diluted cells and drastically improve non-adherent cell plating efficiencies at these low dilutions. In most cases use of these V-bottom plates is indispensable. Single plates are usually sufficient for the higher dilutions but it is usually necessary to prepare up to 10 plates at the two lowest dilutions. This way of deriving single cell colonies relies on a probability distribution and on the plating efficiency of the cell line. To be reasonably confident that all the clones at a given dilution level arose from a single cell, that dilution needs to produce no more than 10 wells showing growth per 96-well plate. Colonies are only expanded from this lowest dilution plate.

4. Incubation in digestion buffer should be allowed to proceed until the initial cloudy appearance of the agarose “worm” has cleared. For some cell types, this may take two days rather than overnight. It is permissible to incubate the DNA/agarose at 50°C in digestion buffer for longer periods of time if convenient.

5. Placing the gel on the benchtop for an hour and then leaving it at 4°C for an hour may seem trivial, however, in our hands this procedure gives reproducibly sharp bands when the gel is run.

6. The 0.8% LMP agarose will be used to seal the digested agarose slices into the wells of the pulsed-field gel. The 0.8% LMP agarose usually takes at least 30 minutes to solidify at room temperature in the 50 ml plastic tube after removal from the boiling water bath, which is easily sufficient time for loading the pulsed-field gel.

7. Sometimes, particularly with extra thin agarose slices, the slice may flick out of the eppendorf tube and land on the razor blade folded in half. If this should happen, the slice can be gently unfolded by pipetting 100 µl 0.5x TBE buffer onto the folded slice in a manner that swirls the liquid. After the slice has unfolded, wick away the excess liquid with a Kimwipe.

8. Since the gel was at 4°C prior to loading, the 0.8% agarose will solidify rapidly after addition to the wells.

9. Never run PFGE with gels that have ethidium bromide already in them – the ethidium bromide intercalates the DNA and will drastically effect run time and results.

10. Be sure that there is no liquid trapped under the gel between the gel and the pan. Gels will shatter upon drying if there is residual trapped liquid.
11. To avoid chain termination by misincorporation of unlabeled dNTPs, it is necessary to decrease the concentration of the unlabeled dNTPs to be similar to that of the limiting radiolabeled dATP (Mertz and Rashtchian 1994).

12. We have found it helpful to store the template in concentrated form and to dilute aliquots suitably for each labeling reaction. Storing very dilute solutions of DNA long-term is not recommended as adherence to the container of the DNA solution can dramatically reduce the DNA concentration and the final yield of radiolabeled product.

13. The degree of radiolabeling can be roughly measured by using a Geiger counter to observe the amount of radioactivity in the liquid flow-through vs that trapped in the spin-column. For a successful radiolabeling reaction, the liquid should have at least twice as much radioactivity as the column.

14. Longer hybridization times make no difference.

15. We often re-expose the gel on the screen for an additional three-day exposure to give a good reduction in background noise and sharper bands.
2.6: FIGURES

Figure 2.1: Schematic of the restriction digestion procedure for liberating intact gene clusters from bulk genomic DNA. Vertical arrows: hypothetical restriction enzyme recognition sites; Open Rectangles: individual repeated genes in a gene cluster. Reproduced from (Stults, Killen et al. 2008).

Figure 2.2: Schematic of expected results. Solid horizontal lines: gene cluster bands detected by Southern hybridization following size separation by pulsed-field electrophoresis; Dotted horizontal lines: potentially allowable cluster lengths constrained by the requirement that recombination change cluster lengths by integer multiples of the unit repeat length; Small horizontal arrows: minor-intensity gene cluster bands indicative of sub-populations within the culture possessing cluster lengths that have been altered by non-allelic homologous recombination (NAHR). Adapted from (Killen, Stults et al. 2009).
Figure 2.3: GCI analysis. Panel A: non-clonal cell lines. Panel B: clonal isolates of the U-2 OS population from A). Brackets: multiple bands of uneven intensity from mixed sub-populations in the U-2 OS parental population. Star: resolution limit of the pulsed field gel.
Figure 2.4: GCI analysis of clonal cell lines: Panel A: subclonal isolates from the clonal U-2 OS-A line shown in Figure 3B. Panel B: subclonal isolates from the clonal U-2 OS-B line shown in Figure 3B. Arrows: minor intensity bands indicative of genomic restructuring during the expansion of the indicated subclonal lines. Open triangles: gene clusters found in the subclonal lines but not found in the parental lines. Open brackets: gene clusters not found in the subclonal lines but present in the parental clonal lines. Star: resolution limit of the pulsed field gel.
2.7: REFERENCES


CHAPTER 3: Configuration and rearrangement of the human GAGE gene clusters.


3.1: SYNOPSIS

The GAGE protein is detected only in cancer and in testis and is expressed from a cluster of nearly identical gene copies on the X-chromosome. We determined the lengths of these GAGE gene clusters from human families, identical twins, and in clinical samples from cancer patients. The GAGE cluster lengths proved to be highly heterogeneous, ranging from 13 to 39 gene copies, with an average content of 20 GAGE genes per cluster. Low levels of meiotic rearrangement in families and mitotic rearrangement in adult solid tumors are detectable. Analysis of Rothmund-Thomson syndrome (RTS) kindreds and probands showed GAGE cluster inheritance and stability indistinguishable from that found in non-RTS individuals. These observations support the concept of evolutionarily rapid rearrangement of clustered repetitive sequences in the human genome.

3.2: INTRODUCTION

GAGE is a member of the CT family of genes, defined as having expression only in cancer and in testis (reviewed in (Scanlan, Simpson et al. 2004)). The founding member of the class, MAGE, was identified in melanoma cells using a T cell epitope cloning technique (van der Bruggen, Traversari et al. 1991). Since that time the list of CT genes has expanded to include over 100 members (Almeida, Sakabe et al. 2009). Importantly, many of the protein products of the CT genes have proven to be immunoreactive. Since testis is an immunoprivileged site, this has opened the door to potential CT protein based immunotherapeutic strategies for cancer treatment. Indeed, several early phase clinical trials have already been conducted with promising results (Marchand, van Baren et al. 1999; Davis, Chen et al. 2004; Jager, Karbach et al. 2006; Odunsi, Qian et al. 2007; Uenaka, Wada et al. 2007).
GAGE is a small, acid protein 117 amino acids (12.9 kDa) and has been found in 24% of melanoma samples, 25% of sarcoma, 19% of non-small cell lung cancers, 19% of head/neck tumors and 12% of bladder cancers (reviewed in (Gjerstorff and Ditzel 2008)). GAGE transcripts are seen in melanoma (De Backer, Arden et al. 1999; Eichmuller, Usener et al. 2002), lung cancer (De Backer, Arden et al. 1999), thyroid carcinoma (Ruschenburg, Kubitz et al. 1999; Maio, Coral et al. 2003), breast cancer (Mischo, Kubuschok et al. 2006), hepatocellular carcinomas (Kobayashi, Higashi et al. 2000) and ovarian cancer (30%) (Russo, Dalerba et al. 1996; Gillespie, Rodgers et al. 1998). Clinically, GAGE expression has been demonstrated to correlate with poor prognosis in stomach cancer (Kong, Koo et al. 2004), esophageal cancer (Zambon, Mandruzzato et al. 2001) and neuroblastoma (Cheung, Chi et al. 2000). The GAGE protein has been characterized as having anti-apoptotic activity, conferring resistance to Fas-ligand, taxol and gamma irradiation (Cilensek, Yehiely et al. 2002). GAGE expression also confers cellular resistance to killing by interferon gamma (IFN-γ) through post-translational downregulation of interferon regulatory factor 1 (IRF1) with subsequent downregulation of caspases 1 and 7, and through stabilization of nucleophosmin (Kular, Yehiely et al. 2009).

The GAGE transcript has five spliced exons, the first of which is untranslated. The gene producing the GAGE transcript is multicopy, with a unit gene repeat of 9556 bp. Over one third of all CT genes have multiple gene copies in the genome, with 12 of the individual CT proteins expressed from at least six gene copies each (Hofmann, Caballero et al. 2008). The multiple copies of the GAGE gene are structured as a tandemly repeated cluster of at least 15 gene copies at Xp11.23. The repeated genes in the GAGE cluster are oriented in a head-to-tail manner, without any intervening sequences. The intimately clustered genomic structure is similar in nature to the clustered ribosomal RNA genes (rDNA) we have already characterized (Stults, Killen et al. 2008; Killen, Stults et al. 2009; Stults, Killen et al. 2009), as well as to other gene clusters elsewhere in the genome (Warburton, Hasson et al. 2008).
The GAGE cluster is “young,” being found only in human, chimpanzee and macaque, is under apparent positive selection for protein function and has undergone steady copy number expansion of the last 4 million years (Liu, Zhu et al. 2008). Other CT-genes on the X chromosome are similar: recent evolutionary additions diverging faster and under stronger selection than either CT-genes found on autosomes, or non-CT genes on the X chromosome (Stevenson, Iseli et al. 2007). In this work, we describe a physical assay we developed to measure the number of gene copies in each GAGE cluster. We used this assay with blood samples from human families, human cancer susceptibility (Rothmund-Thomson syndrome) kindreds and probands, and clinical specimens from human cancers to determine the range of human variation of the GAGE cluster, and the stability of the cluster under normal human meiosis and in the genomic pathology of human cancers.

3.3: MATERIALS AND METHODS

High molecular weight genomic DNA was prepared in the solid phase from cells, blood and tissues embedded in agarose as described previously (Killen, Stults et al. 2009; Stults, Killen et al. 2009). All human studies adhere to University of Kentucky protocols set forth in IRB #06-0576-F2L, “Correlation of Human Phenotypic Variation with Polymorphic rDNA Genomic Structure” filed by Dr. Andrew Pierce. Genomic DNA was subjected to restriction digestion with StuI (New England Biolabs) unless otherwise noted, and loaded onto 1% PFC agarose (Bio-Rad) pulsed field gels. Gels were run in 0.5 x TBE (44.5 mM Tris base, 44.5 mM boric acid, 1 mM EDTA pH 8.0) for 20 hours with 6 V/cm electrical field vectors at 120º separation, with a switching frequency that varied linearly from 10" to 40", at 14°C. Following electrophoresis, gels were equilibrated to 0.5% final concentration glycerol, dried, rehydrated and Southern blotted as described previously (Stults, Killen et al. 2008).

Southern probe templates specific to the GAGE cluster were isolated by PCR from human genomic DNA either with the primer set 1: GAGE-1F: GTCCTCCTTCCCTTCACAGG and GAGE-1R: TTCTCGTGATTGCTGCTTTG or
with primer set 2: GAGE-2F: AGACCCAGTTCAGAGGAGCA and GAGE-2R: CGTGAAGAAACAAAGCACCAA.

Radiolabeled Southern probes were generated by PCR amplification of the template products derived from human genomic DNA described above using the appropriate primer sets. PCR conditions were:

<table>
<thead>
<tr>
<th>Reaction component</th>
<th>PCR program</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0 TP @ 40 µM each</td>
<td>94°C 30 sec</td>
</tr>
<tr>
<td>1.0 µl dATP @ 20 µM (not radioactive)</td>
<td>65°C 30 sec</td>
</tr>
<tr>
<td>0.4 µl TAQ (2U)</td>
<td>72°C 1 min + 2 sec/cycle</td>
</tr>
<tr>
<td>3.6 µl H₂O µl TAQ buffer (10x)</td>
<td>• 40 cycles total</td>
</tr>
<tr>
<td>2.0 µl PCR product @ 10 pg/µl</td>
<td>72°C 7'</td>
</tr>
<tr>
<td>2.0 µl forward primer @ 1 µM</td>
<td>hold at 4°C</td>
</tr>
<tr>
<td>2.0 µl reverse primer @ 1 µM</td>
<td></td>
</tr>
<tr>
<td>5.0 µl α³²P-dATP (50 µCi @ 3000 Ci/mmol)</td>
<td></td>
</tr>
<tr>
<td>2.0 µl d(GTC)</td>
<td></td>
</tr>
</tbody>
</table>

3.4: RESULTS

Our experimental strategy is shown in Fig. 3.1A. For tandemly repeated genes with very high sequence identity, the gene cluster can be liberated from bulk genomic DNA by digestion with a restriction enzyme that lacks a recognition site in the unit gene repeat. The liberated cluster can be size-resolved by pulsed field gel electrophoresis, detected though Southern hybridization, and the number of gene copies inferred from the length of the cluster with a small correction for flanking non-cluster DNA. Each repeat of the GAGE gene is 9556 bp. We initially selected two different Southern probes with predicted minimal cross-reactivity to the reference human genome, and three different restriction enzymes, AvrII, StuI and Swal to liberate the GAGE cluster from bulk genomic DNA (Fig. 3.1B). Anticipating that evolutionarily rapid GAGE cluster restructuring due to homologous recombination would cause a diversity of cluster lengths in the human population, we assayed genomic DNA from two human females, with the
prediction that each X-chromosome would have a different sized GAGE cluster and thereby show two different length clusters, in a manner similar to the 5S rDNA gene cluster at 1q42 (Stults, Killen et al. 2008). We were able to resolve these two predicted GAGE clusters (Fig. 3.1C) using either AvrII or StuI (SwaI failed to efficiently digest the agarose-embedded genomic DNA), and with either Southern probe. The three bands seen with AvrII digestion in the FASHY individual suggests that the GAGE repeats do not share perfect sequence identity with each other, such that a cryptic AvrII site can exist within the cluster, a phenomenon we also observed in analysis of the 5S rDNA clusters. The non-GAGE flanking sequence according to the Human Genome Project is 3859 bp when the clusters are liberated by AvrII digestion (2828 bp telomeric and 1031 bp centromeric with respect to the GAGE cluster) and 9781 bp when liberated by StuI digestions (537 bp telomeric and 9244 bp centromeric with respect to the GAGE cluster). Consistent with the predicted smaller amount of flanking sequence after cluster liberation by AvrII relative to StuI, we observed that the GAGE clusters identified by Southern blotting are slightly shorter when liberated by AvrII than by StuI (Fig. 3.1C). After verifying the DNA sequence of probe 2, we selected StuI and probe 2 as an optimal combination for further experiments.

