Autosomal random asynchronous replication is analogous to X-chromosome inactivation

By

Alexander Wilson Ensminger

B.S. Biology
Haverford College
Haverford, PA 19041

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

at the

Massachusetts Institute of Technology

February, 2006

Signature of Author.....
Department of Biology
October 28, 2005

Certified by.....
Andrew Chess
Associate Professor of Biology
Thesis Supervisor

Accepted by.....
Stephen P. Bell
Chair, Committee on Graduate Students
Department of Biology

Autosomal random asynchronous replication is analogous to X-chromosome inactivation

by

Alexander Wilson Ensminger

Submitted to the Department of Biology on October 28, 2005 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

A number of mammalian genes are expressed from only one of two alleles in either an imprinted or random manner. Those belonging to the random class include X-linked genes subject to X inactivation, as well as a number of autosomal genes, including odorant receptors, immunoglobulins, T-cell receptors, interleukins, natural killer-cell receptors, and pheromone receptors. Random asynchronous replication of DNA in S-phase represents an epigenetic mark that often parallels monoallelic expression. All randomly monoallelically expressed genes discovered to date replicate asynchronously in S-phase, though not all of the genes contained within asynchronous domains are monoallelically expressed. The focus of my work has been on understanding this random choice that cells make between two sequence-identical alleles. Using two-color fluorescent *in situ* hybridization (FISH) analyses, the random asynchronous replication of a large number of human and mouse genes appears to be coordinated at the level of entire chromosomes. This regulatory scheme is reminiscent of random X-chromosome inactivation, the dosage compensation machinery in mammals. We have shown that autosomal coordination responds to trisomy in a fashion similar to X inactivation, with one copy of the trisomic chromosome marked for early replication and the other two rendered late replicating. These observations raise the intriguing possibility that the mechanistic underpinnings of X inactivation and autosomal coordination may also be similar. Furthermore, the existence of chromosome-wide epigenetic differentiation between autosomes has evolutionary implications concerning the establishment of X inactivation as the approach to mammalian dosage compensation.

A crucial event in X inactivation is the random monoallelic expression of a noncoding RNA, *Xist* from one of the two X chromosomes. Noncoding RNA transcripts are enticing candidates for regulating chromatin structure within the mammalian nucleus. We have initiated a screen for novel nuclear, noncoding RNA transcripts. Using expression array profiling, we have identified several broadly expressed nuclear enriched transcripts. In addition to *Xist*, this approach identified two noncoding transcripts, *NEAT1* and *NEAT2* that are located near one another on human chromosome 11 and chromosome 19 of mice. Using a variety of techniques, including RNA FISH and RNA-mediated interference, we have explored the potential regulatory functions of these transcripts.

Thesis Supervisor: Andrew Chess
Title: Associate Professor of Biology

Dedicated to

Julie Claycomb

and

William and Karen Ensminger

Acknowledgements

There are a number of people who have contributed to my progress as a student over the years and have helped me to find this path. I have always felt that things work out ok if you just wait long enough. Graduate school has certainly had its share of challenges, both personal and professional. The ability to look at the opportunities afforded by temporary defeat and to keep things in perspective has been indispensable.

The feeling that everything is going to be okay if you just stick it out is probably part genetic and part environment. Surely, that level of optimism is reinforced if things really do turn out okay in the end. I want to thank my parents for not only instilling that belief, but also for being there to make sure that it comes true. I certainly have had my share of ups and downs over the past several years, but they were there in the peaks and the valleys, helping to keep things in perspective. I feel very lucky to be their son.

One of the major peaks for me during graduate school was the development of some great friendships over these years. One of these friendships, with my wife, Julie Claycomb, deserves particular highlighting. We started out as classmates in 7.50. Years passed (as they often do in grad school) and we became good friends. It is a remarkable experience to have such a partner in every facet of life. We have helped each other through the hard times and celebrated together when experiments actually work. She brings balance to my life and so much more. Julie works harder than anyone I know. Sometimes trying to keep up with her wears me out, but I thrive off of her energy. A life with Julie is worth a thousand lives with anyone else.

Julie is not the only friend I have made in graduate school. Nelson Lau and I have been great friends since our first year at MIT. He has been a great sounding board for life's trials and tribulations over the last several years. He and his wife, Dianne Schwarz, are among the nicest, most generous people I've met.

Another extremely dependable friend over the years has been Tate Kauffman, who I met while working in the lab of Phil Meneely at Haverford College. He and I thought we might be distant cousins once, due to the potential for shared ancestry in central Pennsylvania. Regardless of the biology, Tate, along with his wife Lara, will always be family to me (even if they are both M.D.s).

I would like to thank some of the other friends I have had the pleasure to work with the past several years. Alexander Gimelbrant has been a great lab neighbor for the last 5 years. I will miss his humor and his advice. Nandita Singh did a wonderful job of teaching me the FISH technique. I appreciate all of the work that she and Sasha Ebrahimi did with respect to the first characterization of coordination that laid the groundwork for the rest of my thesis. Michael Tackett and I have suffered together at times, but I have been glad to have someone to complain to. John Hutchinson and Guilherme Neves have been wonderful colleagues and friends, brightening everyone's path that crosses theirs.

These are not the only friends I have made in lab. I would like to express my deepest thanks to Andy Chess for being not only and advisor but a friend. I expect that both of us will be in this business for quite some time and it's nice to know that I will have a very creative friend and advisor to call upon when times get tough.

My thesis committee members, David Page and Paul Garrity have provided me with great support and guidance over the years. I will miss our discussions and the reassurance of having them in my corner through all these years. I also would like to thank the Howard Hughes Medical Institute for supporting me as a predoctoral fellow. It has meant a lot to me to be a part of such a great program.

I would also like to thank my college advisor and genetics professor, Philip Meneely. Phil showed me that biology is a dynamic science. Without him, I know that I would have been an economics major and my life would have taken a totally different path. So I guess what I'm trying to say is that this is all Phil's fault.

TABLE OF CONTENTS

Chapter One:

Introduction: Monoallelic Expression and Asynchronous Replication 7
- The comparative value of haploidy versus diploidy 8
- Dosage compensation mechanisms in worms and flies 10
- Dosage compensation in mammals: X-chromosome inactivation 12
- A number of autosomal genes are monoallelically expressed in mammals 19
- Asynchronous replication coincides with monoallelic expression 28
- Prevalence of asynchronous replication and monoallelic expression 35
- Coordination of asynchronous replication between disparate loci 36
- Summary of Thesis 39
- References 43

Chapter Two:

Coordinated replication timing of monoallelically expressed genes along human autosomes. 59
- Abstract 60
- Introduction 61
- Results 63
  - Asynchronous replication in humans 63
  - Coordination of asynchronous replication in human cells 68
  - Asynchronous replication in trisomies 75
- Discussion 81
- Materials and Methods 85
- Acknowledgements 87
- References 88

Chapter Three:

A search for *Xist*-like nuclear transcripts in the human genome identifies two conserved noncoding RNAs on Chromosome 11 90
- Abstract 91
- Introduction 92
- Results 95
  - An array-based approach to identify ubiquitously expressed nuclear RNAs 95
  - Two noncoding RNAs enriched in the nuclei of human cells 97
  - RNA FISH of *Xist*, *NEAT1*, and *NEAT2* 100
  - Examining transcription around *NEAT1* and *NEAT2* 104
  - Functional analysis of *NEAT1* and *NEAT2* using RNA interference 105
  - Comparative genomic analysis of *hNEAT1* and *hNEAT2* 106
- Discussion 114
- Materials and Methods 117
- References 122

Conclusions and Afterword **129**
Conclusions **130**
Potential mechanisms of coordination **131**
Asynchronous replication as a potential means to maintain genomic stability **133**
Examining the evolutionary relationship between X-chromosome inactivation and autosomal coordination **135**
Generating genetically tractable systems for the study of monoallelic expression **139**
Examining established epigenetic regulators for their role in asynchronous replication, coordination, and monoallelic expression **141**
Other approaches for identifying *Xist*-like RNAs in the mammalian nucleus **142**
Summary **145**
References **146**

Appendix One:
Coordination of the random asynchronous replication of autosomal loci **153**

Appendix Two:
Using RNase-treatment to examine the structural details of RNA transcripts within chromatin **178**

Appendix Three:
Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells **185**

Introduction:

Monoallelic Expression and Asynchronous Replication

The comparative value of haploidy versus diploidy

There is a preponderance of diploidy all around us. Indeed, almost all higher organisms have a predominantly diploid life cycle, an observation which has been suggested as evidence itself of the relative advantage of the diploid genome (Adams and Hansche, 1974). Evolutionarily, the emergence of diploidy resulted in a number of advantages to the diploid cell. However, the manifestation of diploidy also introduced a number of complexities that have necessitated further adaptations.

One main characteristic of the diploid state is that deleterious mutations can often be masked in the heterozygote. Thus, the transition from haploid to diploid can also be seen as the first true disconnect between genotype and phenotype. Individual mutations that would otherwise be lethal or reduce fitness in a haploid are instead often masked as recessive alleles in the heterozygous diploid. This logic stems from the concept of the diploid cell as a pair of haploid genomes. Each gene is represented by two independently functioning alleles with relatively parallel transcription states due to their exposure to the same milieu of transcription factors and regulatory elements. One might consider this the "spare genome" reason for diploidy.

This view of the diploid genome turns out to be overly simplistic in a number of ways. Diploidy is not merely a means by which to carry an extra genome in reserve. For instance, in addition to masking recessive alleles, heterozygosity often leads to new phenotypes. These phenotypes, under certain conditions, are sometimes better adapted to a particular environment than the phenotype of either homozygous state. Perhaps the best-known example of heterozygous advantage is sickle cell anemia (Aidoo et al., 2002; Allison, 1964). In regions where malarial infection is endemic, the sickle-cell

hemoglobin allele persists despite homozygotes suffering the devastating consequences of sickle-cell anemia. This persistence is believed to result from a reduced susceptibility to malaria in heterozygous individuals relative to either homozygote. Similar explanations have been proposed to explain the high incidence of cystic fibrosis in populations of European descent. In particular, a heterozygous advantage may be conferred to carriers of the mutant allele in the form of increased resistance to a number of infectious agents, including the causative agent of typhoid fever, *Salmonella typhi* (Pier et al., 1998; van de Vosse et al., 2005).