To establish the range of normal human variation in GAGE cluster lengths and to determine the degree of meiotic cluster rearrangement, we analyzed the GAGE cluster lengths in human families (Fig. 3.2A). We found strong heterogeneity in the lengths of the GAGE cluster ranging from 135 kb up to 380 kb, corresponding to from 13 to 39 gene copies, after subtracting the predicted 9.7 kb of non-GAGE flanking DNA, with an average cluster length of approximately 20 GAGE repeats. Females generally show two distinct GAGE clusters and males only one, consistent with the X-linkage of the cluster. We observed primarily predicted patterns of Mendelian inheritance without rearrangement, with the exception of one potential meiotic recombination event in the female germline of the PYFI family (Fig. 3.2A, blue dotted circle). To further establish the reproducibility of the assay, we determined GAGE cluster lengths in 10 sets of female identical twins. In all cases, the twins shared the identical GAGE cluster configuration (Fig. 3.2B).
Next we sought to determine whether the GAGE cluster was stable under potentially elevated recombination rates in human cancer. We compared the GAGE cluster configuration in individual human lung and colorectal tumors to matched control DNA from the same individuals isolated from either peripheral blood or surgically excised pathologically confirmed non-tumor tissue. In one male lung cancer (Fig. 3.3A, ‘XEPRY’) we observe evidence of mitotic rearrangement generating a new GAGE cluster length in the tumor not seen in the surrounding non-tumor tissue. We observed the same mitotic rearrangement phenomenon in one of the colon cancers (Fig. 3.3C, ‘DAPEB’). Otherwise the GAGE cluster configuration in tumors was comparable to that in normal tissues, with the exception of loss of one GAGE cluster in a female lung cancer (Fig. 3.3B, ‘TIPOR’), likely due to X-chromosome aneuploidy. The two confirmed mitotic events in these 36 solid tumor sets would indicate a detectable rearrangement frequency of around 5%.

In order to assess potential GAGE restructuring in pediatric leukemia, in addition to the adult solid tumors described above, we compared the GAGE cluster genomic architecture between peripheral blood containing blasts and peripheral blood from the same patients after their disease was in remission. We detected no differences in the GAGE cluster configurations between disease-containing and disease-free blood sample from eight different pediatric patients (Fig. 3.4).

Finally we sought to extend our analysis of GAGE architecture to families with hereditary cancer syndromes. We assayed GAGE architecture in EBV-immortalized lymphocytes from eight kindreds containing Rothmund-Thomson syndrome Type 2 probands, and from four Rothmund-Thomson syndrome Type 1 probands (Fig. 3.5). We found that GAGE cluster lengths segregated with in the normal Mendelian manner in the kindreds, and that the immortalized cells from Rothmund-Thomson probands displayed well-defined stable banding patterns.
3.5: CONCLUSION

We found that the normal human variation in GAGE gene copy number is from 13 to 39 GAGE gene repeats on each X chromosome. This is considerably less variation that we had earlier found in the ribosomal RNA genes that varied from 1 to over 140 copies per cluster (Stults, Killen et al. 2008). The GAGE clusters also recombine with reduced frequency relative to the rDNA clusters. There are several plausible factors that may account for this reduced frequency of rearrangement: the GAGE unit repeat at 9.6 kb is shorter than the 43 kb rDNA repeat and there are only one or two GAGE clusters per genome (in males and females respectively) compared to the ten rDNA clusters. Nevertheless, the high degree of inter-person GAGE cluster heterogeneity indicates that the GAGE clusters are restructuring rapidly in evolutionary time, even if less rapidly than the rDNA. In a similar manner, although we can detect GAGE cluster alterations in adult solid tumors at a frequency of around 5%, this is considerably less than the greater than 50% frequency of rDNA restructuring we observed in human adult solid tumors. Interestingly, both of the cancer patients in which we detected GAGE cluster restructuring also exhibited rDNA cluster restructuring in their disease, suggesting a potential mechanistic link between the two processes (Stults, Killen et al. 2009).

Rothmund-Thomson syndrome arising as a result of mutations in RECQL4 (RTS type 2) confers an increased cancer risk, particularly of osteosarcoma (Siitonen, Sotkasiira et al. 2009). As RECQL4 is a member of the RecQ family of proteins associated with altered replication and recombination phenotypes (Bohr 2008), we hypothesized that a potentially recombination-prone genomic locus like GAGE might display accelerated instability in lines defective for RECQL4. Counter to this hypothesis, we found that cells from Rothmund-Thomson probands either with RECQL4 defects (RTS type 2) or without RECQL4 defects (RTS type 1) display well-defined discrete banding patterns with no overt evidence of instability. GAGE clusters inherited from parents heterozygous for RECQL4 mutation segregated in the normal Mendelian manner. The average human GAGE gene cluster length exhibits strong heterogeneity, varying from 13 to 39 gene copies, with an average of approximate 20 copies per X-chromosome. Both meiotic and mitotic alterations to GAGE cluster length occur at detectable frequencies.
Figure 3.1: Human GAGE gene clusters
A) Experimental schematic showing a model GAGE cluster containing six gene copies. Restriction digestion with enzymes that do not have a recognition site in the unit gene repeat will liberate entire gene clusters (straight blue lines), with a small amount of additional flanking-sequence DNA (wavy blue lines). Open rectangles: unit GAGE clusters; double black lines: non-GAGE genomic DNA. B) GAGE unit gene repeat. Transcription proceeds clockwise from the indicated transcription start site. Exons: thick green bars; introns: thin green bars; high copy genomic repeats: pink regions and thin arrows; unique genomic sequences: black bars. The positions of Southern blot probe 1 and probe 2 are indicated by red arrows. C) Southern blot of GAGE clusters liberated from genomic DNA by the indicated restriction enzymes and detected with the indicated Southern probe. Digests were overnight: AvrII: 12 units at 37C, StuI: 20 units at 37C, SwaI: 20 units at room temperature. Genomic DNA isolated from the peripheral blood from two different unrelated female donors (FASHY, XAVIL). Open star: unresolved DNA due to inefficient digestion with the SwaI enzyme.
Figure 3.2: GAGE clusters in families and identical twins.
A) Families (all names are coded). Circles: females, squares: males. A potential meiotic recombination event (dotted blue oval) is shown in the first PYFI son (blue square).
B) 10 sets of female identical twins.
Figure 3.3: GAGE clusters in adult solid tumors.
A) Lung cancer in males. B: blood, T: tumor, N: non-tumor surrounding tissue. A mitotic recombination event is shown in the XEPRY tumor sample (red arrow).
B) Lung cancer in females: B: blood, T: tumor, N: non-tumor surrounding tissue. Loss of one allele seen in the peripheral blood of TIPOR is indicated (red pushpin).
C) Colorectal cancer. B: blood; T: tumor; N: non-tumor surrounding tissue. A mitotic recombination even that altered GAGE cluster length is seen in the DAPEB tumor samples (red arrow).

Figure 3.4: GAGE clusters in pediatric leukemia.
Peripheral blood either with or without blasts is analyzed for GAGE clusters by Southern blotting. No alterations between the blasts and the normal cells was detected.
Figure 3.5: GAGE clusters in Rothmund-Thomson kindred and probands. Circles: females, squares: males. R-T probands indicated in red.
3.7 REFERENCES


Copyright © Michael Wayne Killen 2011
CHAPTER 4: LOSS OF BLOOM SYNDROME PROTEIN DESTABILIZES HUMAN GENE CLUSTER ARCHITECTURE*


doi: 10.1093/hmg/ddp282

4.1: SYNOPSIS

Bloom syndrome confers strong predisposition to malignancy in multiple tissue types. The Bloom syndrome patient protein (BLM) defective in the disease biochemically functions as a Holliday junction dissolvase and human cells lacking functional BLM show 10-fold elevated rates of sister chromatid exchange. Collectively, these phenomena suggest that dysregulated mitotic recombination drives the genomic instability underpinning the development of cancer in these individuals. Here we use physical analysis of the highly repeated, highly self-similar human ribosomal RNA gene clusters as a sentinel biomarker for dysregulated homologous recombination to demonstrate that loss of BLM protein function causes a striking increase in spontaneous molecular level genomic restructuring. Analysis of single-cell derived sub-clonal populations from wild-type human cell lines shows that gene cluster architecture is ordinarily very faithfully preserved under mitosis, but is so unstable in cell lines derived from BLMs as to make gene cluster architecture in different sub-clonal populations essentially unrecognizable one from another. Human cells defective in a different RecQ helicase, the WRN protein involved in the premature aging Werner syndrome, do not exhibit the gene cluster instability (GCI) phenotype, indicating that the BLM protein specifically, rather than RecQ helicases generally, holds back this recombination-mediated genomic instability. An ataxia-telangiectasia defective cell line also shows elevated rDNA GCI, although not to the extent of BLM defective cells. Genomic restructuring mediated by dysregulated recombination between the abundant low-copy repeats in the human genome may prove to be an important additional mechanism of genomic instability driving the initiation and progression of human cancer.
4.2: INTRODUCTION

Early onset of multiple malignancies is a clinical hallmark of Bloom syndrome, with mean age of death in Bloom syndrome patients (BS) younger than 24 years (German 1997) suggesting a critical loss of genomic stability. Cytologically, cells from BLMs show a 10-fold elevation in rates of sister chromatid exchange (SCE) when differentially stained following incorporation of various thymidine analogs, as well as a spontaneous increase in chromosomal abnormalities such as quadradial structures (German, Archibald et al. 1965; Chaganti, Schonberg et al. 1974). Although dramatic in appearance when visualized in the standard assay for SCE (Perry and Wolff 1974), exchanges between sister chromatids are genetically silent, and therefore not directly contributory to genomic instability per se. Nevertheless, the class of genomic alteration represented by SCE, the physical relocation of genetic material, without associated change in either the amount or sequence of this material, is an important but poorly examined mechanism of chromosomal alteration, particularly when these alterations are in the sub-microscopic size range. In this report, we present a sensitive assay of ribosomal gene cluster instability (GCI) that demonstrates exactly this kind of non-silent molecular level alterations to chromosomal architecture resulting from the absence of the Bloom syndrome protein. We also use this assay to show that unlike loss of the Bloom syndrome protein, neither loss of the BLM-related Werner protein nor inactivation of either transcription-coupled nucleotide excision repair or non-homologous end-joining (NHEJ) appreciably increase this GCI.

The BLM protein inactivated in BS is a member of the RecQ helicase family of helicases (Ellis, Groden et al. 1995). In humans, there are five known RecQ homologs implicated in the maintenance of genomic stability (reviewed in (Bohr 2008)). Mechanistically, the BLM protein, in conjunction with binding partners RMI1, RMI2 and TOPOIIα (Raynard, Zhao et al. 2008; Xu, Guo et al. 2008), can function as a Holliday junction dissolvasome (Raynard, Bussen et al. 2006; Wu, Bachrati et al. 2006) that reverses the strand exchanges in a double Holliday junction structure to prevent potential junction resolution involving crossover products. In the absence of sufficient BLM-
complex activity, such recombination intermediates may be subject instead to resolution by the human Holliday junction resolvase GEN1 (Ip, Rass et al. 2008), with the risk of production of both crossover and non-crossover products.