Heterozygous advantage is merely one example of how the view of the diploid genome as two haploid genomes is overly simplistic. Further complexities in the model arise from the fact that, for many higher eukaryotes, the entire genome is not necessarily diploid. In particular, it is often the case that different sexes of a particular species have slightly different karyotypes. In mammals, for instance, males have only one copy of the X chromosome, whereas females have two copies. Left unregulated, such a difference in copy number of each X-linked gene would mean that males would produce half as much X-linked transcripts as females. Instead, the level of X-linked gene expression needs to be compatible with both sexes, despite differences in their karyotype. Throughout evolution, many species have dealt with this dilemma in a number of ways, known collectively as dosage compensation (Lucchesi et al., 2005; Lyon, 1986; Parkhurst and Meneely, 1994).

Dosage compensation mechanisms in *C. elegans* and *D. melanogaster*

One of the hallmarks of dosage compensation in mammals is the chromosome-wide silencing of half of the X-linked genes in female cells (Lyon, 1961; Lyon, 1986). Though this strategy is highly effective at normalizing the amount of X-linked transcription of females to that of males, it is not the only strategy found in nature. Indeed, while the end result of dosage compensation is relatively conserved between species, there are many different mechanisms utilized to normalize transcription between different sexes of the same species. The equilibration of transcription between sexes with different karyotypes can occur by reducing transcription in homogametic (XX) individuals to levels found in heterogametic individuals (XY or XO). Alternatively, transcription can be upregulated in individuals with one copy of the chromosome to approximate the levels found in individuals with two copies. Within these different general approaches, many different strategies exist.

Rather than reducing the amount of X-linked transcription in XX females to the level of XY males, dosage compensation in *Drosophila melanogaster* involves an increase in X-linked transcription in male flies (Belote and Lucchesi, 1980a). This hypertranscription in males depends on a number of factors, including the ratio of X chromosomes to autosomes (Baker and Ridge, 1980; Bridges, 1921), a cascade of regulatory events, and the assembly of a regulatory complex, MSL, on the male X chromosome (Meller et al., 2000). The MSL (Male Specific Lethal) complex consists of five proteins that are essential for male development but dispensable in females (Belote and Lucchesi, 1980b; Gorman et al., 1993). In addition, at least two noncoding RNAs, *roX1* and *roX2* serve a role in the formation of the MSL complex (Amrein and Axel,

1997; Franke and Baker, 1999; Park et al., 2002). These noncoding RNAs seem to function at least somewhat redundantly in the developing male, because only males lacking both *roX* RNAs experience lethality (Franke and Baker, 1999). This redundancy occurs despite very little homology between the two transcripts as well as a great discrepancy in size (the *roX1* transcript is greater than 4 kb long whereas *roX2* is only 0.6 kb long). While the function of the roX genes is not entirely understood, it is believed that they play a crucial role in targeting the MSL complex to the male X. Indeed, sites of *roX1* and *roX2* transcription are themselves points of initial association between the MSL assembly and the male X chromosome (Kelley et al., 1999).

In the roundworm, *Caenhorhabditis elegans*, dosage compensation utilizes an approach opposite that seen in *Drosophila*. Rather than increasing X-linked transcription in males, hermaphroditic (XX) X-linked transcription is reduced to levels similar to that of males (XO) (Meyer and Casson, 1986). As seen in flies, the ratio of X chromosomes to autosomes also plays a crucial role in dosage compensation (Madl and Herman, 1979), but the rest of the pathway is so dissimilar as to strongly suggest different evolutionary histories for the two mechanisms. In worms, a complex of proteins known as the dosage compensation complex (DCC) regulates transcription from the X chromosome in hermaphrodites (Wood et al., 1997). The downstream effects of many factors result in the condensation of X-linked chromatin, finely tuned to reduce the levels of X-linked transcription by 50% on each X chromosome found in hermaphrodites such that the sum total of the two is equal to that of XO males.

Mammalian dosage compensation takes the same general approach as worms in terms of the equilibration of the level of X-linked transcription between heterogametic

males (XY) and homogametic females (XX). Like worms, dosage compensation is targeted not to males, but to females; and like worms, the end result of this dosage compensation machinery is to reduce the level of X-linked gene expression in females (hermaphrodites) to a level equal to that in males. Whereas the hermaphroditic *C. elegans* nucleus accomplishes this reduction while treating the two X chromosomes the same, the mammalian female nucleus treats its two X chromosomes in a highly inequitable manner.

It is interesting to note that the approach of reducing X-linked expression in females to that of males seems somewhat paradoxical. Such a strategy does not address the insufficient transcription of X-linked genes in XO males relative to autosomal genes. On the surface, it would seem that males have a problem, not females. After all, autosomal monosomy is not well tolerated, so why should the loss of one X chromosome by males be any less detrimental to development? With this in mind, male hypertranscription seems a more appropriate evolutionary response to this imbalance, elevating X-linked male transcripts to their previous levels when they were part of the autosomal transcriptome. It has been suggested that X inactivation may have been established to correct an imperfect solution to the initial problem of male monosomy. If during evolution, X-linked hypertranscription was not limited to males, mechanisms would need to be employed in order to reduce female X-linked transcription to its normal levels (Charlesworth, 1996).

Dosage compensation in mammals: X-chromosome inactivation

Like the species discussed above, mammalian females are the homogametic sex (XX), whereas males are the heterogametic sex (XY). In the mammalian female nucleus, one X chromosome is transcriptionally active (X_a) and the other X chromosome is transcriptionally inactive (X_i) (Lyon, 1961). Established around the time of implantation, the inactive X chromosome is, among other things, late replicating (Priest et al., 1967; Takagi, 1974), coated with a noncoding RNA (Brockdorff et al., 1992; Brown et al., 1992; Clemson et al., 1996), highly condensed, methylated (Wolf et al., 1984), packaged with hypoacetylated histones H4, H2A, and H3 (Belyaev et al., 1996; Jeppesen and Turner, 1993), as well as the histone variant macroH2A (Costanzi and Pehrson, 1998). The extreme epigenetic inequality resulting from X-chromosome inactivation occurs despite the fact that each copy of the X chromosome is extremely similar genetically. Indeed, these X chromosomes begin development as relative equals. The voyage each takes is a journey of chance followed by inexorable fate. This voyage is well understood between some waypoints and surprisingly uncharacterized at others.

One of the earliest differences in development between the two X chromosomes is a difference in replication timing. Strikingly, both alleles of most DNA segments usually replicate at roughly the same point in S-phase (Goren and Cedar, 2003). Different parts of the genome will replicate at different times in S-phase, but the two alleles of a DNA segment will be copied within a short interval during S-phase. This type of DNA replication is known as synchronous replication. In the case of the two X chromosomes, all of the DNA segments on one X chromosome tend to replicate later than their allelic counterparts on the other X (Schmidt and Migeon, 1990). This difference in replication

timing between two alleles is known as asynchronous replication. Late replication is often associated with inactive transcriptional states (Simon et al., 2001), so it might make sense that the inactive X chromosome should replicate later in S-phase than the active X chromosome. This is in fact the case, but it has been unclear whether the late replication is a consequence of all the epigenetic modifications on the inactive X chromosome, or an early mark to distinguish the future inactive X chromosome from the active X chromosome. Strikingly, asynchronous replication is observed between the two X chromosomes in embryonic stem (ES) cells, prior to the majority of epigenetic modifications awaiting one of the two X chromosomes (Gribnau et al., 2005). This asynchronous replication, while chromosome-wide, does not seem fixed in development, as ES lines which will later make a non-random choice with respect to X inactivation still seem to randomly choose which allele will replicate late. One caveat of such studies is that the disruption that makes X inactivation non-random in these experiments may in fact be so severe that it interferes with the normal chain of events leading to X inactivation. Specifically, the ES cells examined have disruptions in one copy of the noncoding RNA, *Xist*, which is necessary early in development for X-chromosome inactivation (Wutz and Jaenisch, 2000).

The large noncoding RNA, *Xist* (X inactive-specific transcript), is indispensable for creating an inactive X chromosome. The *Xist* locus produces a spliced, polyadenylated transcript of up to 19.3 kb in humans and 17.9 kb in mice (Brockdorff et al., 1992; Brown et al., 1992; Hong et al., 2000). The discovery of *Xist* was greatly facilitated by previous studies that identified a region of the X chromosome that was required for establishing an inactive X chromosome in *cis* (Brown et al., 1991b).

Disruptions in the X inactivation center (Xic) prevent inactivation of the chromosome containing the deletion and result in nonrandom inactivation of the intact X chromosome. When a large, noncoding transcript was identified within this region, a number of observations suggested a functional role for *Xist* in the creation of an inactive X (Brown et al., 1991a).

The first of these observations was that *Xist* was expressed in female (XX) cells, but not in male (XY) cells. Furthermore, the level of *Xist* expression increased proportionally as the number of X chromosomes increased. For instance, XXX cells expressed twice as much *Xist* as XX cells. This was revealing, because it was already known that only one X chromosome remains active in a cell, thus XXX cells have two inactive X chromosomes compared to the one inactive X chromosome in XX cells. In addition, it has been shown that *Xist* is transcribed exclusively from the inactive X chromosome. Transcription of *Xist* is necessary for establishing an inactive X chromosome, but is dispensable for its maintenance in differentiated somatic cells (Brown and Willard, 1994; Csankovszki et al., 1999; Wutz and Jaenisch, 2000).

Perhaps the most striking characteristic of *Xist* is its localization in the cell. Using fluorescence *in situ* hybridization (FISH), *Xist* RNA was observed to coat the entire length of the inactive X chromosome (Brown et al., 1992; Clemson et al., 1996). In mouse ES cells, *Xist* RNA is detected by FISH as a punctate signal at the point of transcription in each X chromosome in females and on the one X chromosome in male cells (Panning et al., 1997; Sheardown et al., 1997). With differentiation, this signal is transformed as *Xist* RNA rapidly spreads across the presumptive inactive X chromosome in females and disappears from the active X in female and male cells (Lee and Jaenisch,

1997; Lee et al., 1999b; Panning et al., 1997). The spreading of *Xist* RNA is not well understood, but it has been hypothesized that some property of the X chromosome might facilitate spreading, as X;autosomal translocations show greatly reduced spreading of *Xist* RNA from ectopic autosomal sites (Hall et al., 2002). One model put forth is that repetitive elements such as LINEs might serve as way stations for *Xist* spreading (Hansen, 2003; Lyon, 1998).

Many of the early clues about X chromosome inactivation came from cytological studies, made possible due to the unique structure of the inactive X chromosome. While it took nearly forty years to discover that a noncoding RNA, *Xist*, coated the inactive X, in the 1950's, M. Barr and colleagues were observing a highly condensed structure in the nucleus (Barr and Moore, 1957). The "Barr body" stains brightly with the fluorescent DNA stain, 4'-6-Diamidino-2-phenylindole (DAPI). Cytologists noticed that the number of Barr bodies in a nucleus was always equal to one less than the total number X chromosomes (Grumbach et al., 1963; Harden, 1961). This observation gave rise to the so-called "N-1" rule, which described the number of inactive X chromosomes as the total number of X chromosomes minus one. While this "N-1" rule makes sense from a cytological standpoint, it is perhaps more relevant to consider the number of active X chromosomes in a given nucleus. Regardless of the total number of X chromosomes in the nucleus, the number of active X chromosomes is always equal to one.

A number of chromatin modifications have been identified which proceed to transform the *Xist*-coated X chromosome into the impenetrable heterochromatin of the Barr body (Lucchesi et al., 2005). These epigenetic events parallel many of the

chromatin modifications that are associated with facultative heterochromatin found interspersed throughout the mammalian genome.

Various histone modifications, often termed the "histone code," have been associated with silent transcriptional states and are likewise used in the creation of a stable inactive X chromosome (Jenuwein and Allis, 2001). One of these modifications is methylation of histones. In X inactivation, histone H3 is dimethylated at lysine 9 (H3K9me2) and there is evidence that trimethlyation at lysine 27 (H3K27me3) is also necessary for silencing. This epigenetic mark depends on the activity of the Class I Polycomb Group (PcG) proteins, Eed, Ezh2, and Suz12 (de la Cruz et al., 2005; Erhardt et al., 2003; Plath et al., 2003; Silva et al., 2003). Recent evidence suggests that both H3 lysine 27 trimethylation and histone H4 lysine 20 monomethylation may serve as a silencing-independent chromosomal memory, established by Xist expression prior to differentiation (Kohlmaier et al., 2004). Histone methylation is known to recruit epigenetic modulators responsible for establishing heterochromatin. One of these proteins, heterochromatin protein 1 (HP1) binds to H3K9me2 and is thus recruited to the X_i (Chadwick and Willard, 2003).

In addition to methylation of specific residues, another modification of histones is associated with heterochromatin: hypoacetylation of histones H3 and H4 (Richards and Elgin, 2002; Turner, 1998). Indeed, histones on the inactive X chromosome are highly hypoacetylated by histone deacetylases (HDACs), which may contribute to the heterochromatic nature of the inactive X (Belyaev et al., 1996; Boggs et al., 1996; Jeppesen and Turner, 1993; Keohane et al., 1996). Another key modification of the inactive X is the utilization of a histone variant, macroH2A. Immunofluorescent

microscopy has demonstrated the accumulation of macroH2A along the length of the inactive chromosome (Costanzi and Pehrson, 1998). This histone variant may silence gene expression either by interfering with transcription factor binding or by preventing nucleosome remodeling by SWI/SNF complexes (Angelov et al., 2003).

Modifications to the histones packaged with the inactive X chromosome are also paralleled by changes to the chromosomal DNA itself. Methylation of DNA, particularly CpG islands in promoters upstream of transcriptional start sites is associated with gene silencing throughout the genome. This canonical epigenetic modification also occurs on the inactive X chromosome, with hypermethylation of CpG islands on the X_i and hypomethylation of the same sequences on the X_a. Proteins known as DNA methyltransferases (DNMT) methylate DNA and mutational analysis indicates an important role for DNMT3b in establishing the hypermethylated state CpG islands on the inactive X (Hansen, 2003; Hansen et al., 2000; Hansen et al., 1999).

While *Xist* transcription is crucial for the establishment of X-chromosome inactivation, the numerous chromatin modifications which follow are likely responsible for the remarkable stability of the inactive X throughout successive somatic cell divisions.

The end result of inactivating one of the X chromosomes in female cells in mammals results in the reduction of X-linked gene expression to a level comparable to that in male cells. Unlike dosage compensation in *C. elegans*, which treats each X chromosome as equals in XX animals, the inequity inherent in X inactivation has other fundamental consequences to gene expression. Chief among these is that in any given female cell, that cell is effectively hemizygous at each X locus. The result of this is that

in each female cell, X-linked genes are expressed from only one of two alleles. This is known as monoallelic expression and differentiates X-linked gene expression in mammals from that in other organisms such as *C. elegans*. Monoallelic expression of X-linked genes is mosaic, as each progenitor cell in the blastocyst makes an independent, random choice. As a consequence, usually half the cells within a female mammal will express paternally derived X-linked genes, and half the cells will have made the opposite choice, thereby transcribing off the maternally inherited X chromosome. It is estimated that greater than 15% of X-linked genes "escape" X inactivation and are therefore biallelically expressed (Carrel and Willard, 2005). It is curious to note that *Xist* itself is monoallelically expressed from the inactive X chromosome in a fashion opposite to the rest of X-linked genes.

The hallmark of mammalian dosage compensation is that X chromosome inactivation is a stably maintained, chromosome-wide choice. It is the result of a random decision that creates epigenetic inequity from an initial state of equality. The major differences that exist between flies, worms, and mammals suggest that dosage compensation has evolved independently a number of times. Strategies such as those employed by *C. elegans*, which treat the two X chromosomes equally, demonstrate that gene expression on the X chromosome can be reduced without invoking a chromosome-wide epigenetic choice. The chromosome-wide nature of X chromosome-inactivation seems to be explained by the spreading in *cis* of the *Xist* transcript along the length of one of two X chromosomes. While many of the events subsequent to stable *Xist* transcription on one X chromosome have been deciphered, the events prior to stable *Xist* expression remain relatively enigmatic.

The description of another noncoding transcript within the Xic of mice has helped explain some of the earlier events in X inactivation. *Tsix*, so named because it is transcribed antisense to the *Xist* transcript, serves to silence *Xist* transcription in *cis* (Lee et al., 1999a). Transcription of *Tsix* is antagonistic to the expression of *Xist* from the same allele. In this way, *Tsix* is expressed from the active X chromosome, where *Xist* is silent; *Tsix* is transcriptionally silent on the inactive X chromosome, where *Xist* is expressed. These observations, while important, do not change one of the fundamental questions surrounding X inactivation. How does each cell randomly choose between two X chromosomes whose DNA sequence is nearly identical? As this question remains unanswered, perhaps some insight can be garnered by examining other examples of random epigenetic choices.

A number of autosomal genes are monoallelically expressed in mammals

While the result of X inactivation is monoallelic expression of X-linked genes in each female cell, monoallelic expression is by no means limited to the X chromosome in mammals. Indeed, a number of autosomal loci, scattered across the mammalian genome, are also expressed from only one of two alleles.

Certainly, the majority of genes in the mammalian genome are biallelically expressed. Our current understanding of gene expression as a function of exposure to specific transcription factors helps explain this phenomenon. If the transcription of a particular gene depends predominantly on whether upstream regulatory elements are exposed to the proper set of trans-acting transcription factors, biallelic expression is the likely result. This is because both alleles of a gene will have the same upstream

regulatory sequences, which, as a result of being in the same nucleus together, will be exposed to the same set of transcription factors at any given time. Thus, if the conditions exist to activate one allele for transcription, they also favor the activation of the other allele. This simplistic model of gene expression probably describes the regulation of 95% of autosomal mammalian genes. The remainder of this introduction will concentrate on the other 5%.

The functional redundancy afforded by biallelic expression is certainly beneficial with regard to most mammalian genes. If one copy of a gene is disrupted, the other allele, expressed in a parallel fashion, can often compensate for the loss. For a number of mammalian genes, however, such benefits cannot outweigh the detriment caused to the organism by expressing two copies of these genes at the same time. One class of genes that cannot be biallelically expressed without harming the organism are those that reside on the X chromosome of females. As discussed above, biallelic expression of these X-linked genes would result in a dosage imbalance of X-linked transcription between males and females. In addition to X-linked genes, a number of autosomal genes function in such a way that biallelic expression would also be detrimental to the overall fitness of the organism.

Autosomal mammalian genes that are randomly monoallelically expressed include odorant receptors (Chess et al., 1994; Ishii et al., 2001; Serizawa et al., 2000), immunoglobulins (Davie et al., 1971; Litwin, 1972; Pernis et al., 1965; Wolf et al., 1971), T-cell receptors (Roehm et al., 1985), interleukins (Bix and Locksley, 1998; Hollander et al., 1998; Rhoades et al., 2000), natural killer-cell receptors (Held et al., 1995), and pheromone receptors (Belluscio et al., 1999; Rodriguez et al., 1999). The expression of

these genes is distinct from that of imprinted genes. Imprinted genes are monoallelically expressed in a non-random manner (Efstratiadis, 1994). Whereas random monoallelic expression represents a random choice, made independently by each cell at some point in development, there is no choice of which allele to express in the case of imprinted genes. Rather, the monoallelic expression of any given imprinted genes is stereotypically from same allele.

The immune system seems especially dependent on random monoallelic expression for its function. The way the immune system generates antibody diversity is a good example of monoallelic expression and the pitfalls that biallelic expression could introduce in some circumstances. For instance, the monoallelic expression of immunoglobulin genes in B cells, also known as allelic exclusion, is crucial for the formation of a functioning adaptive immune system (Bergman and Cedar, 2004). Adaptive immunity depends on the production of specific antibodies to antigens encountered throughout the lifespan of an individual. Each B cell makes a specific antibody, which consists of two copies of a light chain immunoglobulin protein and two copies of a heavy chain immunoglobulin. (Although it is not critical for this discussion, there are two light chain genes: immunoglobulin kappa (IGK) and immunoglobulin lambda (IGL) which seem to serve predominantly redundant roles in antibody formation, as either gene (but not both) can be used in the formation of one antibody.) Each immunoglobulin expressed in a B cell is the result of a random recombination event within the genomic DNA at that immunoglobulin locus (Hozumi and Tonegawa, 1976). Since each allele of a given immunoglobulin gene would produce different transcripts due to independent, random rearrangement, biallelic expression of the loci would mean

there were two completely different immunoglobulin chains expressed at the same time. Such a situation would greatly interfere with the function of the adaptive immune system, as it gets its specificity from B cell clones that target specific antigens by ensuring that each B cell expresses only one type of antibody. By monoallelically expressing the two light chain immunoglobulins IGK, and IGL as well as the heavy chain IGM, B cells ensure that they express only one rearrangement for any particular chain and thus produce only one antibody. As similar rearrangements are a requisite part of T-cell receptor development (Chien et al., 1984; Hedrick et al., 1984a; Hedrick et al., 1984b), it is perhaps not surprising that T-cell receptors are also monoallelically expressed in T-cells (Roehm et al., 1985).