Although crossover recombination between precisely aligned sister chromatids is not inherently destabilizing, the human genome contains abundant low-copy repeats (LCRs) with sufficiently high levels of sequence similarity as to be susceptible to non-allelic homologous recombination (NAHR) (Consortium 2004). Structural alterations to genomic architecture caused by NAHR between LCRs include deletions, inversions and translocations with the possible generation of both acentric and dicentric chromosomes (reviewed in (Gu, Zhang et al. 2008)). Although the consequences of recombination-mediated genomic restructuring can be dire, spontaneous occurrences of NAHR are rare, hindering the identification of genetic components important for suppressing this class of genomic instability. Early work with BLM defective cells gave phenotypic evidence that the BLM protein could suppress NAHR in repetitive genomic loci (Kyoizumi, Nakamura et al. 1989; Langlois, Bigbee et al. 1989), later supported by molecular evidence of loss of heterozygosity between parentally homologous chromosomes (Groden, Nakamura et al. 1990; Ellis, Lennon et al. 1995), and in instability of a 40 bp repetitive mini-satellite sequence (Groden and German 1992). We now extend this work to much larger gene clusters distributed on multiple chromosomes.

Large gene clusters possess attributes predicted to be conducive to recombinational alteration. The repeated genes that make up clusters tend to have very high levels of sequence identity, sufficient length to be substrates for recombination, and occur in high local concentration with respect to each other. The gene clusters encoding the 45S precursor transcript to the 18S, 5.8S and 28S ribosomal RNA molecules (collectively the ‘rDNA’) are perhaps the largest clustered gene arrangement in the human genome, with approximately 600 repeats of the unit 43 kb gene (Gonzalez and Sylvester 1995) divided among the five pairs of acrocentric chromosomes (Henderson, Warburton et al. 1972). We recently characterized the physical lengths of these rDNA gene clusters in human individuals and found essentially complete cluster length heterozygosity, both between clusters on different acrocentric chromosomes and also between clusters on individual parental homolog chromosome pairs (Stults, Killen et al. 2008). This cluster length
heterogeneity is driven by strong meiotic recombination at a rate of over 10% per cluster per meiosis. We were also able to detect spontaneous mitotic rearrangement in these clusters, suggesting that they would be suitable as sentinel biomarkers for dysregulated recombination. Accordingly, we screened a panel of cell lines representing diverse human chromosomal instability syndromes for hyper-recombination in the rDNA gene clusters and used both shRNA knockdowns and complemented mutant cell lines to establish that the BLM protein is a crucial suppressor of GCI.

4.3: RESULTS

We assay dysregulated recombination by ascertaining the stability of the length of gene clusters. The lengths are determined by restriction digestion of high molecular weight genomic DNA with enzymes that do not have a recognition site in the repeated gene unit. Since the repeated genes in clusters are generally highly conserved, an enzyme that does not cut in one repeat tends not to cut in any and can thereby liberate intact gene clusters from bulk genomic DNA, with the addition of a small amount of randomly sized flanking DNA (Fig. 4.1A). Liberated gene clusters can then be separated by size using pulsed-field gel electrophoresis and identified by Southern blotting. The goal is to utilize this pulsed-field based assay to quantify the GCI in human cells. In any cell population derived from a single cell, including human individuals, genomic gene clusters will have a well-defined initial length. We call the size-resolved pattern of these initial cluster lengths the ‘major banding pattern’ (Fig. 4.1B, ‘Initial Pattern’). As the cell population expands, if the gene clusters are completely stable, the initial cluster lengths found in the progenitor cell will be faithfully transmitted to all subsequent daughter cells (Fig. 4.1B, ‘No GCI’). Alternatively, recombination in the expanding population can generate sub-populations with altered gene cluster lengths. Since these sub-populations only represent a fraction of the total population, bands detected by Southern blotting will be reduced in intensity accordingly. We call these reduced-intensity bands the ‘minor banding population’ (Fig. 4.1B, ‘Low GCI’). The amount of this minor-intensity banding found in any cell population is indicative of the degree of GCI in that population. Since recombination requires precise alignment of homologous sequences, cluster lengths can only change by integer multiples of the unit repeat length. This constraint upon allowable
gene cluster lengths (Fig. 4.1B, dotted lines) means that very high levels of instability will generate a ladder-like pattern of bands in a Southern blot (Fig. 4.1B, ‘High GCI’), consistent with a recombination-based mechanism. If cluster length alterations were due to random breakage and rejoining, a smear would be observed, rather than a ladder. We screened a panel of cell lines including various known human chromosomal instability syndromes (Fig. 4.1C). Stable gene clusters result in a well-defined pattern of major bands, with largely empty intervening space and few minor bands. As expected, since every individual human has a unique pattern of major bands (21), the human cell lines also possess unique patterns of major bands. Instability will be indicated by the presence of minor-intensity bands. Four lines wild-type for major DNA repair pathways, BJ-5ta, CGM1, GM06990 and HeLa S3 (left side of the panel), largely exhibit such gene cluster stability, with only a few minor-intensity bands observed. BJ-5ta is immortalized by the expression of telomerase, CGM1 and GM06990 by EBV and HeLa S3 is a cultured cancer line, collectively indicating that cellular transformation and immortalization per se is generally not destabilizing to the faithful transmission of gene cluster lengths. Of the SV40-transformed lines from a panel of patients exhibiting defects in a variety of DNA repair pathways (right side of the panel), only that derived from a BLM is strongly destabilized (rounded box), with a clear ladder-pattern consistent with integer multiple differences of 43 kb: the size of the repeated rRNA genes. In contrast, the wild-type, DNA ligase I-defective, DNA ligase IV-defective and Nijmegen breakage syndrome lines appear to be gene cluster stable. The ATM cells show minor banding consistent with a GCI-low phenotype; the HeLa and ATM cells are explored in more detail in Figure 4.7.

The basic GCI screen (Fig. 4.1) measures the amount of GCI that has occurred in the history of a given cell population rather than the rate at which GCI occurs. In this context, we can understand the presence of minor bands in the GCI profile of a cell population as being the result of a gradual cumulative stochastic process of recombination-mediated alterations from the initial major banding pattern subjected to genetic drift. The fraction of cells in a cultured population that has gene cluster lengths not represented in the major banding pattern is a function of the inherent GCI rate of the cells in question, the total
number of cell divisions undergone by the culture, and the degree to which the culture
has been subjected to population bottlenecks.

We can uncover the GCI rate by unrestricted expansion of single-cell derived sub-
clones. This procedure, in essence Luria–Delbruck fluctuation analysis (Luria and
Delbruck 1943), is sensitive only to the GCI rate since freely expanded sub-clonal
populations experience the same number of cell divisions and no population bottlenecks.
Sub-cloning cell populations from single cells can be applied iteratively: minor bands in
any parental population will produce new major bands in derivative sub-clonal
populations with a frequency of occurrence determined by the fractional intensity of the
minor bands in the parental population. New spontaneous GCI in any sub-clonal
population will appear as de novo minor bands. This clonal GCI analysis of the wild-type
GM00637 cell line (Fig. 4.2A) shows relatively stable gene cluster lengths. The lane ‘WT
clonal’ is derived from a single cell. Single-cell derived sub-clones from this clonal
population are shown in the right of the panel. Changes to the major banding pattern in
the sub-clones are seen in sub-clones ‘C’ and ‘H’ (triangle and brackets indicating new
major band and missing major bands, respectively). These represent mitotic gene cluster
alterations occurring in the expansion of the ‘WT clonal’ cells that occurred too late in
the expansion to yield visible minor bands in the ‘WT clonal’ lane. A minor band (arrow)
is seen in sub-clone ‘E’, from a spontaneous gene cluster length alteration in the
expansion of this individual sub-clone. We can reliably detect minor bands with an
intensity of <10% of the major bands. For 10% of the population to show an altered
gene cluster length, the alteration had to occur sometime in the initial three or four
rounds of cell division, i.e. before the population derived from a single cell had
expanded to a total of 10 cells. In contrast to the rare minor banding of the wild-type
population, the Bloom syndrome line GM08505 (Fig. 4.2B) shows such extensive minor
banding that for practical purposes no major banding pattern of rDNA gene clusters can
be defined. Gene cluster lengths in the BLM cells are changing on the order of every
mitotic division.

To establish that the genetic defect causing the GCI-high phenotype in the BLM-
syndrome GM08505 cells is lack of BLM, we surveyed BLM lines derived from three
other individual BLMs (Fig. 4.3A and 4.3B). The bulk wild-type cell populations appear
stable with well-defined major bands and little or no minor banding evident. In contrast, each of the three BLM-patient derived lines, BLM Ash.1, BLM Ash.2 and BLM F.C., shows the ladder-pattern indicative of the GCI-high phenotype. BLM Ash.1 and BLM Ash.2 are from patients of Ashkenazi Jewish descent, each homozygous for the 6 bp deletion/7 bp insertion frameshifting Ashkenazi founder mutation. BLM F.C. is from a French-Canadian Bloom patient and is homozygous for a S595X translation terminating mutation. All three lines show a distinct ladder-pattern of minor bands (rounded boxes). In order to determine whether the gene cluster destabilization preceded either transformation or tumorigenesis, we analyzed cells from primary fibroblast explants (Fig. 4.3B). Primary cells from an apparently healthy 11-year-old girl (GM10652) show well-defined major bands, and no minor banding, whereas primary cells from a 28-year-old male BLM (GM02932) show minor band laddering, indicating that GCI is inherent to Bloom syndrome cells generally and therefore is a potential contributor to the genomic instability that initiates tumor development in these patients. The contrast-enhanced right-hand panel allows direct comparison of the spacing of the BLM gene cluster laddering with the well-defined length lambda phage concatemers. The BLM ladder (arrows) shows slightly closer spacing than the lambda ladder, consistent with the 43.0 kb rDNA unit gene repeat versus the 48.5 kb lambda genome size.

Complementation of the BLM defect restores rDNA gene cluster stability. Minor band laddering is seen in the bulk populations of two BLM derived lines: BLM-comp and BLM-vec (Fig. 4.3A). These non-clonal lines are the product of transfection and stable selection of BLM-defective BLM Ash.1 cells with either the pcDNA3 vector expressing full-length wild-type BLM or the pcDNA3 vector alone, respectively (Gaymes, North et al. 2002). Since these are non-clonal lines, the BLM-comp population shows the minor banding pattern of instability that was generated during the BLM-null history of this line. The current stability of the BLM-comp line is revealed by clonal analysis (Fig. 4.4A). As expected, the minor band laddering in the BLM-comp bulk population resolves as variability in the pattern of major bands in the BLM-comp clonal isolates. The clonal isolates themselves, however, contain a greatly reduced amount of de novo minor banding relative to the BLM-vec non-complemented isogenic line (Fig. 4.4B), indicating that restoration of the BLM protein has stabilized the rDNA gene clusters.
Since the BLM-comp cells are not completely complemented for the elevated SCE phenotype of Bloom syndrome cells (Gaymes, North et al. 2002) (Figure 4.9) and still show some minor banding indicative of ongoing GCI in the clonal sub-lines (Fig. 4.4A), we wanted to verify the importance of the BLM protein in promoting gene cluster stability by assaying loss of stability in an otherwise stable line upon loss of the BLM protein. Accordingly, we used shRNA (Stegmeier, Hu et al. 2005) to knockdown expression of BLM in a clonal wild-type GM00637 cell population. We used a ‘semi-stable’ knockdown technique where the selectable shRNA expression plasmid was transfected and selected for 7 days to kill untransfected cells, followed by dilution plating to single cells and unselected sub-clonal expansion. We used this semi-stable technique to prevent eventual overgrowth by cells losing shRNA expression. After 7 days of selection, BLM levels measured by western blotting were decreased by 70% (Fig. 4.5C), at which time sub-clonal lines were rederived from single cells. Sub-clonal lines from the clonal cell population transfected with the negative-control shRNA construct pCPM-neg show only limited gene cluster length alterations (Fig. 4.5A: arrows, brackets and triangles), whereas sub-clonal lines from the same clonal cell population transfected instead with the BLM knockdown shRNA construct pCPM-234 show an approximately 3-fold greater number of gene cluster length alterations (Fig. 4.5B), again implicating the expression of BLM as a critical stabilizer of these rDNA gene clusters.