The first random monoallelic expression described outside the immune system was the monoallelic expression of odorant receptors in olfactory neurons (Chess et al., 1994). Both the mouse and human genomes contain a large number of odorant receptors. Genomic analysis shows that there are roughly 1500 odorant receptor genes in mice and 950 in humans (Glusman et al., 2001; Young et al., 2002; Young and Trask, 2002). Nearly two-thirds of human odorant receptor loci are thought to be pseudogenes (Niimura and Nei, 2003; Niimura and Nei, 2005), though there is some evidence that non-functional pseudogenes can be expressed (Crowe et al., 1996). Each mature olfactory neuron is believed to express only one odorant receptor gene, the result of a random choice between a slightly restricted set of the complete repertoire of receptors (Nef et al., 1992; Ressler et al., 1993; Vassar et al., 1993). This pattern of expression is further restricted by the fact that only one allele of the chosen odorant receptor is expressed (Chess et al., 1994; Ishii et al., 2001; Serizawa et al., 2000).

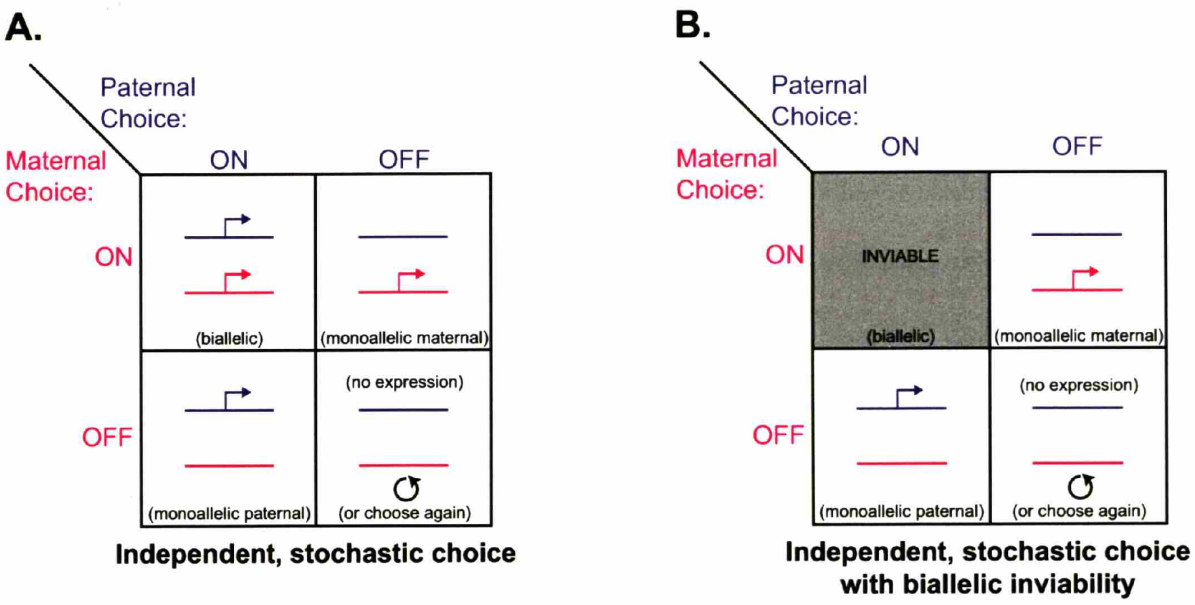
Conceptually, monoallelic expression might occur in a number of different ways (Figure 1). Monoallelic expression can result from a situation in which each allele of a given gene makes an independent, stochastic choice to either be active or inactive. If the probability of expression at each allele is 50%, 4 types of expression patterns will result. In equal numbers, cells will express both alleles, express only maternal allele, express only the paternal allele, or express neither allele. If some sort of feedback loop requires the expression of at least one allele of the gene, 1/3 of the cells express the maternal allele, 1/3 express the paternal allele, and 1/3 express both alleles (Figure 1A). This type of expression pattern describes what is observed for a number of genes, including natural killer cell receptors, interleukins, Tlr-4, and p120-catenin (Gimelbrant et al., 2005; Rhoades et al., 2000; Saleh et al., 2004). In the case of natural killer cell receptors, this phenomenon results from a bidirectional switch at each locus, resulting in an independent choice between active and inactive transcription at each allele (Saleh et al., 2004).

Some monoallelically expressed genes are never biallelically expressed, however. These genes include immunoglobulins, T-cell receptors, and odorant receptors. This exclusively monoallelic expression is important, as biallelic expression of these genes would be particularly detrimental. Exclusive monoallelic expression has implications for the mechanisms underpinning such regulation. One mechanism for this type of tight regulation of monoallelic expression might be that allelic transcription states are still determined independently, but that biallelic transcription results in cell death (Figure 1B). However, exclusive monoallelic expression could also occur without the need for such measures. Changing the efficiency of choice at both alleles reduces the likelihood of

Figure 1.

Models of Monoallelic Expression. **(A)** Monoallelic expression of a gene can result if the probability of each allele's activation is independent. If the maternal allele is expressed in 1/2 of cells, and the paternal allele is expressed in 1/2 of cells, four patterns of expression will result. Biallelic expression, exclusively maternal expression, exclusively paternal expression, and no expression are equally likely outcomes. In situations where lack of expression is not seen, additional rounds of selection may occur until stable expression of at least one allele results. **(B)** If biallelic expression is detrimental to the fitness of the cell, cells which independently arrive at expression of both alleles may not be present, due to cell death. **(C)** By modifying the probability of each allele's activation, the likelihood of biallelic expression can be reduced, without requiring cell death. When the probability of each allele's expression is reduced by n, the probability of biallelic expression is reduced by n^2. One consequence of reducing the probability of expressing each allele is that the most likely outcome becomes expression from neither allele. This could be achieved by successive cycles of choice, followed by feedback inhibition of further rounds after one allele is expressed. **(D)** The presence of a singularity within the nucleus can explain monoallelic expression. If only one copy of a necessary *cis*-acting transcriptional activator existed in the nucleus, it would be physically constrained to activate only one locus. If such an activator was involved in odorant receptor expression, it could simultaneously explain the random choice of one receptor and the monoallelic expression of the chosen locus.

FIGURE 1


A.
Paternal Choice:
Maternal Choice:
ON
OFF
ON
OFF
(biallelic)
(monoallelic maternal)
(no expression)
(monoallelic paternal)
(or choose again)
Independent, stochastic choice
B.
Paternal Choice:
Maternal Choice:
ON
OFF
ON
OFF
INVIABLE
(biallelic)
(monoallelic maternal)
(no expression)
(monoallelic paternal)
(or choose again)
Independent, stochastic choice with biallelic inviability
C.
Paternal Choice:
Maternal Choice:
ON
OFF
ON
OFF
Low frequency of activation with successive rounds of choice
D.
or
(monoallelic paternal)
(monoallelic maternal)
Singularity of activating complex results in monoallelic expression ("One Blob")

biallelic expression at a greater rate than the monoallelic expression of either allele (Figure 1C). In particular, the absence of biallelic expression could result from independent activation of each allele of a given gene, so long as the efficiency of activation is low and there is some sort of feedback mechanism that prevents the activation of the second allele once one is active. Indeed, such negative feedback inhibition exists in immunoglobulin rearrangement (Rusconi and Kohler, 1985) and recent evidence suggests the existence of one with respect to odorant receptor choice as well (Serizawa et al., 2003).

Another mechanism to achieve exclusive monoallelic expression is through the requirement of a singular activating complex. Such a complex must be required for expression and present as only one copy in the nucleus. One might refer to this as the "one blob" hypothesis (Figure 1D). In such a model, one blob could not activate both alleles of a given gene because of the physical constraints of the nucleus. Remarkably, evidence for such a method of gene regulation exists in the causative agent of human sleeping sickness, *Typanosoma brucei* (Navarro and Gull, 2001).

Trypanosomes are coated with a single type of variant surface glycoproteins (VSG). There are 20 VSG expression sites scattered across the trypanosome genome, but only one VSG is expressed at any time (Borst and Ulbert, 2001). This stochastic choice has many parallels to the choice of one odorant receptor out of a repertoire of over a thousand, and until recently, the mechanisms of trypanosome coat switching were equally unresolved. Recent evidence suggests that VSG expression depends on the presence of a transcriptional complex with striking singularity within the trypanosome nucleus (Navarro and Gull, 2001). This complex localizes to the active VSG locus, and the

presence of only one complex within the nucleus likely explains the monoallelic expression of the VSG coat.

Studies into the mechanistic underpinnings of monoallelic expression have been greatly complicated by some of the characteristics of random monoallelic expression. The most problematic of these features is the highly restricted pattern of expression for most of the genes discussed above. This restriction to specific cell types has hindered efforts to study random choice in easily tractable systems, such as mammalian cell culture. Investigation of random choice establishment in endogenous tissues is also complicated by the fact that different cells within the same tissue make independent choices. Thus, the study of random choice in tissues has often been restricted to single cell analysis, which is replete with its own set of hurdles.

Random monoallelic expression of a number of genes is clonally inherited in subsequent cell divisions (Gimelbrant et al., 2005; Mostoslavsky et al., 2001; Singh et al., 2003). The creation of a large number of clonal cell lines from human and murine lymphoblastoid cells, has facilitated the identification of genes which are monoallelically expressed in more experimentally tractable cell types. In particular, immunoglobulins, interleukins, Toll-like receptor 4 (Tlr-4) and at least one gene outside the immune system, p120-catenin are monoallelically expressed in some of these clonal lines (Gimelbrant et al., 2005). The type of monoallelic expression observed in these cells seems to be the nonexclusive type, as some clones express the maternal allele, some express the paternal allele, and others express both. This type of monoallelic expression may indicate that the transcriptional state of each allele is determined independently, a characteristic which can likely be examined in such cell lines. Notably, in the case of p120-catenin and Tlr-4,

monoallelic expression appears to be cell-type specific, as the same genes are biallelically expressed in all fibroblastoid clones examined (Gimelbrant et al., 2005).

Asynchronous replication coincides with monoallelic expression

As mentioned earlier, the default mode of DNA replication in mammalian cells is such that alleles of the same gene replicate within the same small window of S-phase, known as synchronous replication. However, a fraction of the genes within the genome do not replicate both their alleles at the same time. When one allele replicates earlier in S-phase than the other, this is known as asynchronous DNA replication. Thus, as the inactive X chromosome replicates much later in S-phase than the active X chromosome, X-linked genes replicate asynchronously (Schmidt and Migeon, 1990).

The time at which particular genes replicate in S-phase often gives an indication of the transcriptional state of those genes. In particular, late replication is typically associated with less accessible chromatin states and the transcriptional silencing they confer. Correspondingly, early replication is associated with more transcriptionally active loci. This relationship between replication timing and expression is suggested not only from studies of the X chromosome, but from autosomal loci as well.

Early studies of the relationship between replication timing showed a relationship between DNase I sensitivity (an indication of open chromatin) and early replicating bands on chromosomes (Kerem et al., 1983). With the development of new techniques such as fluorescent *in situ* hybridization, studies of individual genes provided additional evidence of the link between replication timing and expression.

One locus that has been fundamentally important with respect to our understanding of the importance of a number of epigenetic phenomena, including replication timing, is the beta globin locus (Chakalova et al., 2005). Beta globin is expressed in the erythroid lineage, where the locus replicates early in S-phase, but is both transcriptionally silent and late replicating in other cell types (Dhar et al., 1989; Epner et al., 1988; Simon et al., 2001). This relationship between replication timing and expression has also been shown for the cystic fibrosis (CF) locus (Selig et al., 1992). This connection has also been reinforced through the use of an interesting microinjection technique to inject DNA into the nuclei at different portions of S-phase (Zhang et al., 2002). DNA that was injected early in S-phase was more likely to be expressed than the same DNA injected in late S-phase, suggesting that factors present in late S-phase are inhibitory towards gene expression. More recently, genome-wide surveys of replication timing in *Drosophila* and humans have also suggested a link between the replication timing in S-phase and transcriptional activity (Schubeler et al., 2002; Woodfine et al., 2004).

Differences in replication timing can be observed not only for the same loci within two different cell types, but can also exist between alleles of a single gene. As mentioned earlier, X-chromosome inactivation results in the asynchronous replication of X-linked genes. A surprisingly large proportion of autosomal genes are also asynchronously replicated. As is the case for developmental differences in replication timing, this asynchronous replication has implications with respect to expression as well. All monoallelically expressed genes heretofore examined are also asynchronously replicating. Random asynchronous replication has been observed for X-linked genes,

odorant receptors, immunoglobulins, T-cell receptors, natural killer cell receptors, interleukins, Tlr-4, and p120-catenin (Bix and Locksley, 1998; Chess et al., 1994; Ensminger and Chess, 2004; Gimelbrant et al., 2005; Mostoslavsky et al., 2001; Schmidt and Migeon, 1990; Singh et al., 2003). In addition, genes that are monoallelically expressed in a non-random, imprinted manner are asynchronously replicated in a non-random manner (as the paternal allele is typically early replicating) (Kitsberg et al., 1993).

Unlike the developmentally dynamic replication timing of beta globin, monoallelically expressed genes are asynchronously replicated in all cell types examined, including fibroblastoid and lymphoblastoid cells. Thus, asynchronous replication can be observed in cell types in which neither allele is expressed. This has proved useful in the study of monoallelically expressed genes, because many of these genes are only expressed in very restricted, relatively experimentally intractable cell types. For instance, odorant receptors are only expressed in specific, post-mitotic olfactory neurons, yet the DNA encoding these genes replicates asynchronously in fibroblasts, lymphoblasts, and embryonic stem cells (Chess et al., 1994; Ensminger and Chess, 2004; Singh et al., 2003).

The link between replication timing and expression has been more difficult to study in monoallelically expressed genes than in developmentally regulated genes such as beta globin. Obviously, it is impossible to examine the replication timing of odorant receptors in olfactory neurons, since post-mitotic neurons do not normally undergo DNA replication. However, some other genes are monoallelically expressed in mitotic cell types. Where it has been examined, asynchronous replication is random when monoallelic expression is random (Chess et al., 1994; Ensminger and Chess, 2004;

Mostoslavsky et al., 2001; Singh et al., 2003); it is non-random when monoallelic expression is imprinted (Kitsberg et al., 1993; Simon et al., 1999). Notably, many imprinted loci have both maternally and paternally imprinted genes at the same loci, making an absolute link between early replication and expression difficult to establish.

The clearest evidence for a direct link between monoallelic expression and asynchronous replication timing comes from the study of the monoallelically expressed and asynchronously replicating light and heavy chain immunoglobulin genes (Mostoslavsky et al., 2001). These studies took advantage of a two-color FISH assay that detected rearrangement of immunoglobulin kappa (IGK) in one color and replication of the constant region in another. In mouse B-cells with one IGK rearrangement, the late replicating allele of IGK corresponded to the unrearranged locus in 83% of cells examined. Similar results were observed for the heavy chain immunoglobulin mu. Asynchronous replication of IGK was observed in transgenic lines in which neither endogenous allele rearranged or in which both alleles were rearranged. These results were interpreted as indication that the relative replication timing of each allele is an early epigenetic mark with influence over which allele is chosen for rearrangement later in development. However, without a better understanding of the causes of asynchronous replication, in particular its relationship to the presence of repetitive elements, a less direct connection between replication timing and allelic choice should not be ruled out.

Asynchronous replication and monoallelic expression likely represent different manifestations of the same set of epigenetic marks, established early in development. Each gene family probably responds to these epigenetic marks in different ways. For instance, in some cases, like immunoglobulins, the connection between replication timing

and expression may be stronger than others. Indeed, while all monoallelically expressed genes identified to date are asynchronously replicated, the inverse is not true. Many genes are asynchronously replicated because they are located near monoallelically expressed genes. Even genes that are monoallelically expressed in some tissues (such as Tlr-4 and p120-catenin) are biallelically expressed in other tissues (Gimelbrant et al., 2005). This suggests that while asynchronous replication is a requirement for monoallelic expression, other variables ultimately contribute to whether a gene is monoallelic. Perhaps different promoters are more or less susceptible to the epigenetic marks underlying asynchronous replication. Tissue specific enhancers may be responsible for why expression is affected by these marks in some cell types and not in others.

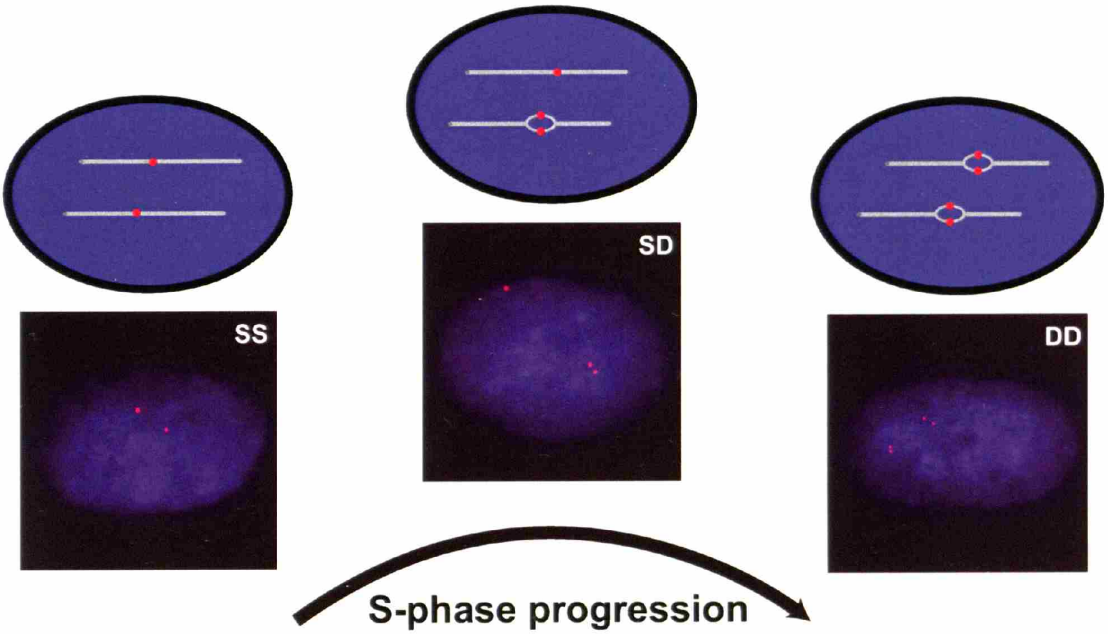
Asynchronous replication can be measured by a number of methods, the most widely used being fluorescent *in situ* hybridization (FISH) (Selig et al., 1992). FISH analysis of replication timing begins by pulse-labeling unsynchronized cells with the nucleotide analog BrdU. S-phase nuclei can be identified by immunofluorescent techniques due to being BrdU-positive. Using probes to specific DNA sequences, the number of copies of a particular locus can be examined in each S-phase nucleus (Figure 2A). Some nuclei display two single hybridization signals. This single-single (SS) pattern indicates that neither allele has replicated in that particular cell. A second class of nuclei displays two double dots, which indicates that both alleles have replicated by that point in S-phase (a DD pattern). Another FISH pattern is found in cells with one single dot and one double dot (a SD pattern). This third class of nuclei represents cells in which one allele has replicated (the doublet) whereas the other allele has not (the singlet).

Figure 2.