Having established a role for BLM in the maintenance of gene cluster genomic integrity, we wanted to determine whether RecQ helicases other than BLM were also involved. We performed clonal GCI analysis on SV40-immortalized fibroblasts from a Werner syndrome patient (Fig. 4.6) homozygous for a truncating R368X mutation in the WRN gene, a RecQ helicase defective in a strong progeroid syndrome (Yu, Oshima et al. 1996). We see no minor bands in a clonal derivative of these WRN-deficient AG11395 cells, and neither changes to the major banding pattern nor new minor bands in single-cell derived sub-clonal populations derived from the clonal cells. We conclude that the suppression of rDNA GCI is not a property of the RecQ helicase family generally, although we have not yet assayed cells deficient in the remaining RecQ helicases: RECQL, RECQL4 and RECQL5.
An ataxia-telangiectasia mutated cell line has more GCI than HeLa cells. The minor banding we observed in a bulk population of ATM-deficient cells (Fig. 4.1C) motivated us to use clonal analysis to determine whether there was ongoing GCI in this line. Starting with a clonal isolate each of the apparently gene cluster stable HeLa cells and the apparently gene cluster unstable ATM cells, we generated sub-clonal lines for analysis. In ten sub-clonal HeLa lines (Fig. 4.7A), we observed one altered major band (triangle) and two new minor bands (arrows) showing the general stability of the HeLa rDNA gene cluster architecture. In contrast, in eight ATM-deficient sub-clones (Fig. 4.7B), we found one new major band (triangle) and seven newly generated minor bands (arrows), suggesting an elevated rate of GCI in the ATM-deficient line.

We also found that defects in transcription-coupled nucleotide excision repair do not greatly enhance GCI. The Cockayne's syndrome B (CSB) protein ERCC6 required for transcription-coupled nucleotide excision repair (Hanawalt and Spivak 2008) has been implicated in the stability of gene clusters with the observation that defects in CSB cause aberrant metaphase figures at gene clusters (Yu, Fan et al. 2000). Therefore to ascertain whether a CSB defect obviously compromised stability of the rDNA gene clusters, we generated a clonal population of CSB defective (GM00739) cells. This clonal population ‘CSB clonal’ exhibited one visible minor band (Fig. 4.8A, arrow) that is reflected as a major banding pattern gain of this cluster length in four of twelve sub-clones, ‘D’, ‘F’, ‘H’ and ‘L’ (triangles). Since these four sub-clones also all lose the same major band (bracket), a reasonable interpretation is that around the third mitosis in the expanding ‘CSB clonal’ population, a spontaneous recombination event in one of the cells caused a gain of two repeats in the cluster indicated by a closed circle to generate the new longer cluster indicated by an arrow. We also observe two additional major band changes (triangles) in sub-clones ‘F’ and ‘H’ that do show a corresponding visible minor band in the parental population at the open circle indicators. Evidently, the events that lead to these major banding changes happened late enough in the parental population expansion that the fraction of cells carrying the change is below the limit of detection for a minor band in the ‘CSB clonal’ population. Nevertheless, the number of minor and major banding alterations in the CSB-defective cells is not greatly larger than that seen in similar clonal analysis with either HeLa (Fig. 4.7A) or wild-type SV40-transformed
fibroblasts (Fig. 4.2A), leading us to conclude that the CSB-dependent repair pathway is not significantly involved in rDNA gene cluster stability.

We similarly detect no large increase in instability with loss of the NHEJ pathway. In experiments that quantify double-strand break mediated non-crossover gene conversion, loss of critical NHEJ proteins such as XRCC6, XRCC5, PRKDC (Ku70/80/DNA-PKcs) or XRCC4 cause an increase in recombination of between 3-fold and 5-fold (Pierce, Hu et al. 2001). Nevertheless, a non-clonal population of cells derived from a patient with a defect in DNA ligase IV, the obligate heterodimeric partner of XRCC4 (Wu, Frit et al. 2009), appears gene cluster stable, with one clearly defined resolved major band and no apparent minor banding (Fig. 4.1C). To confirm this apparent stability, we compared rDNA banding patterns in clonal human isogenic NALM-6 cells, and NALM-6 cells in which DNA ligase IV was disrupted by gene targeting (Iiizumi, Nomura et al. 2006). In the wild-type NALM-6 cells, we observe one minor band in the non-clonal bulk population (Fig. 4.8B, arrow) that gave an altered major band in sub-clone ‘C’ of the three sub-clonal NALM-6 lines, likely the result of loss of one repeat from the gene cluster indicated by the closed circle. We likewise see a major band change in NALM-6 clone ‘B’ from a late event arising from a small sub-population of the bulk parental line (open circle). In the DNA ligase IV knockout NALM-6 cells, we see one major band alteration in sub-clone ‘C’, again arising from a small sub-population in the ‘LIG4 ko bulk’ population. In none of the six clonal lines from either NALM-6 or NALM-6 lines knocked out for DNA ligase IV do we see minor bands, indicating that rapid ongoing recombination-mediated rearrangement of rDNA gene clusters is not occurring, and ruling out DNA ligase IV as a strong player in maintaining rDNA gene cluster stability.

4.4: DISCUSSION

The GCI assay is philosophically similar to the microsatellite instability (MSI) assay for defective mismatch repair. In MSI+ lines, polymerase slippage at repetitive mono-, di- and tri-nucleotide short tandem repeats (STRs) generates expansions and contractions of the STR length, which are not subsequently corrected by mismatch repair (reviewed in (Laghi, Bianchi et al. 2008)). In contrast, the repeated sequences assayed here for GCI are thousands of base pairs in length, requiring a completely different underlying
biochemistry for stability than the microsatellites. We propose that the mechanism causing GCI is dysregulated homologous recombination, where the usual strong bias in favor of short-tract gene conversion without crossing over (Paques, Leung et al. 1998) has been lost or attenuated. The kind of crossover recombination reactions capable of producing the GCI we observe in the BLM-deficient cells is ordinarily highly suppressed. The known biochemical ability of BLM and partner proteins to dissolve double Holliday junctions, in combination with the ladder-like pattern of instability we describe here, is consistent with a crossover recombination mechanism for gene cluster destabilization in the absence of BLM. Nevertheless, we cannot rule out the possibility that gene cluster length alteration is a product of break-induced replication (BIR) (reviewed in (Llorente, Smith et al. 2008)), either in whole or in part. Indeed, a microhomology mediated break-induced replication model has been recently proposed as a mechanism for the generation of many human genomic copy number variations (Hastings, Lupski et al. 2009).

It is also formally possible that in the absence of BLM, the rDNA gene clusters we assay in this work are being broken and then rejoined by a translocation-like mechanism driven by NHEJ. Since NHEJ does not require alignment of homologous sequences as does recombination, generation of the single repeat 43 kb changes in gene cluster length we observe would additionally require the gene clusters to break at one specifically defined sequence within a unit repeat. Conceivably some kind of BLM-suppressed chromosome fragility structure could produce this kind of sequence-specific breakage pattern, potentially quadruplex-G sequences (Huber, Duquette et al. 2006) or hemicatenated replication structures (Fricke and Brill 2003). The lack of a significant increase in GCI in the absence of DNA ligase IV, however, argues that if rearrangements are initiated by frank chromosomal breaks, NHEJ is not a competing pathway for these breaks.

Accordingly we do not favor the notion of frank double-strand breaks as the primary initiating events in rDNA GCI but prefer the previously suggested idea (Pierce, Hu et al. 2001) that recombination can be initiated from one-sided double-strand ends arising from collapsed replication forks. It is attractive to speculate that this putative replication fork collapse is caused by the formation of quadruplex-G sequences in the G/C-rich rDNA, since the BLM protein efficiently unwinds quadruplex-G DNA (Sun, Karow et al. 1998). It will be interesting to determine whether loss of the BRIP1 (FANCJ) protein, which has
a similar ability to unwind quadruplex-G (London, Barber et al. 2008), will likewise cause rDNA GCI.

The human BLM protein has been characterized as facilitating the EXO1 mediated 5′–3′ resection of double-strand breaks prior to allowing the loading of the RAD51 recombinase (Gravel, Chapman et al. 2008; Nimonkar, Ozsoy et al. 2008). The highly elevated recombination phenotypes observed in the absence of BLM, however, argue against the requirement for this activity of BLM in either rDNA gene cluster recombination or SCE. Likewise, human BLM also possesses a single-stranded DNA (ssDNA) annealing activity (Cheok, Wu et al. 2005; Machwe, Xiao et al. 2005) that appears uninvolved in these particular recombination processes. In contrast, BLM also possesses an anti-recombinagenic capacity to remove invaded 3′ ssDNA tails from D-loops (van Brabant, Ye et al. 2000; Bachrati, Borts et al. 2006), including the disruption of a pre-formed RAD51 filament (Bugreev, Yu et al. 2007). Loss of this BLM activity likely contributes to elevated recombination phenotypes in Bloom syndrome cells.

We assayed the repeated 43 kb genes comprising the rDNA gene clusters in the belief that these clusters were the most likely to show spontaneous recombination-mediated alterations due to their abundance, length, degree of sequence conservation and number of potentially interacting chromosomes. Intriguingly, the size of chromatin loops in HeLa cells has been characterized as averaging 86 kb (Jackson, Dickinson et al. 1990), which is the exact length of two rDNA genes and may be important in the propensity of recombination to alter rDNA gene cluster lengths. Looped chromatin generally may be an important intermediate in gene cluster recombination, and the tendency for chromatin loop domains to be long may protect shorter gene clusters from intramolecular recombination. It will be interesting to determine if the spontaneous gene cluster reorganization seen here in the rDNA can also be observed in smaller, less abundant non-rDNA gene clusters (Warburton, Hasson et al. 2008), although we would predict having fewer potentially recombining gene clusters of shorter overall length would decrease the ability to sensitively detect alterations in other gene cluster loci. Consistent with a role in protecting particularly susceptible classes of gene clusters such as the rDNA from recombination, BLM protein sub-cellular localization includes the nucleolus (Sanz, Proytcheva et al. 2000), possibly for the specific purpose of counteracting the potential
for recombination to destabilize the rDNA gene clusters that are also located in the nucleolus. Seen in this light, the nucleolus may function specifically as a recombination-suppression sub-nuclear zone. It will also be of interest for future studies to determine whether chemical agents known to induce SCE lead to a parallel increase in GCI, and whether the GCI assay can be a useful measure of sub-microscopic genomic toxicity.

To a first-order approximation, by comparing the number of observed minor bands per clonal isolate, we find the rate of spontaneous alterations in gene cluster architecture to be over 100-fold elevated in cells lacking BLM (9.6 minor bands per clone) and 10-fold elevated in cells lacking ATM (0.9 minor bands per clone) compared with wild-type controls (0.08 minor bands per clone). It is intriguing that the rDNA GCI we observe in BLM and to a lesser extent in ATM cells parallels the increased cancer predisposition in Bloom syndrome and ataxia-telangiectasia patients. Particularly, because crossover recombination between low-copy human genomic repeats has the potential to generate enormous genomic instability through formation of dicentric and acentric chromosomes, it remains to be seen whether or not GCI, whether caused by functional loss of BLM or otherwise, is a common mechanism of genomic instability driving the etiology and progression of human cancer.

4.5: METHODS

Cell lines

Cell lines GM00637 (wt), GM09607 (ATM), GM08505 (BLM), GM13136 (FANCC), GM16097 (LIG1), GM16089 (LIG4), GM15989 (NBN), AG11395 (WRN), GM00739 (CSB) were from the Coriell Cell Repository and maintained in minimal essential (MEM) with 10% fetal bovine serum (FBS). The lines GM06990 (wt), GM01652 (wt), GM02932 (BLM), GM16375 (BLM) and GM03403 (BLM), also from the Coriell Cell Repository, were maintained in either MEM or RPMI media with 15% FBS. BJ-5ta and HeLa S3 cells were from ATCC (CRL-4001 and CCL-2.2, respectively), grown in MEM with 10% FBS. CGM1 cells were from RIKEN, also grown in MEM with 10% FBS. The isogenic cell lines BLM-comp (GM08505 + BLM cDNA, also known as PSNF5) and BLM-vec (GM08505 + vector control, also known as PSNG13) were kindly provided by Hickson.
(Gaymes, North et al. 2002) and maintained in MEM with 10% FBS with the addition of 0.35 mg/ml G418. NALM-6 DNA ligase IV knockout cells have been previously described (Iiizumi, Nomura et al. 2006). All cell lines were grown at 37°C and in 5% CO₂ in a humidified incubator.

**DNA isolation and Southern analysis**

High molecular weight genomic DNA was isolated from cells in agarose, digested with EcoRV to liberate intact rDNA gene clusters, resolved on pulse-field gels with resolution from 10 kb to 1 Mb and probed with radiolabeled rDNA sequences according to methods given by Stults et al. (Stults, Killen et al. 2008).

**Western blotting**

Cells were lysed with RIPA buffer [50 mm Tris–HCl (pH 7.4), 150 mm NaCl 1% NP-40, 1% deoxycholate, 0.1% SDS and 1 mm EDTA] containing a cocktail of protease inhibitors (Pierce, cat. 78410) for 5 min at 4°C. Whole-cell lysates were separated on 8% SDS–polyacrylamide gels, and proteins transferred to a nitrocellulose membrane and subjected to western blotting analysis with rabbit anti-BLM antibody (Calbiochem, cat. DR1034) or rabbit anti-β-tubulin (Thermo, cat. RB-9249-P). The secondary antibody used was HRP-conjugated donkey anti-rabbit IgG (Pierce, cat. 31458). Blots were detected with ECL plus western blotting detection system (Pierce, cat. RPN2132) and visualized on a Storm 860 PhosphorImager (Molecular Dynamics) via chemifluorescence.