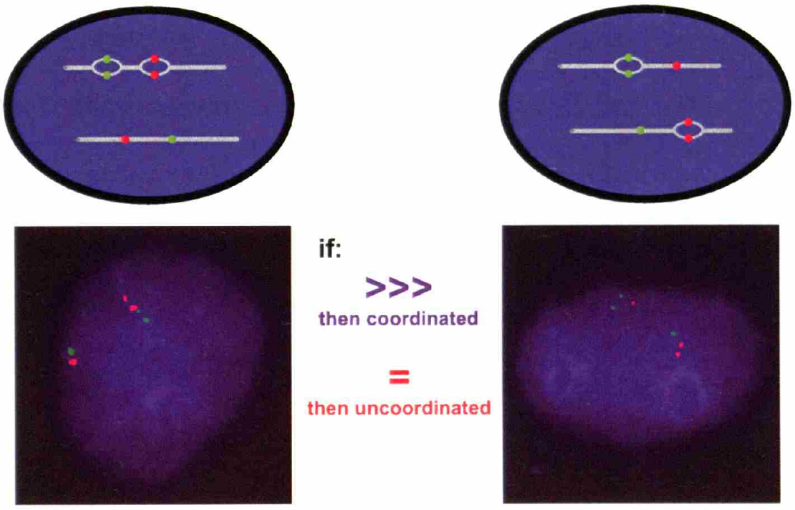
The FISH based assay of asynchronous replication. **(A)** Loci can be identified as replicating either synchronously or asynchronously using a fluorescence *in situ* hybridization (FISH) assay of replication timing. FISH analysis of interphase nuclei pulse-labeled with BrdU allows selective examination of cells in S-phase. Using a probe to a particular chromosomal site, some cells display two single hybridization dots indicating that neither allele has replicated (an SS pattern) while cells of a second class display two double dots indicating that both alleles have replicated (a DD pattern). A third class has cells with one single dot and one double dot indicating replication of only one of the two alleles (an SD pattern). Asynchronously replicating genes show the SD pattern in S-phase cells 30-40% of the time; this is higher than what is observed for synchronously replicating genes, which typically present this pattern in only 10-20% of S-phase cells. **(B)** The level of coordination of two distant genes on a particular chromosome can be examined by using two-color FISH analysis and scoring cells which display a single dot-double dot (SD) signal for both genes. This type of pattern can be found if the two genes replicate in an overlapping portion of S-phase. If the two genes are coordinated, the double dots for both genes should reside on the same chromosome (either maternal or paternal) and thus will be near each other in the nucleus. If the two genes are not coordinated, then the double dots for both genes should be on the same chromosome only 50% of the time.

FIGURE 2

A.


SS
SD
DD
S-phase progression

B.


if:
>>>
then coordinated
=
then uncoordinated

Asynchronously replicating loci show the intermediate, SD pattern in roughly 30-40% of cells in S-phase. Synchronously replicating loci, on the other hand, show a consistently lower percentage of SD cells (10-20%). The FISH assay has enabled the classification of a number of genes as either asynchronously replicating or synchronously replicating. Other, more cumbersome assays such as PCR-based measurements of replication timing have consistently agreed with the more flexible FISH assay (Gimelbrant et al., 2005; Gribnau et al., 2003; Mostoslavsky et al., 2001; Singh et al., 2003).

Prevalence of asynchronous replication and monoallelic expression

While there is continued uncertainty at to the number of asynchronously replicating loci in the genome, the number of odorant receptors and other known monoallelically expressed genes can be used for estimation purposes. Odorant receptors make up as much as 3-4% of the genes in the mammalian genome (Glusman et al., 2001; Young et al., 2002; Young and Trask, 2002). Therefore at least 3-4% of mammalian genes replicate asynchronously in S-phase.

While the exact size and delineation of replication domains are poorly characterized in mammals, it is estimated that mammalian replication domains span approximately 1 megabase in size and consist of several replicons that work in parallel (Edenberg and Huberman, 1975; Hand, 1978; Holmquist, 1987; Selig et al., 1992). If asynchronously replicating domains are this big, many loci may replicate asynchronously merely due to their proximity to odorant receptors and other known monoallelically expressed genes. Some of these genes, perhaps with weak promoters or enhancers, may be monoallelically expressed due to a position effect, reminiscent of the position effect

variegation first observed in *Drosophila melanogaster* (Schultz, 1950; Wakimoto, 1998). In position effect variegation, genes that are normally actively transcribed in euchromatic regions of the genome are silenced if they are placed near heterochromatic domains. Perhaps some of the monoallelically expressed genes located within asynchronously replicating clusters would be biallelically expressed if they were moved to other parts of the genome (either through transgenics, evolution, or transposition).

Coordination of asynchronous replication between disparate loci

Monoallelically expressed, asynchronously replicating genes are scattered throughout the genome, with other genes interspersed between them. The largest group of known asynchronously replicating genes, the odorant receptors, are located on most chromosomes, and exist in clusters varying in size from 1 gene to over 100 genes (Glusman et al., 2001). This dispersed genomic arrangement raises questions as to the regulation of these genes and the relationship between random asynchronous replication at different clusters. Random asynchronous replication represents an epigenetic readout of a random choice made early in development. Is that choice made at the level of individual loci, clusters, chromosomes, or is it genome-wide?

X-chromosome inactivation is the product of a chromosome-wide, random choice. Early evidence for the chromosome-wide nature of X inactivation came in the form of cytological observations of the highly condensed Barr body. Without such clear cytological differences between early and late replicating autosomal alleles, other approaches were needed to determine the nature of the random choice represented by autosomal asynchronous replication.

In the absence of gross cytological differences between asynchronously replicating alleles, modified FISH assays have proved extremely useful in the simultaneous examination of multiple loci within the same nucleus (Ensminger and Chess, 2004; Singh et al., 2003). Two-color FISH approaches, in which fluorophores with different emission wavelengths are used to label different probes, have provided great insights into the nature of asynchronous replication at autosomal loci.

Two-color FISH can be applied to determine whether random asynchronously replicating loci on the same chromosome represent a series of independent random choices or whether there is coordination between the random choice made at one locus and the choice made at others (Figure 2B). When two probes target linked segments of the same chromosome, the signals from those probes will be located near one another in the nucleus. So long as the distance between probe targets is not too great, signal from these linked segments will be closer to each other than they will be to their allelic counterparts. Signals representing DNA sequences with the same parent-of-origin can thus be distinguished from signals from the other homolog of that particular chromosome. This logic has allowed for the examination of a number of asynchronously replicating genes. By examining cells that display a singlet-doublet (SD) pattern for two different linked genes, one can determine whether the random asynchronous replication of those genes is coordinated.

When applied to X-linked loci, this application of two-color FISH confirms coordination between two disparate loci. Strikingly, just as disparate loci were coordinated on the X chromosome, coordination was also observed between loci on a number of autosomes in humans and mice (Ensminger and Chess, 2004; Singh et al.,

2003). Several autosomal asynchronously replicating genes were determined to be coordinated with each other, so long as they were located on the same chromosome. This coordination of asynchronous replication in mammalian autosomes provided evidence of a chromosome-wide process on the autosomes that was analogous to X-chromosome inactivation, at least with respect to replication timing.

It is important to remember that asynchronous replication only occurs in a fraction of the replication domains on any given autosome. Thus, asynchronously replicating domains can be thought of as islands of asynchronous replication in a sea of synchronously replicating sequence. Indeed, the observation of synchronous replication between coordinated asynchronous loci confirms that these loci reside in different replication domains. Insight into this may come by analogy to X inactivation, where a number of genes escape X inactivation (Carrel and Willard, 2005). Likewise, a large number of genes also escape autosomal asynchronous replication, though just as it is unclear why some genes escape X inactivation, it is not known why the majority of mammalian genes replicate synchronously in S-phase.

Another similarity between autosomal coordination and X-chromosome inactivation is the way that both systems respond to specific aneuploidy. As mentioned earlier, X inactivation has proven to be remarkably resilient with respect to the number of X chromosomes in a nucleus. Regardless of the number of X chromosomes in an otherwise diploid cell, there is always only one active X chromosome (Grumbach et al., 1963; Harden, 1961). This feature of X inactivation, known as the "N-1" rule, was also observed using the FISH assay of replication timing (Ensminger and Chess, 2004). This allowed the extension of analyses to autosomal trisomies, which showed that random

asynchronous replication of loci on that autosome behave in a similar way. Even with three copies, each of these loci has only one early replicating allele, with the other two replicating later in S-phase (Ensminger and Chess, 2004).

Summary of thesis

For my doctoral research, I have studied the epigenetic regulation of monoallelic expression in mammalian systems. A number of mammalian genes are expressed from only one of two alleles in either an imprinted or random manner. Those belonging to the random class include X-linked genes subject to X inactivation, as well as a number of autosomal genes, including odorant receptors, immunoglobulins, T-cell receptors, interleukins, natural killer-cell receptors, and pheromone receptors. The monoallelic expression of these autosomal genes is crucial to the establishment of cell-identity in many tissues. In order to dissect the processes involved in monoallelic expression, I have used a variety of biochemical, genomic, and cell biology-based approaches. This work began with the initial observation by our lab that suggested that in mice, chromosome-wide processes were shaping the regulation of these autosomal genes, analogous to what was known concerning X inactivation. I followed this work by focusing my analyses on human cells, which had a number of characteristics that allowed for the elucidation of further analogies between the autosomes and the X. Afterwards, I directed my focus on understanding the mechanisms behind these similarities.

The study of X-chromosome inactivation is a field with a rich history of many different approaches. Many X-linked genes are widely expressed, allowing for direct comparisons between the transcriptional state of one locus and that of the other.

However, members of the autosomal class of known monoallelically expressed genes present a problem when it comes to this sort of analysis, as they are often very restricted with regard to their expression patterns. For instance, any potential link between the choice of which allele of IGK is going to be expressed in B-cells and which allele of a particular odorant receptor is going to be expressed in a given olfactory neuron is masked by the fact that the two genes are never expressed in the same cell. Thus, we took advantage of another difference between the alleles of monoallelically expressed genes, their random asynchronous replication in S-phase.

Since randomly monoallelically expressed genes are scattered throughout the genome, we asked whether or not the choice made at one locus between two alleles influences the choice made at other loci. This was first examined in mice, where the FISH assay of replication timing showed that the random asynchronous replication of distributed autosomal genes is coordinated at the whole chromosome (but not genome) level.