**Cytogenetics**

SCEs were visualized in stained metaphase spreads according to Perry and Wolff (Perry and Wolff 1974) with minor modifications.

**shRNA**

The plasmids pCPM-234 (deplete BLM) and pCPM-neg (negative control) were derived from plasmids V2HS_89234 and RHS1707 (Open Biosystems), respectively. The shRNA sequences were sub-cloned into the 3’ untranslated region of a puromycin
resistance gene and expressed in a derivative of the pCAGGS vector (Niwa, Yamamura et al. 1991) modified to be incapable of episomal replication. For semi-stable knockdowns, cells were transfected with 10 µg shRNA plasmids pCPM-234 or pCPM-neg using a BTX ECM 830 square-wave electroporator: nine pulses, 150 V, 7 ms pulse duration, 1 s pulse interval, total volume 750 µl growth media. Stable integration of the plasmid was selected via puromycin resistance (300 ng/ml) for 1 week, at which time BLM knockdown was measured by western blotting followed by limited dilution sub-cloning and expansion.
4.6: FIGURES

A

Figure 4.1: Screening cell lines for gene cluster instability (GCI)

(A) Experimental strategy. Digestion of genomic DNA with restriction enzymes that do not cut within an individual gene cluster repeat liberates intact gene clusters from bulk genomic DNA. Panel from Stults et al. (Stults, Killen et al. 2008; Bernstein, Gangloff et al. 2010) (used with permission). (B) Schematic of GCI patterns in clonal cell lines. A single cell shows a well-defined pattern of gene cluster lengths (‘Initial Pattern’—thick bands). Allowable, but currently unrepresented gene cluster lengths are shown as dotted lines. As mitotic division expands the clonal cell population in the absence of instability, gene cluster lengths are faithfully preserved (‘No GCI’). Alternatively, instability generates sub-populations within the expanding population with altered cluster lengths giving rise to lower intensity ‘minor bands’ (‘Low GCI’—thin bands indicated by arrows). High levels of instability generate a ladder-like pattern of minor banding with individual bands on the ladder differing by integer multiples of the unit repeat length (‘High GCI’). (C) Screening non-clonal cell populations. Cluster lengths are resolved up to 1 Mb in length, with sizes shown from yeast chromosome markers. Unresolved clusters larger than 1 Mb run together at the gel resolution limit (star). Cell lines indicated at the top. Minor bands indicated by arrows. The ladder-like banding pattern indicative of high GCI is seen in the gene clusters from the BLM deficient cells (rounded box). BJ-5ta: wild-type newborn foreskin fibroblasts immortalized by ectopic telomerase expression; CGM1,
GM06990: EBV-transformed wild-type lymphoblast lines; HeLa S3: cervical carcinoma cells; SV40-transformed fibroblast lines: WT: wild-type (GM00637); ATM: ataxia-telangiectasia mutated (GM09607); BLM: Bloom syndrome (GM08585); Fanc-C: Fanconi anemia complementation group C mutated (GM13136); Lig I: DNA ligase 1 defective (GM16097); Lig IV: DNA ligase IV defective (GM16089); NBS: Nijmegen breakage syndrome (GM15989).

Figure 4. 2: GCI clonal analysis. (A) Wild-type SV40-transformed GM00637 fibroblast cells. The gel resolves gene clusters up to 1 Mb in length (star). A clonal population derived from a single cell is shown in the leftmost lane. Sub-clones were derived by dilution plating to single cells with subsequent expansion. Minor bands indicated by arrows. Changes to the major banding pattern shown by triangles (new bands) and brackets (missing bands). (B) SV40-transformed fibroblast GM08505 Bloom syndrome cells. The bulk population as received from Coriell is shown in the far left. Clonal populations derived from the expansion of plated single cells from this bulk population are shown in the right. Minor bands indicated by arrows.
Figure 4.3: Bloom syndrome GCI. (A) rDNA gene clusters in non-clonal cell line populations resolved to 1 Mb (star). WT: wild-type lines: SV40-transformed (GM00637), EBV-transformed (GM06990). BLM Ash.1: Ashkenazi Jewish, Bloom syndrome registry #42 (GM08505); BLM F.C.: French-Canadian, Bloom syndrome registry #81 (GM16375); BLM Ash.2: Ashkenazi Jewish, Bloom syndrome registry #9 (GM03403); BLM-comp: same as BLM Ash.1 stably complemented with BLM cDNA; BLM-vec: same as BLM Ash.1 stably transfected with empty vector control. Rounded boxes show laddering pattern indicative of a GCI-high phenotype. (B) rDNA gene clusters in primary cells resolved to 1 Mb (star). WT primary: untransformed wild-type fibroblasts (GM01652); BLM primary: untransformed Bloom syndrome fibroblasts, Ashkenazi Jewish, Bloom syndrome registry #3 (GM02932); Lambda: lambda phage concatemers (unit genome size=48 502 bp). Arrows show the minor banding pattern in BLM primary cells in the contrast-enhanced rightmost panel.
Figure 4.4: GCI in isogenic BLM and BLM-complemented human cells. (A) BLM-comp bulk population (leftmost lane) and single-cell derived clones resolved to 1 Mb. On the right is a densitometric trace of the first four clonal BLM-comp lines with minor bands indicated by arrows. (B) BLM-vec bulk population (leftmost lane) and single-cell derived clones resolved to 1 Mb (star). On the right is a densitometric trace of the first four clonal BLM-vec lines with minor bands indicated by arrows. Notice the much larger quantity of minor bands observed in the vector-control BLM-defective clonal lines relative to the BLM complemented but otherwise isogenic clonal lines in (A).
Figure 4.5: shRNA expressing clonal lines. Gene clusters resolved to 1 Mb (star). Changes to the major banding pattern are shown by open triangles (new bands) and brackets (missing bands). Arrows indicate new minor bands. (A) Sub-clones of a clonal wild-type GM00637 cell population, semi-stably transfected with a negative control shRNA. (B) Sub-clones of the same GM00637 clonal cell population, semi-stably transfected with shRNA to knockdown BLM expression. (C) Western blot showing shRNA knockdown of BLM expression 1 week post-transfection. WT: clonal isolate of wild-type GM00637 cells; BLM: Bloom syndrome GM08505 cells; BLM shRNA: WT cells semi-stably transfected with pCPM-234 (knockdown BLM); neg shRNA: WT cells semi-stably transfected with pCPM-neg (negative control). β-Tubulin is used as a loading control.
Figure 4.6: Werner syndrome clonal lines. rDNA gene clusters from WRN-defective SV40-transformed fibroblasts (AG11395) resolved to 1 Mb (star). A clonal population derived from a single cell is shown in the leftmost lane. Sub-clones were derived by dilution plating to single cells with subsequent expansion. No alterations in major gene cluster lengths are observed, and neither are any minor-intensity bands.
Figure 4.7: HeLa versus ATM-deficient clonal analysis. (A) HeLa cells. (B) ATM deficient cells. Gene clusters are resolved to 1 Mb (star). Changes to major banding patterns are indicated by open triangles. Arrows show new minor bands indicative of rearrangement during clonal expansion.
Figure 4.8: CSB-deficient and wild-type versus DNA-ligase IV knockout cell clonal analysis. (A) CSB-deficient cells. (B) Wild-type NALM-6 cells and NALM-6 cells gene targeted to inactivate DNA ligase IV. Gene clusters are resolved to 1 Mb (star). Changes to major banding patterns are indicated by open triangles. Brackets indicate loss of major bands. Minor bands are shown by arrows. Locations of predicted but unobserved minor bands in parental populations based on new major bands observed in clonally derived sub-populations (triangles) are shown by open circles. The fraction of cells in the parental population carrying clusters of these lengths is below the detection capability of the assay.
Figure 4.9: SCE analysis of BLM cells complemented with BLM cDNA BLM-comp cells are not completely complemented for the elevated SCE phenotype of Bloom syndrome cells
4.6: REFERENCES


CHAPTER 5: *Escherichia coli* RECG IS A FUNCTIONAL ANALOG OF BLM, THE HUMAN BLOOM SYNDROME PROTEIN

*This chapter has been published: Killen MW, Stults DM, Wilson WA, Pierce AJ. (Submitted)*

5.1: SYNOPSIS

Defects in human *BLM* cause Bloom syndrome, notable for early development of tumors in a broad variety of tissues. On the basis of sequence similarity, *BLM* has been identified as one of the five human homologs of *RecQ* from *Escherichia coli*. Nevertheless, biochemical characterization of the BLM protein indicates far greater functional similarity to the *E. coli* RecG protein. To explore the possibility that BLM and RecG are functional analogs, we characterized cellular phenotypes of human BLM-deficient cells stably expressing *E. coli* RecG. We find that RecG largely complements the definitive elevated sister chromatid exchange phenotype of *BLM* cells and the more recently demonstrated gene cluster instability phenotype of *BLM* cells, both in a dose-dependent manner. In contrast, expression of RecG has no impact on these phenotypes in human cells with functional BLM. Although human BLM protein has homology to *E. coli* RecQ, the biochemistry of BLM is a much closer match to that of *E. coli* RecG. We demonstrate here the physiological relevance of these related biochemistries by showing that ectopic expression of RecG in human cells complements many of the functional cellular phenotypes associated with BLM deficiency. Our complementation of human cellular disease phenotypes with a non-homologous bacterial protein provides new insights into the physiological function of the BLM protein and into the broader evolutionary biology of genomic stabilization. RecG and BLM appear to be functional analogs, and are a potential example of convergent evolution.
5.2: INTRODUCTION

Human cells possess five proteins with clear sequence homology to the *E. coli* RecQ protein: BLM, WRN, RECQL, RECQL4 and RECQL5. These proteins are all implicated in preserving genomic integrity (reviewed in (Bernstein, Gangloff et al. 2010; Monnat 2010)). Functionally, inherited homozygous defects in *BLM*, *WRN* or *RECQL4* cause human disease: Bloom syndrome, Werner syndrome and Rothmund-Thomson/RAPADILINO/Baller-Gerold syndromes respectively. Bloom syndrome is particularly striking for its predisposition to early-onset malignancy with a broad distribution of cancer types similar to that seen with sporadic tumors in the general population (German 1997).

Sequence homology of BLM with RecQ notwithstanding, characterization of the *in vitro* activities of BLM demonstrates significant similarities to the biochemistry of the RecG protein. Both BLM (Machwe, Xiao et al. 2006; Ralf, Hickson et al. 2006) and RecG (McGlynn, Lloyd et al. 2001; Robu, Inman et al. 2004) bind to and regress multi-stranded DNA structures that model stalled replication forks. Similarly, both BLM (Karow, Constantinou et al. 2000; Plank, Wu et al. 2006) and RecG (Whitby, Vincent et al. 1994; Grove, Harris et al. 2008) bind to and branch migrate Holliday junctions. Both BLM (van Brabant, Ye et al. 2000; Bachrati, Borts et al. 2006) and RecG (McGlynn, Al-Deib et al. 1997) dismantle D-loops where a 3'-OH ssDNA has invaded a homologous DNA duplex, although the mechanism by which RecG carries out this reaction is less well established (Briggs, Mahdi et al. 2004). The manner by which RecG accomplishes these tasks is in large part made clear by its crystal structure (Singleton, Scaife et al. 2001): a RecG monomer binds at a model replication fork by inserting a C-terminal protein wedge domain into the fork. The body of RecG then functions as a double stranded DNA translocase to pull the DNA template strands together through the body of the protein. At such time as the nascent DNA strands encounter the wedge domain, they are stripped off and allowed to anneal together resulting in the formation of a Holliday junction. As RecG continues to translocate on the dsDNA, the branch point of the Holliday junction is effectively migrated. In the absence of high-resolution structural information it remains unclear precisely how BLM carries out these activities.
The BLM protein also possesses activities it is not known to share with RecG. BLM can act in concert with EXO1 at double stranded DNA ends to cause a 5'-3' single stranded resection that exposes a free ssDNA 3' end suitable for loading with Rad51 (Nimonkar, Ozsoy et al. 2008), reminiscent of the combined activities of the *E. coli* RecQ helicase and RecJ 5'-3' exonuclease (Courcelle, Crowley et al. 1999; Handa, Morimatsu et al. 2009). BLM has strong unwinding activity on G-quadruplex DNA structures (Sun, Karow et al. 1998) as well as both ssDNA annealing (Cheok, Wu et al. 2005) (Machwe, Xiao et al. 2005) and/or strand exchange activities (Chen and Brill 2010). Notably, BLM has many well-characterized protein-protein interactions, including those with RMI1, C16orf75 (RMI2) and TOPOIIIα (Raynard, Bussen et al. 2006; Wu and Hickson 2006; Singh, Ali et al. 2008; Xu, Guo et al. 2008) that collectively mediate double-Holliday junction dissolution, as well as direct interaction with the Rad51 recombinase (Wu, Davies et al. 2001) and with the multi-component Fanconi anemia protein containing BRAFT complex (Meetei, Sechi et al. 2003). In contrast, RecG functions in a largely monomeric manner (McGlynn, Mahdi et al. 2000).