In order to determine whether coordination of asynchronous replication extends to humans, I used the FISH-based assay of replication timing to demonstrate that the human homologs of a number of odorant receptors, interleukin genes, and the kappa immunoglobulin (IGK) all replicate asynchronously in human fibroblasts. Two-color FISH was utilized to show coordination occurred between genes on the same chromosome for 6 autosomes, using, for the first time, the X chromosome as a control. In addition to demonstrating that coordination is conserved between humans and mice, studying human cells allowed for the extension of earlier analyses in several important ways. First, mouse chromosomes are universally telocentric, which makes it impossible

to examine genes on opposite arms of the chromosome in mouse cells. However, human chromosomes have two arms of varying length, allowing for the demonstration that coordination can cross centromeres, thereby affecting genes on opposite arms of chromosomes. Also, the availability of several trisomic human cell lines facilitated the examination of what occurs to the choice of "one early allele, one late allele" when three copies of a particular gene are present. By modifying the standard FISH assay, I showed that autosomal trisomies tend to replicate one copy of an asynchronously replicating gene early and replicate the remaining two copies later in S-phase. These results were indistinguishable from what I observed in XXX trisomies, and are reminiscent of the "N-1 rule" of X inactivation. Such observations have significant implications regarding the way in which the choice of one early allele is chosen.

Due to similarities between X inactivation and autosome-pair nonequivalence, I explored the possibility that the early events in X inactivation may also occur on autosomes; specifically, that a *cis*-regulating RNA (like *Xist*) may be responsible for the creation of epigenetic inequality between the two homologs of each chromosome. *Xist* has a number of unique properties that might be utilized in the isolation of other *Xist*-like RNAs.

Our first attempt at such an isolation depended on a purely bioinformatics approach, scanning EST libraries for transcripts which lacked significant open reading frames. Unfortunately, due to the large number of introns and other noncoding sequences represented in these libraries, such an approach produced an unacceptably high number of false positives. I next sought to take advantage of the enrichment of *Xist* within the nucleus of cells. RNA purified from human nuclei was compared to cytoplasmic RNA

from the same cells using Affymetrix expression arrays. This approach correctly identified *Xist* ESTs as significantly enriched in the nuclear fractions. Strikingly, two other noncoding RNAs, located 60 kb apart on human chromosome 11 were identified as widely-expressed, nuclear-enriched transcripts. We have named these transcripts *hNEAT1* and *hNEAT2* (nuclear enriched autosomal transcripts). Both *NEAT1* and *NEAT2* are conserved within the mammalian lineage and are also enriched within the nuclei of mouse cells. I performed RNA FISH on these nuclear RNAs as well as *Xist* in primary human female fibroblasts. Whereas the *Xist* RNA signal is present as a single large focus covering the inactive X, these two autosomal RNAs appear to accumulate in dozens of small foci throughout the nucleus. The signal for these two RNAs, though qualitatively similar, is non-overlapping with respect to each other. In an attempt to elucidate the function of these noncoding RNAs, I have used RNA-mediated interference (RNAi) to knockdown the expression of these transcripts in HeLa cells. Expression profiling of these cells on Affymetrix arrays should indicate whether particular pathways or gene families are affected by the reduction of either *NEAT1* or *NEAT2*.

REFERENCES

Adams, J., and Hansche, P. E. (1974). Population studies in microorganisms. I. Evolution of diploidy in Saccharomyces cerevisiae. Genetics *76*, 327-338.

Aidoo, M., Terlouw, D. J., Kolczak, M. S., McElroy, P. D., ter Kuile, F. O., Kariuki, S., Nahlen, B. L., Lal, A. A., and Udhayakumar, V. (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality. Lancet *359*, 1311-1312.

Allison, A. C. (1964). Polymorphism and Natural Selection in Human Populations. Cold Spring Harb Symp Quant Biol *29*, 137-149.

Amrein, H., and Axel, R. (1997). Genes expressed in neurons of adult male Drosophila. Cell *88*, 459-469.

Angelov, D., Molla, A., Perche, P. Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Mol Cell *11*, 1033-1041.

Baker, B. S., and Ridge, K. A. (1980). Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. Genetics *94*, 383-423.

Barr, M. L., and Moore, K. L. (1957). Chromosomes, sex chromatin, and cancer. Proc Can Cancer Conf *2*, 3-16.

Belluscio, L., Koentges, G., Axel, R., and Dulac, C. (1999). A map of pheromone receptor activation in the mammalian brain. Cell *97*, 209-220.

Belote, J. M., and Lucchesi, J. C. (1980a). Control of X chromosome transcription by the maleless gene in Drosophila. Nature *285*, 573-575.

Belote, J. M., and Lucchesi, J. C. (1980b). Male-specific lethal mutations of Drosophila melanogaster. Genetics *96*, 165-186.

Belyaev, N., Keohane, A. M., and Turner, B. M. (1996). Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells. Hum Genet *97*, 573-578.

Bergman, Y., and Cedar, H. (2004). A stepwise epigenetic process controls immunoglobulin allelic exclusion. Nat Rev Immunol *4*, 753-761.

Bix, M., and Locksley, R. M. (1998). Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. Science *281*, 1352-1354.

Boggs, B. A., Connors, B., Sobel, R. E., Chinault, A. C., and Allis, C. D. (1996). Reduced levels of histone H3 acetylation on the inactive X chromosome in human females. Chromosoma *105*, 303-309.

Borst, P., and Ulbert, S. (2001). Control of VSG gene expression sites. Mol Biochem Parasitol *114*, 17-27.

Bridges, C. B. (1921). Triploid intersexes in *Drosophila melanogaster*. Science *54*, 252-254.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive

X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., and Willard, H. F. (1991a). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., and Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527-542.

Brown, C. J., Lafreniere, R. G., Powers, V. E., Sebastio, G., Ballabio, A., Pettigrew, A. L., Ledbetter, D. H., Levy, E., Craig, I. W., and Willard, H. F. (1991b). Localization of the X inactivation centre on the human X chromosome in Xq13. Nature *349*, 82-84.

Brown, C. J., and Willard, H. F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154-156.

Carrel, L., and Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400-404.

Chadwick, B. P., and Willard, H. F. (2003). Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Hum Mol Genet *12*, 2167-2178.

Chakalova, L., Carter, D., Debrand, E., Goyenechea, B., Horton, A., Miles, J., Osborne, C., and Fraser, P. (2005). Developmental regulation of the beta-globin gene locus. Prog Mol Subcell Biol *38*, 183-206.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. Curr Biol *6*, 149-162.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834.

Chien, Y. H., Gascoigne, N. R., Kavaler, J., Lee, N. E., and Davis, M. M. (1984). Somatic recombination in a murine T-cell receptor gene. Nature *309*, 322-326.

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol *132*, 259-275.

Costanzi, C., and Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Crowe, M. L., Perry, B. N., and Connerton, I. F. (1996). Olfactory receptor-encoding genes and pseudogenes are expressed in humans. Gene *169*, 247-249.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat Genet *22*, 323-324.

Davie, J. M., Paul, W. E., Mage, R. G., and Goldman, M. B. (1971). Membrane-associated immunoglobulin of rabbit peripheral blood lymphocytes: allelic exclusion at the b locus. Proc Natl Acad Sci U S A *68*, 430-434.

de la Cruz, C. C., Fang, J., Plath, K., Worringer, K. A., Nusinow, D. A., Zhang, Y., and Panning, B. (2005). Developmental regulation of Suz12 localization. Chromosoma *114*, 183-192.

Dhar, V., Skoultchi, A. I., and Schildkraut, C. L. (1989). Activation and repression of a beta-globin gene in cell hybrids is accompanied by a shift in its temporal replication. Mol Cell Biol *9*, 3524-3532.

Edenberg, H. J., and Huberman, J. A. (1975). Eukaryotic chromosome replication. Annu Rev Genet *9*, 245-284.

Efstratiadis, A. (1994). Parental imprinting of autosomal mammalian genes. Curr Opin Genet Dev *4*, 265-280.

Ensminger, A. W., and Chess, A. (2004). Coordinated replication timing of monoallelically expressed genes along human autosomes. Hum Mol Genet *13*, 651-658.

Epner, E., Forrester, W. C., and Groudine, M. (1988). Asynchronous DNA replication within the human beta-globin gene locus. Proc Natl Acad Sci U S A *85*, 8081-8085.

Erhardt, S., Su, I. H., Schneider, R., Barton, S., Bannister, A. J., Perez-Burgos, L., Jenuwein, T., Kouzarides, T., Tarakhovsky, A., and Surani, M. A. (2003). Consequences

of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. Development *130*, 4235-4248.

Franke, A., and Baker, B. S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell *4*, 117-122.

Gimelbrant, A. A., Ensminger, A. W., Qi, P., Zucker, J., and Chess, A. (2005). Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. J Biol Chem *280*, 1354-1359.

Glusman, G., Yanai, I., Rubin, I., and Lancet, D. (2001). The complete human olfactory subgenome. Genome Res *11*, 685-702.

Goren, A., and Cedar, H. (2003). Replicating by the clock. Nat Rev Mol Cell Biol *4*, 25-32.

Gorman, M., Kuroda, M. I., and Baker, B. S. (1993). Regulation of the sex-specific binding of the maleless dosage compensation protein to the male X chromosome in Drosophila. Cell *72*, 39-49.

Gribnau, J., Hochedlinger, K., Hata, K., Li, E., and Jaenisch, R. (2003). Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. Genes Dev *17*, 759-773.

Gribnau, J., Luikenhuis, S., Hochedlinger, K., Monkhorst, K., and Jaenisch, R. (2005). X chromosome choice occurs independently of asynchronous replication timing. J Cell Biol *168*, 365-373.

Grumbach, M. M., Morishima, A., and Taylor, J. H. (1963). Human Sex Chromosome Abnormalities In Relation to DNA Replication and Heterochromatinization. Proc Natl Acad Sci U S A *49*, 581-589.

Hall, L. L., Clemson, C. M., Byron, M., Wydner, K., and Lawrence, J. B. (2002). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum Mol Genet *11*, 3157-3165.

Hand, R. (1978). Eucaryotic DNA: organization of the genome for replication. Cell *15*, 317-325.

Hansen, R. S. (2003). X inactivation-specific methylation of LINE-1 elements by DNMT3B: implications for the Lyon repeat hypothesis. Hum Mol Genet *12*, 2559-2567.

Hansen, R. S., Stoger, R., Wijmenga, C., Stanek, A. M., Canfield, T. K., Luo, P., Matarazzo, M. R., D'Esposito, M., Feil, R., Gimelli, G., *et al.* (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum Mol Genet *9*, 2575-2587.

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M., and Gartler, S. M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A *96*, 14412-14417.

Harden, D. G. (1961). Nuclear Sex in Triploid XXY Human Cells. Lancet *2*, 488.