The mechanistic similarities between BLM and RecG have led us and others (Amor-Gueret 2006) to speculate that *E. coli* RecG and human BLM may be functional analogs. In order to test this hypothesis and to determine the extent to which the shared biochemical activities of BLM and RecG are responsible for suppressing the functional cellular phenotypes observed in human cells lacking BLM, we reasoned that suitable expression of RecG might complement a BLM defect. The best characterized cellular phenotype of BLM deficiency is a 10-fold elevated frequency of sister chromatid exchanges (Chaganti, Schonberg et al. 1974), thought to represent a hyper-recombination phenotype indicative of elevated crossing-over and overall genomic instability. In addition, we have recently demonstrated that BLM deficiency causes a striking destabilization of the highly repetitive human ribosomal RNA gene clusters (the ‘rDNA’), with recombination-mediated genomic restructuring of these clusters increased 100-fold over cells wild-type for BLM function (Killen, Stults et al. 2009). Accordingly, we engineered several semi-humanized RecG protein expression systems and stably introduced these constructs into human cells either wild type or defective for the BLM
protein. We then assayed the resulting RecG transgene expressing cells for changes in these two phenotypes.

5.3: RESULTS

i) RecG expression complements the BLM elevated SCE phenotype
We found a significant RecG dose-dependent decrease in SCEs per chromosome in BLM cells (Fig. 5.2A,C and Table 5.1) with high level expression of RecG reducing the elevation of SCEs by 75%. In our hands, expression of the human BLM protein from a cDNA construct in BLM cells reduces this characteristic elevated SCE phenotype by 80% (Table 5.1 and (Killen, Stults et al. 2009)), so E. coli RecG expression is nearly as effective as the human protein. RecG complemented the elevated BLM SCE phenotype either when expressed as an EGFP fusion protein or as a separate polypeptide co-translated with EGFP. Expression of EGFP alone without RecG in BLM cells has no statistically significant effect on SCE levels (Fig. 5.5, Table 5.1). The RecG complementation of the BLM SCE phenotype was not restricted to the Ashkenazi Jewish founder BLM mutation (6-bp del/7-bp ins) because RecG also reduced the elevated SCE phenotype of a French-Canadian derived (S595X) BLM line (Fig. 5.3A, Table 5.1). RecG is expressed at a low level in “BLM FC: RecG fuse” cells (Fig. 5.3B), resulting in a less dramatic, 25% reduction in the elevated SCE phenotype, nevertheless in a statistically highly significant manner (P=0.0004, Mann-Whitney two-tailed test).

ii) RecG has no effect on SCE levels in cells wild-type for the BLM protein
In order to establish the BLM specificity of the RecG SCE reduction effect, we carefully examined SCE levels in two different wild-type lines that stably express high levels of either the RecG-EGFP fusion protein, or the bicistronic RecG/EGFP construct (Fig. 5.4). We found that high level RecG expression (Fig. 5.2B) had no effect on either the median levels of SCEs per chromosome or in the distribution of SCEs seen in either wild-type line (Table 5.1).

iii) Loss of RecG expression restores the elevated SCE phenotype to BLM cells
One of the clonal BLM lines stably expressing medium levels of the RecG-EGFP fusion protein (Fig. 5.5B, “BLM: RecG fuse med”) gradually lost EGFP expression during two
months of continuous cell culture in the absence of selection, presumably due to epigenetic silencing of the RecG-EGFP transgene. In the resulting “BLM: lost RecG fuse” line, SCE levels were re-elevated to nearly those of either the parental BLM line, or the control BLM line containing a transgene for EGFP alone “BLM: EGFP” (Fig. 5.5A, Table 5.1), confirming that RecG expression is solely responsible for complementing the elevated SCE phenotype in BLM cells.

iv) RecG expression reduces the elevated gene cluster instability (GCI) of BLM cells. Loss of BLM very strongly increases the rate of spontaneous recombination-mediated genomic restructuring in the highly repetitive human rRNA gene clusters (Killen, Stults et al. 2009). In BLM cells, recombination alters gene cluster lengths on the order of every round of mitotic cell division, an approximate 100-fold increase over the rate in wild-type cells. This rapid randomization of cluster lengths produces a diagnostic ladder-like electrophoretic karyotype with each rung on the ladder separated by the 48.5 kb unit length of the human rDNA repeat, particularly in the range of gene cluster lengths from 250 kb to 550 kb (Fig. 5.6 – shaded in red, red side bracket). The intensity of individual cluster bands is related to how early in the clonal expansion the mitotic recombination event took place, with more intense bands having arisen earlier in the expansion of a clonal population (Killen, Stults et al. 2009). Expression of RecG complements this elevated GCI phenotype: the rate of spontaneous cluster restructuring is greatly reduced (Fig. 5.6 – red carets) although as with complementation of the elevated SCE phenotype of BLM cells, there is a residual low level of GCI even with high levels of RecG expression. Notably, expression of RecG is particularly effective at restabilizing gene clusters in the previously highly unstable 250 kb to 550 kb size range, as seen by the significant reduction of gene cluster lengths detected in this range.

v) RecG expression does not affect gene cluster instability in wild-type cells. Expression of RecG as an independent polypeptide co-translated with EGFP did not have an obvious effect on gene cluster instability in a cell line wild-type for BLM protein (Fig. 5.7) even though RecG is expressed at high level when measured by western detection of the co-translated EGFP protein in this line (Fig. 5.2B: “WT: RecG 2a”). The changes to the banding pattern in the subclones relative to that of the parental clonal population represent mitotic recombination events that occurred sufficiently late in the expansion of
the parental clonal population as to be undetectable as visible minor-intensity banding. The relative small number of changes to major intensity bands in the subclones and the lack of detectable minor bands in any of the cell populations is consistent with the low levels of spontaneous gene cluster restructuring seen in other wild-type transformed human cell lines (Killen, Stults et al. 2009).

5.4: DISCUSSION

We show here that expression of the *E. coli* RecG protein is able to complement two functional cellular phenotypes of BLM deficiency: elevated sister chromatid exchange, and elevated gene cluster instability. Although this result is in accord with the several shared biochemical activities of these two proteins, this functional complementation across different kingdoms of biology is surprising, particularly when considering that there appears to be no relationship of evolutionary sequence homology between RecG and BLM. Both RecG and BLM are members of the DEXDc superfamily of helicase proteins, and both contain both DEXDc and HELICc subdomains (Marchler-Bauer, Lu et al. 2011), however, the *E. coli* RecQ protein is the clear evolutionary homolog of BLM on the basis of sequence similarity within these two conserved domains. Both RecG and BLM have extended amino acid sequences N-terminal to their DEXDc superfamily domains. In BLM, this N-terminal half of the protein is not well-characterized; it seems likely that significant structure/function similarities to RecG may reside in this region.

The *E. coli* RecG protein, with essentially no sequence similarity to human BLM, would seem unlikely to engage in any of the well-characterized protein-protein interactions that are important for function of the BLM protein. The capacity with which RecG can complement BLM cellular phenotypes suggests therefore that the primary physiological role of BLM is to perform the same molecular reactions that can be performed in a human cell by the RecG monomer alone, namely direct manipulations of DNA structures. Secondary effects such as signaling through protein-protein interactions are likely to be of lesser physiological importance, at least at the cellular level. Similarly, it would appear that precise regulation of the reactions carried out by the BLM protein are likewise largely unnecessary for function of the protein since RecG was constitutively
expressed in our experiments. Conceivably, targeting of BLM by protein-protein interactions to DNA structures on which RecG can act may reduce the amount of BLM needed by the cell. Activities of BLM not known to be shared by RecG, such as the unwinding of G-quadruplex DNA, must play a minor role in both sister chromatid exchange and gene cluster instability suppression, or alternatively, a G-quadruplex activity of RecG may remain to be discovered.

The role of BLM in human genomic stabilization is well-established. In contrast, a major mechanistic role of RecG in *E. coli* genomic stabilization has only recently been elucidated. In the absence of RecG, PriA initiates spurious replication forks throughout the bacterial genome, causing poorly-controlled genomic over-replication and compromised viability (Rudolph, Upton et al. 2009). In the absence of RecG, cells have an absolute requirement for a ssDNA exonuclease activity, unless PriA is also eliminated (Rudolph, Mahdi et al. 2010). It will be interesting to determine whether human *BLM* cells share this over-replicated phenotype. Curiously, the *E. coli* RecQ protein rather than protecting the cell from aberrant recombination structures as the human BLM and *E. coli* RecG proteins do, seems to act instead to promote their formation (Magner, Blankschien et al. 2007). Cells lacking the UvrD-mediated inhibition of recombination and also lacking RecG are killed by formation of intermolecular recombination intermediates (IRIs) that interfere with correct genome segregation to daughter cells. Formation of IRIs is caused by the action of RecQ and partner proteins and causes “death by recombination”. Deletion of RecQ restores cellular viability (Fonville, Blankschien et al. 2010). An analogous phenotype in human cells might be the elevated formation of anaphase bridges in cells deficient for BLM (Chan, North et al. 2007). One evolutionary interpretation consistent with our functional data here would be that although the ancient common protein ancestor of both RecQ and BLM was preserved in evolutionary decent through both lineages to provide a core helicase domain, the functionality of these proteins diverged in opposite directions. Since there are no RecG homologs by sequence similarity in human cells, it would appear that at least one of the five human RecQ paralogs, BLM, was co-opted to perform the important recombination molecular transactions provided to *E. coli* by the RecG protein, and is an example of convergent evolution.
5.5: METHODS

i) Protein expression constructs

The coding sequence for RecG was isolated from *E. coli* TOP10 (Invitrogen) genomic DNA with the addition of a consensus Kozak sequence (Kozak 1994) and the SV40 large T-antigen nuclear localization signal (PKKKRKV) by PCR using primers 5'-gggggggaatccagccaccatggctccaaaaaagcgcaaagtggcgatgaaaggtcgcctg and 5'-gggggggatatcgcggccgc (RecG sequences underlined) followed by cloning into the pCAGGS mammalian constitutive expression vector. A cryptic polyadenylation sequence in the RecG coding sequence 5'-AATAAA (AsnLys) was removed via silent mutagenesis to 5'-AACAAG (AsnLys). Carboxy-terminal additions of the enhanced green fluorescent protein (EGFP) coding sequence were constructed either to produce a RecG-EGFP fusion protein with a GSG linker peptide (Fig. 5.1A) “RecG fuse” (predicted molecular weight: 104.7 kDa), or to produce two proteins in a bi-cistronic manner: RecG fused to the Strep-tag II for potential affinity purification (Maier, Drapal et al. 1998) followed by a GSG linker, the 2A polyprotein ‘self-cleaving’ sequence from *Thosea asigna* virus (Donnelly, Hughes et al. 2001; Szymczak, Workman et al. 2004), a VAT peptide linker and EGFP (Fig. 5.1B) “RecG 2a”. In the bi-cistronic construct the predicted molecular weight of the RecG polypeptide is 80.8 kDa and the predicted molecular weight of the EGFP polypeptide is 27.3 kDa. All final construct sequences were verified by direct DNA sequencing. The EGFP expression vector used for subcloning the EGFP coding sequence was pEGFP-N1 (Clontech).

ii) Cell lines

Cell lines used that are wild-type for BLM include the SV40-transformed fibroblast line GM00637 (Coriell) “WT” and the cervical carcinoma line HeLa S3 (ATCC) “HeLa.” Cell lines used that are deficient in BLM protein include GM08505 (Coriell) which are SV40 transformed fibroblasts derived from a Bloom syndrome patient homozygous for the Ashkenazi Jewish founder BLM mutation (6-bp del/7-bp ins) at nucleotide 2281 of the open reading frame, hereafter referred to as “BLM.” A second BLM deficient line used was GM16375 (Coriell) which are EBV transformed lymphocytes from a French-
Canadian Bloom syndrome patient homozygous for a C>A transversion resulting in a (S595X) termination mutation, hereafter referred to as “BLM FC”. The BLM-defective line stably complemented by either BLM cDNA expression, or by a control empty vector are the lines PSNF5 “BLM: cDNA” and PSNG13 “BLM: vec” respectively (Gaymes, North et al. 2002) (kind gift from Ian Hickson) both derived from the GM08505 line. Lines were generally grown in MEM with 10% fetal bovine serum, with L-glutamine and antibiotic supplementation at 37°C in a humidified 5% CO2 incubator. Stable cell lines were generated by electroporating either a RecG expression construct or a control EGFP expression construct into cells, followed by unselected cell population expansion, one round of flow-sorting enrichment for green fluorescent cells, further unselected expansion, and finally a second round of flow-sorting enrichment for green fluorescent cells. Clonal and subclonal derivatives of these highly enriched fluorescent populations were subsequently derived by limiting dilution. All of the transgene expressing cell lines generated and used in this work are either clonal or subclonal, with the exception of the “BLM FC: RecG fuse” line only: BLM FC cells were transduced by a high-titer lentivirus (Welgen, Inc.) containing an expression cassette for the RecG-EGFP fusion construct (Fig. 5.1A) and separated into green fluorescent and non-fluorescent populations by flow sorting.

iii) Sister chromatid exchange assays

Sister chromatid exchange assays were performed largely according to (Perry and Wolff 1974) with minor modifications (Killen, Stults et al. 2009). Cellular metaphase spreads were imaged and scored individually by counting the number of visible exchanges and the number of chromosomes in each metaphase. The resulting SCEs/chromosome figures were binned and plotted. All statistical tests were performed using unbinned data.

iv) Western blotting

Protein extracts were prepared using RIPA buffer as described previously (Killen, Stults et al. 2009). All resolving SDS-PAGE gels used 9% acrylamide and were blotted onto Hybond-ECL nitrocellulose membrane (Amersham Biosciences, cat. #RPN68D). Primary antibodies used were: rabbit anti-GFP (Cell Signaling Technologies, cat. #2555), rabbit anti-β-actin (Cell Signaling Technologies, cat. #4970), rabbit anti-β-
tubulin (NeoMarkers, cat. #RB-9249-PO). The secondary antibody was ImmunoPure Antibody donkey anti-rabbit IgG conjugated with horseradish peroxidase (Pierce, cat. #31458). Blots were developed using an Amersham™ ECL Plus western blotting detection system (GE Healthcare, cat. #RPN2132) and imaged with a Storm 860 PhosphorImager (Molecular Dynamics).

v) GCI analysis

Gene cluster instability analysis was carried out as described previously (Killen, Stults et al. 2009). Briefly, genomic DNA was prepared in the solid phase by digesting single cell suspensions in agarose with proteinase K in the presence of sarkosyl and EDTA, rinsed thoroughly and equilibrated in 50% glycerol/10 mM Tris/1 mM EDTA pH 8.0 and stored at -20°C. 10 ml agarose slices containing approximately 1 mg genomic DNA were equilibrated in suitable restriction digestion buffer and digested overnight with EcoRV (New England Biolabs). Digested DNA still in solid form was loaded into a 1% PFC agarose (Bio-Rad) gel and run in 0.5x TBE buffer (44.5 mM Tris base, 44.5 mM boric acid, 1 mM EDTA pH 8.0) using a CHEF-MAPPER system (Bio-Rad) at 14°C. Pulsed-field electrophoretic conditions were a field strength of 6 V/cm with 120° separation between field vectors. Field switch times varied from 3 seconds to 90 seconds with a ‘ramp factor’ of 0.357. Gels were run for 24 hours, then dried, rehydrated, probed with a radiolabeled probe specific for the human rDNA (Stults, Killen et al. 2008) and imaged with a Storm 860 PhosphorImager (Molecular Dynamics).
5.6: FIGURES

Figure 5.1: RecG expression constructs. A) RecG expressed as a fusion protein with EGFP. Lettering colors: black: linker amino acids, red: nuclear localization signal, blue: RecG coding sequence, green: EGFP coding sequence. B) RecG co-translated with EGFP as two separate polypeptides. Lettering colors: black: linker amino acids, red: nuclear localization signal, blue: RecG coding sequence, brown: strep-tag II, purple: 2a ‘self-cleaving’ peptide, green: EGFP coding sequence. Dotted line represents the division of the precursor polyprotein into two independent polypeptides.
Figure 5.2: RecG expression complements the BLM elevated SCE phenotype. A) Normalized sister chromatid exchange frequencies. “RecG 2a” denotes cells express RecG and EGFP as separate co-translated proteins. “RecG fuse” denotes cells express RecG and EGFP together as a fusion protein. B) RecG expression levels. Protein extracts from the indicated cell lines are blotted with an anti-GFP antibody and with anti-β-actin as a loading control. RecG-EGFP fusion proteins (105 kDa) are indicated with blue ovals. EGFP is detected as a proxy biomarker for RecG in the “RecG 2a” co-translated lines. C) Representative sister chromatid exchange metaphase spreads with sister chromatids differentially stained. The harlequin staining pattern of the BLM cells is greatly reduced upon expression of RecG.
Table 5.1: The effect of RecG expression on SCEs per Chromosome

<table>
<thead>
<tr>
<th>Cells</th>
<th>Median SCE/Chr.</th>
<th>$P_{(2)}$ vs BLM: EGFP $^a$</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>0.10 $^b$</td>
<td>&lt;0.0001</td>
<td>2,4</td>
</tr>
<tr>
<td>WT: RecG 2a</td>
<td>0.10 $^b$</td>
<td>&lt;0.0001</td>
<td>4</td>
</tr>
<tr>
<td>HeLa</td>
<td>0.09 $^b$</td>
<td>&lt;0.0001</td>
<td>4</td>
</tr>
<tr>
<td>HeLa: RecG fuse</td>
<td>0.08 $^b$</td>
<td>&lt;0.0001</td>
<td>4</td>
</tr>
<tr>
<td>BLM</td>
<td>0.98 $^d$</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>BLM: empty vector</td>
<td>1.05 $^d$</td>
<td>0.9283</td>
<td>-</td>
</tr>
<tr>
<td>BLM: cDNA</td>
<td>0.28 $^c$</td>
<td>&lt;0.0001</td>
<td>2</td>
</tr>
<tr>
<td>BLM: RecG 2a high</td>
<td>0.33 $^c$</td>
<td>&lt;0.0001</td>
<td>2</td>
</tr>
<tr>
<td>BLM: RecG fuse high</td>
<td>0.34</td>
<td>&lt;0.0001</td>
<td>2</td>
</tr>
<tr>
<td>BLM: RecG 2a low</td>
<td>0.68</td>
<td>&lt;0.0001</td>
<td>2</td>
</tr>
<tr>
<td>BLM: RecG fuse low</td>
<td>0.85</td>
<td>0.0001</td>
<td>2</td>
</tr>
<tr>
<td>BLM: EGFP</td>
<td>1.02 $^d$</td>
<td>1</td>
<td>2,5</td>
</tr>
<tr>
<td>BLM FC: RecG fuse</td>
<td>0.70 $^e$</td>
<td>&lt;0.0001</td>
<td>3</td>
</tr>
<tr>
<td>BLM FC</td>
<td>0.91 $^e$</td>
<td>0.0011</td>
<td>3</td>
</tr>
<tr>
<td>BLM: RecG fuse med</td>
<td>0.43 $^f$</td>
<td>&lt;0.0001</td>
<td>5</td>
</tr>
<tr>
<td>BLM: lost RecG fuse</td>
<td>0.86 $^f$</td>
<td>0.0067</td>
<td>5</td>
</tr>
</tbody>
</table>

$^a$ Mann-Whitney two-tailed test.

$^b$ Differences between these distributions cannot be established. $P=0.5539$ Kruskal-Wallis test for $k=4$.

$^c$ $P_{(2)}=0.0332$ for “BLM: cDNA” vs “BLM: RecG 2a high” by Mann-Whitney two-tailed test.

$^d$ Differences between these distributions cannot be established. $P=0.99$ Kruskal-Wallis test for $k=3$.

$^e$ $P_{(2)}=0.0004$ Mann-Whitney two-tailed test.

$^f$ $P_{(2)}<0.0001$ Mann-Whitney two-tailed test.
Figure 5.3: RecG expression reduces SCE in a French-Canadian *BLM* line.  A) Normalized sister chromatid exchange frequencies. “BLM FC”: EBV-immortalized lymphocytes homozygous for an inactivating (S595X) mutation. “BLM FC: RecG fuse”: the same cells expressing low levels of the RecG-EGFP fusion protein. B) Western blot showing relative protein levels. Blue oval: RecG-EGFP fusion protein. “BLM FC: RecG fuse”: protein extract from French-Canadian *BLM* cells stably expressing low levels of the RecG-EGFP fusion protein. “BLM FC”: protein from French-Canadian *BLM* cells lacking a RecG transgene. “BLM: RecG fuse low”: protein extract from Ashkenazi Jewish *BLM* cells complemented with RecG-EGFP (see also Fig. 2B for comparison of protein levels). Notice the different intensities of the β-actin loading control between the “BLM FC” extracts and the “BLM” extract.
Figure 5.4: RecG expression has no effect on SCE levels in wild-type cells. Normalized sister chromatid exchange frequencies are shown.
Figure 5.5: Loss of RecG expression in BLM cells restores the elevated SCE phenotype. A) Normalized sister chromatid exchange frequencies are shown. B) Loss of expression of RecG-fuse is shown by western blotting (blue ovals).
Figure 5.6: RecG stabilizes the gene cluster instability of BLM cells. Southern blotting detection of the rDNA clusters with resolution up to 1 Mb is shown. Cluster lengths are calibrated by *S. cerevisiae* chromosome size markers. Open star is the 1 Mb resolution limit of the gel: all clusters larger than 1 Mb co-localize to this band. Clonal populations are shown on the left of each panel, with a collection of subclones derived from this clonal population to the right. Red carets: mitotic recombination indicated by minor intensity gene cluster bands. The zone of particularly high instability in the 250 kb to 550 kb size range is indicated by red shading and by the red dotted bracket at the right. A) “BLM: RecG 2a high”: BLM cells stably expressing high levels of cotranslated RecG/EGFP. B) The parental *BLM* line from which the “BLM: RecG 2a high” cells were derived.
Figure 5.7: RecG has no effect on gene cluster instability in wild-type cells. Southern blotting detection of the rDNA clusters in wild-type cells stably expressing high levels of co-translated RecG/EGFP, “WT: RecG 2a”. Cluster lengths resolved up to 1 Mb are shown. Cluster lengths are calibrated by *S. cerevisiae* chromosome size markers. Open star is the 1 Mb resolution limit of the gel: all clusters larger than 1 Mb co-localize in this region. See Fig. 2B for RecG expression levels by western blotting. A clonal cell population is shown in the far left lane of the gel. A collection of subclonal populations each derived from a single cell of this clonal population are shown to the right. Plum circles: relocated major band clusters. Blue braces: former location of major band clusters from the parental clone.
5.7: REFERENCES


CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation genomic instability driven by dysregulated homologous recombination is explored as a potential mechanism of carcinogenesis. A new assay is presented that employs gene cluster instability (GCI) as a sentinel biomarker of dysregulated homologous recombination. This assay was used to demonstrate the incidence of this type of genomic instability in tissue from human cancer patients and to establish gene cluster instability as a primary phenotype of cells deficient in BLM protein. It was shown that the lack of BLM protein is solely responsible for the phenotype in the BLM deficient cell lines. Since Bloom’s syndrome cells have been characterized as hyper-recombinatory and BS patients present with an extreme predisposition to a broad spectrum of cancers, BLM cells were explored as a model of generalized cancer to gain insight into the role that dysregulated homologous recombination might play in carcinogenesis. Since the GCI phenotype is common to both BS and human cancer some of the unique biochemical characteristics of BLM that are thought to play a role in suppressing genomic instability were explored. Using the gene cluster instability assay along with sister chromatid exchange analysis it was determined that these phenotypes of BLM deficient cells could be functionally complemented by the non-homologous bacterial protein RecG. The fact that these proteins share specific biochemical activities but no structural or sequence homology has led to the speculation that it is these shared biochemical activities are important to suppressing this type of instability. Therefore it is plausible that these same biochemical pathways might be responsible for the instability observed in the gene clusters in BS patients and human cancer tissue.

The development of a new assay that uses gene cluster instability (GCI) as a sentinel biomarker for dysregulated homologous recombination is described in Chapter 2. In short, this assay employs pulse-field gel electrophoresis (PFGE) to evaluate size changes of intact gene cluster liberated through enzymatic digestion of mega-base scale DNA preps. The gene clusters are detected via an in-gel southern blotting technique. This assay allows the user to evaluate clonal isolates of a cell line or to compare cancerous tissue to normal blood and tissue. Several gene clusters were explored in order to
determine the most useful prognosticator. This assay was previously verified on blood from normal human volunteers while characterizing the human ribosomal gene cluster (rDNA) (Stults, Killen et al. 2008). The results provided direct evidence of restructuring in the gene cluster architecture and validated that GCI was an extremely sensitive method to detect dysregulated homologous recombination. In another publication this assay was used to investigate instability in the rDNA gene clusters in human cancer samples, where it was found that 50% of the human cancer samples tested had a GCI phenotype (Stults, Killen et al. 2009). Of the gene clusters explored the rDNA gene cluster seemed to demonstrate the most sensitivity; however, due to its unique characteristics it was necessary to investigate these results in what would be considered a more normal gene cluster.

In chapter 3 the GAGE gene cluster was evaluated in normal human families and in human cancer. This series of experiments were performed to complement the work mentioned above that demonstrated mitotic instability in the rDNA gene clusters in human cancer (Stults, Killen et al. 2008; Stults, Killen et al. 2009). Although the rDNA proved to be the most sensitive substrate the unique qualities of the rDNA gene clusters required that other clustered sequences be considered. As mentioned in chapter 1 the rDNA clusters are composed of more than 600 copies of a 43kb unit repeat that is dispersed across 5 different chromosomes and they are greater than 99.5% similar between copies. These qualities, high local concentration, high sequence similarity and multiple copies make the rDNA clusters ideal substrates for HR. (Stults, Killen et al. 2008) However the rDNA has some other characteristics that might contribute to its behavior that are not common to most genes. The rDNA is largely nucleolar and is transcribed by RNA Polymerase I, unlike the majority of genes, which are normally transcribed but RNA Polymerase II (Russell and Zomerdijk 2006). The GAGE gene cluster is considered here as a more typical example of repetitive sequence found in the genome that would provide a substrate for dysregulated homologous recombination. The genomic structure of the GAGE gene cluster is similar to the rDNA in that it has a high local concentration, high sequence similarity and multiple copies. However, the unit repeat of the gene that produces the GAGE transcript is only 9.6 kb making it
considerably smaller than the rDNA unit repeat (Hofmann, Caballero et al. 2008). The multiple copies of the GAGE gene are structured as a tandemly repeated cluster of at least 15 gene copies at Xp11.23 and are oriented in a head-to-tail manner, without any intervening sequences. Thus they are also similar to the rDNA genomic structure although they have significantly fewer unit repeats which make these loci more similar to gene clusters elsewhere in the genome (Warburton, Hasson et al. 2008). Most importantly, the GAGE gene clusters have not been observed to have any of the replication or localization abnormalities found in the rDNA gene clusters.

The results in chapter 3 demonstrate low levels of meiotic rearrangement in families and detectable mitotic rearrangement in adult solid tumors which concurs with the data reported previously (Stults, Killen et al. 2008; Stults, Killen et al. 2009). GAGE cluster alterations were detected in 5% of the adult solid tumors tested. This is greatly reduced from the greater than 50% restructuring found in the rDNA gene clusters of the same patients. But observing these events in 5% of the population is remarkable compared to the rDNA since the rDNA clusters compose more than 0.5% of the human genome and the GAGE clusters account for less than 0.007% of the genome. That is a 70-fold reduction in available sequence yet there is only a 10-fold decrease in observed events. It is also an interesting observation that the cancer patients in which GAGE cluster restructuring occurred also exhibited rDNA restructuring in their disease, suggesting a potential mechanistic link (Stults, Killen et al. 2009). Additionally, restructuring events were only observed in the diseased tissue, never in the normal tissue or blood as was the observation in the rDNA (Stults, Killen et al. 2009).

Both the GAGE data here and the rDNA data presented previously provide a strong causal argument for Gene Cluster Instability and by extension dysregulated homologous recombination driving carcinogenesis. However, that argument in itself suffers the crux of whether genomic instability was the cause or the consequence of tumorgenesis. This causality dilemma is further complicated by the fact that the mechanism by which GCI occurs is unknown. Although the mechanistic requirements of HR make it the logical choice by which these restructuring events could occur this has not been verified. To
investigate this relationship a new series of experiments is necessary. Ideally, one could
induce GCI in human cells and determine if it leads to carcinogenesis. However, how to
accomplish this is not immediately clear. Fortunately, the work in Chapter 4 led to a
potential cell culture model that provides a corresponding human phenotype.

In Chapter 4 the newly developed GCI assay was employed to characterize cells
deficient for the BLM protein associated with human Bloom’s syndrome. Initially we
screened a panel of cell lines derived from patients exhibiting defects in a variety of DNA
repair pathways, including various known human chromosomal instability disorders.
Only the cell line derived from a BS patient was strongly destabilized. This is very
interesting since one of the most prevalent phenotypes associated with BS is a strong
predisposition to cancer. This predisposition is not associated with a specific type of
cancer, so this disease is often used as a model for generalized cancer by us and others.
To use this as a model it became important to establish that the absence of the BLM
protein was responsible for the GCI phenotype observed.

The results presented in chapter 4 show that a high-GCI phenotype is present in
multiple BLM cell lines, some of which have different mutations, all of which abolish
BLM protein production. In fact, high GCI was observed in every BLM cell line tested
including a primary cell line. There was a 100-fold increase in the cellular GCI
phenotype in cell lines derived from BS patients. To complement these observation two
methods were used to demonstrate that BLM was solely responsible for the phenotype
observed. First shRNA was used to knock down BLM in wild-type fibroblast. This
caused a significant increase in both GCI and SCE. Additionally, a BS cell line was
obtained whose BLM expression was reconstituted by stable expression of the full length
BLM cDNA and it was used to show that reconstitution of the protein complemented the
genomic instability phenotypes demonstrated previously. From this data it was
determined that the absence of BLM directly leads to genomic instability as demonstrated
by GCI and SCE. Taken together with the fact that there is a 100-fold increase in cancer
risk for Bloom’s Syndrome patients compared to normal individuals, this data suggests
that GCI is likely a precursor to cancer and one of the mechanisms that serves as a driving force of carcinogenesis.

The data from chapter 3 and 4 support the hypothesis that Gene Cluster Instability could be a driving mechanism of carcinogenesis; however, there are alternatives that must be considered. A deficiency in the BLM protein causes cancer, but how it causes cancer is unclear. The drastic increase in GCI and SCE is strongly suggestive that NAHR is responsible, but this evidence is still circumstantial. Although it has now been established that both an increase in GCI and SCE are phenotypes of BLM cells, it is not absolute that these two phenotypes arise from the same mechanism. It has long been hypothesized that the increase in SCE in BS is a result in dysregulated homologous recombination. It is likely that this is also the mechanism underpinning gene cluster instability. However, there is no evidence that GCI and SCE arise from the same consequences. Therefore both types of instability could contribute to carcinogenesis, but neither can be established as a root cause until it is demonstrated that they are caused by a common mechanism. This hypothesis also discounts the other know defects associated with BS. For instance Bloom’s Syndrome is known to result in a drastic increase in loss of heterozygosity (LOH). This defect could also drive carcinogenesis. Although, it is hypothesized that LOH is also a result of dysregulated homologous recombination. So it is left to determine if dysregulated homologous recombination is the underlining mechanism behind all of these phenotypes. Some insight might be gained by comparing normal and cancer tissue from Bloom’s Syndrome patients. This might indicate the prevalence of GCI, SCE and LOH in the cancer tissue. However, it is unlikely that this experiment will be undertaken in the foreseeable future. The small patient pool would make the logistics of collecting a sufficient amount of samples nearly impossible. Another potential would be to adapt the GCI assay to mice in order to exploit the mouse model of Bloom’s syndrome. However, perhaps a more straight forward method of determining the involvement of dysregulated homologous recombination in BLM cells would be to employ a reporter gene. It was recently reported that BLM knockdown in cell lines contain an I-SCEI driven reporter construct specifically increased the frequency of HR events that produced deletions by crossovers while leaving the over-all frequency
of HR events unchanged. This work provides some evidence that loss of BLM can increase genomic instability by provoking an increased frequency of HR events of a potentially deleterious nature (Wang, Smith et al. 2011). However, since the reporter system used is based on selection, it does not accurately reflect the concise population that results from repair in this mutant background. To address this we developed a new reporter gene, XRDRE that functions to repair a single DSB introduced by the homing endonuclease I-SCEI. XRDRE is designed to quantify dysregulated homologous recombination as a result of crossing over and distinguish it from gene conversion (GC). Additionally, the repair product is easily assayable by flow cytometry since it produces function red fluorescent protein. Unfortunately, there was not a sufficient amount of time to collect this data. However, this is a definite avenue of investigation for future work.

In order to explore the link between GCI and dysregulated homologous recombination in BLM cells Chapter 5 explores the physiological relevance of the established biochemistry of BLM. BLM is a RecQ helicase and is a homolog of E. coli RecQ. BLM and RecQ share some biochemical activities, however, the activities that are believed to explain BLM’s role in HR are not shared with RecQ. Those biochemical interactions overlap with the known biochemical properties of another E. coli protein, RecG. BLM and RecG are both branch migration-specific helicases and display greater helicase activity on four-way junction DNA substrates such as Holliday junctions than they do on other DNA substrates. These activities are thought to be essential for normal HR. Chapter 5 explores the experimental complementation of BLM deficient cells via ectopic expression of RecG. Since BLM and RecG share no homology in their amino acid sequences any interactions with accessory proteins can be ruled out. Thus any complementation can be attributed solely to their shared biochemical activities. Indeed ectopic expression of RecG complements both the SCE and GCI phenotypes of BLM deficient cells to the same level as complementation with the BLM protein. RecG and BLM appear to be functional analogs, and are a potential example of convergent evolution. This type of convergent evolution is not unheard of, even within the recombination repair pathway there are other examples. For instance RecO and RAD52 are functional analogs despite almost no sequence similarity (Kolodner, Fishel et al.
1985). More importantly from this it becomes apparent that the prevalent biological involvement of BLM in HR stems from these shared biochemical activities.

Taken together this data points to one of several conclusions, first it is possible that the loss of BLM leads to the initiation of more DNA lesions because of the loss of capacity to unwind quadruplex-G DNA. This might cause additional recombination initiation in the G/C-rich rDNA repeats. Likewise, if replication forks stall and collapse at quadruplex-G structures that might have been processed by BLM and this could lead to more lesions. A second possibility is that the loss of BLM decreases the cells capacity to dismantle Holliday junctions resulting in more recombination intermediates being repaired with crossing-over rather than GC. It is also possible that although the biochemistry for BLM is well established, it may not accurately depict the physiological relevant enzymatic functions of BLM. With this in mind the data supports the hypothesis that the overlapping biochemical activities of BLM and RecG are germane to BLM’s primary physiological function and could explain how BLM acts to suppress genomic instability. Still there is no direct proof that BLM and RecG are functioning through the same pathway. It will be necessary to establish if RecG functions through the same mechanism as BLM. Since the cells that are functionally complemented with RecG express a protein tagged with EGFP they lend themselves to microscopic studies. Thus exploration of RecG localization to sites of DNA damage should be explored. Another recently established association of BLM is its localization to anaphase bridges (Chan, North et al. 2007). This too could be investigated in the RecG-EGFP expressing cells. Since the data also indicates that one or all of the three biochemical activities common to both BLM and RecG is germane to observed phenotype it would also be pertinent to attempt disrupting these pathways using the tools presented here to evaluate their physiological prevalence.

It was my intentions for this work to provide insight into carcinogenesis and outline how dysregulated homologous recombination might contribute to this process. Although, the work herein encroaches upon these goals, it provides just as many questions. Fortunately it also provides some new tools to begin investigating these questions. Since
I have already integrated XRDRE into BLM deficient cells the most obvious and immediate work that should be done is to differentiate between GC and NAHR in these cells in order to determine if the loss of BLM protein increases homology directed repair with associated cross-over. Also as a result of my work I have integrated XRDRE and DRGFP into HeLa cells. These cells could easily be used to screen siRNA libraries to determine which genes modulate HR. The role of genome instability in the development of cancer continues to evoke intense debate. To address these critical issues it is clearly important to understand the mechanisms that give rise to genome instability. Through these studies it might be possible to establish the mechanism by which BLM suppresses genomic instability. Resolving this mechanism may provide insight to how instability can arise from HR and how it can lead to the gross genomic alterations often associated with cancer.
REFERENCES:


NAME: Michael W. Killen, MS
DATE OF BIRTH: September 03, 1979
PLACE OF BIRTH: Portsmouth, OH

EDUCATION

2006-2011 Ph.D Candidate, Curriculum in Microbiology, Immunology, and Molecular Genetics, University of Kentucky College of Medicine
2006-2008 Medical Sciences M.S, University of Kentucky College of Medicine
1997-2004 Biology B.S., University of Kentucky
1997-2004 Chemistry B.S., University of Kentucky
1997-2004 Mathematics B.S., University of Kentucky

PUBLICATIONS