Hedrick, S. M., Cohen, D. I., Nielsen, E. A., and Davis, M. M. (1984a). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature *308*, 149-153.

Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I., and Davis, M. M. (1984b). Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature *308*, 153-158.

Held, W., Roland, J., and Raulet, D. H. (1995). Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. Nature *376*, 355-358.

Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998). Monoallelic expression of the interleukin-2 locus. Science *279*, 2118-2121.

Holmquist, G. P. (1987). Role of replication time in the control of tissue-specific gene expression. Am J Hum Genet *40*, 151-173.

Hong, Y. K., Ontiveros, S. D., and Strauss, W. M. (2000). A revision of the human XIST gene organization and structural comparison with mouse Xist. Mamm Genome *11*, 220-224.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc Natl Acad Sci U S A *73*, 3628-3632.

Ishii, T., Serizawa, S., Kohda, A., Nakatani, H., Shiroishi, T., Okumura, K., Iwakura, Y., Nagawa, F., Tsuboi, A., and Sakano, H. (2001). Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurones. Genes Cells *6*, 71-78.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jeppesen, P., and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*, 281-289.

Kelley, R. L., Meller, V. H., Gordadze, P. R., Roman, G., Davis, R. L., and Kuroda, M. I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell *98*, 513-522.

Keohane, A. M., O'Neill L, P., Belyaev, N. D., Lavender, J. S., and Turner, B. M. (1996). X-Inactivation and histone H4 acetylation in embryonic stem cells. Dev Biol *180*, 618-630.

Kerem, B. S., Goitein, R., Richler, C., Marcus, M., and Cedar, H. (1983). In situ nick-translation distinguishes between active and inactive X chromosomes. Nature *304*, 88-90.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., Nicholls, R. D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. Nature *364*, 459-463.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol *2*, E171.

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999a). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet *21*, 400-404.

Lee, J. T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275-279.

Lee, J. T., Lu, N., and Han, Y. (1999b). Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain. Proc Natl Acad Sci U S A *96*, 3836-3841.

Litwin, S. D. (1972). Allelic and class exclusion of membrane-associated immunoglobulins of human lymphocytes. J Immunol *108*, 1129-1131.

Lucchesi, J. C., Kelly, W. G., and Panning, B. (2005). Chromatin Remodeling in Dosage Compensation. Annu Rev Genet.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

Lyon, M. F. (1986). X chromosomes and dosage compensation. Nature *320*, 313.

Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet *80*, 133-137.

Madl, J. E., and Herman, R. K. (1979). Polyploids and sex determination in Caenorhabditis elegans. Genetics *93*, 393-402.

Meller, V. H., Gordadze, P. R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R. L., and Kuroda, M. I. (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. Curr Biol *10*, 136-143.

Meyer, B. J., and Casson, L. P. (1986). Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell *47*, 871-881.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221-225.

Navarro, M., and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. Nature *414*, 759-763.

Nef, P., Hermans-Borgmeyer, I., Artieres-Pin, H., Beasley, L., Dionne, V. E., and Heinemann, S. F. (1992). Spatial pattern of receptor expression in the olfactory epithelium. Proc Natl Acad Sci U S A *89*, 8948-8952.

Niimura, Y., and Nei, M. (2003). Evolution of olfactory receptor genes in the human genome. Proc Natl Acad Sci U S A *100*, 12235-12240.

Niimura, Y., and Nei, M. (2005). Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. Gene *346*, 23-28.

Panning, B., Dausman, J., and Jaenisch, R. (1997). X chromosome inactivation is mediated by Xist RNA stabilization. Cell *90*, 907-916.

Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I., and Meller, V. H. (2002). Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. Science *298*, 1620-1623.

Parkhurst, S. M., and Meneely, P. M. (1994). Sex determination and dosage compensation: lessons from flies and worms. Science *264*, 924-932.

Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. J Exp Med *122*, 853-876.

Pier, G. B., Grout, M., Zaidi, T., Meluleni, G., Mueschenborn, S. S., Banting, G., Ratcliff, R., Evans, M. J., and Colledge, W. H. (1998). Salmonella typhi uses CFTR to enter intestinal epithelial cells. Nature *393*, 79-82.

Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Priest, J. H., Heady, J. E., and Priest, R. E. (1967). Delayed onset of replication of human X chromosomes. J Cell Biol *35*, 483-487.

Ressler, K. J., Sullivan, S. L., and Buck, L. B. (1993). A zonal organization of odorant receptor gene expression in the olfactory epithelium. Cell *73*, 597-609.

Rhoades, K. L., Singh, N., Simon, I., Glidden, B., Cedar, H., and Chess, A. (2000). Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. Curr Biol *10*, 789-792.

Richards, E. J., and Elgin, S. C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.

Rodriguez, I., Feinstein, P., and Mombaerts, P. (1999). Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. Cell *97*, 199-208.

Roehm, N. W., Carbone, A., Kushnir, E., Taylor, B. A., Riblet, R. J., Marrack, P., and Kappler, J. W. (1985). The major histocompatibility complex-restricted antigen receptor on T cells: the genetics of expression of an allotype. J Immunol *135*, 2176-2182.

Rusconi, S., and Kohler, G. (1985). Transmission and expression of a specific pair of rearranged immunoglobulin mu and kappa genes in a transgenic mouse line. Nature *314*, 330-334.

Saleh, A., Davies, G. E., Pascal, V., Wright, P. W., Hodge, D. L., Cho, E. H., Lockett, S. J., Abshari, M., and Anderson, S. K. (2004). Identification of probabilistic transcriptional switches in the Ly49 gene cluster: a eukaryotic mechanism for selective gene activation. Immunity *21*, 55-66.

Schmidt, M., and Migeon, B. R. (1990). Asynchronous replication of homologous loci on human active and inactive X chromosomes. Proc Natl Acad Sci U S A *87*, 3685-3689.

Schubeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., and Groudine, M. (2002). Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. Nat Genet *32*, 438-442.

Schultz, J. (1950). Interrelations of factors affecting heterochromatin-induced variegation in *Drosophila*. Genetics *35*, 134.

Selig, S., Okumura, K., Ward, D. C., and Cedar, H. (1992). Delineation of DNA replication time zones by fluorescence in situ hybridization. Embo J *11*, 1217-1225.

Serizawa, S., Ishii, T., Nakatani, H., Tsuboi, A., Nagawa, F., Asano, M., Sudo, K., Sakagami, J., Sakano, H., Ijiri, T., *et al.* (2000). Mutually exclusive expression of odorant receptor transgenes. Nat Neurosci *3*, 687-693.

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., and Sakano, H. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. Science *302*, 2088-2094.

Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E., Formstone, E. J., Arkell, R. M., Nesterova, T. B., Alghisi, G. C., Rastan, S., and Brockdorff, N. (1997). Stabilization of Xist RNA mediates initiation of X chromosome inactivation. Cell *91*, 99-107.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., Jenuwein, T., Otte, A. P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell *4*, 481-495.

Simon, I., Tenzen, T., Mostoslavsky, R., Fibach, E., Lande, L., Milot, E., Gribnau, J., Grosveld, F., Fraser, P., and Cedar, H. (2001). Developmental regulation of DNA replication timing at the human beta globin locus. Embo J *20*, 6150-6157.

Simon, I., Tenzen, T., Reubinoff, B. E., Hillman, D., McCarrey, J. R., and Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. Nature *401*, 929-932.

Singh, N., Ebrahimi, F. A., Gimelbrant, A. A., Ensminger, A. W., Tackett, M. R., Qi, P., Gribnau, J., and Chess, A. (2003). Coordination of the random asynchronous replication of autosomal loci. Nat Genet *33*, 339-341.

Takagi, N. (1974). Differentiation of X chromosomes in early female mouse embryos. Exp Cell Res *86*, 127-135.

Turner, B. M. (1998). Histone acetylation as an epigenetic determinant of long-term transcriptional competence. Cell Mol Life Sci *54*, 21-31.

van de Vosse, E., Ali, S., Visser, A. W., Surjadi, C., Widjaja, S., Vollaard, A. M., and Dissel, J. T. (2005). Susceptibility to typhoid fever is associated with a polymorphism in the cystic fibrosis transmembrane conductance regulator (CFTR). Hum Genet, 1-3.

Vassar, R., Ngai, J., and Axel, R. (1993). Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. Cell *74*, 309-318.

Wakimoto, B. T. (1998). Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. Cell *93*, 321-324.

Wolf, B., Janeway, C. A., Jr., Coombs, R. R., Catty, D., Gell, P. G., and Kelus, A. S. (1971). Immunoglobulin determinants on the lymphocytes of normal rabbits. 3. As4 and As6 determinants on individual lymphocytes and the concept of allelic exclusion. Immunology *20*, 931-944.

Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T., and Migeon, B. R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci U S A *81*, 2806-2810.

Wood, W. B., Streit, A., and Li, W. (1997). Dosage compensation: X-repress yourself. Curr Biol *7*, R227-230.

Woodfine, K., Fiegler, H., Beare, D. M., Collins, J. E., McCann, O. T., Young, B. D., Debernardi, S., Mott, R., Dunham, I., and Carter, N. P. (2004). Replication timing of the human genome. Hum Mol Genet *13*, 191-202.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L., and Trask, B. J. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. Hum Mol Genet *11*, 535-546.

Young, J. M., and Trask, B. J. (2002). The sense of smell: genomics of vertebrate odorant receptors. Hum Mol Genet *11*, 1153-1160.

Zhang, J., Xu, F., Hashimshony, T., Keshet, I., and Cedar, H. (2002). Establishment of transcriptional competence in early and late S phase. Nature *420*, 198-202.

Chapter Two

Coordinated Replication Timing of Monoallelically Expressed Genes Along Human Autosomes

Alexander W. Ensminger and Andrew Chess

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142
Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

This chapter was published in Human Molecular Genetics 13(6): 651-658 (2004).

Alexander W. Ensminger performed all of the experiments leading to the publication of this chapter.

ABSTRACT

A number of genes in the mammalian genome are expressed from only one of two alleles in either an imprinted or random manner. Those belonging to the random class include X-linked genes subject to X inactivation, as well as a number of autosomal genes, including odorant receptors, immunoglobulins, T-cell receptors, interleukins, natural killer-cell receptors, and pheromone receptors. Monoallelically expressed genes display the unusual property of asynchronous replication and for those genes whose transcription is randomly monoallelic, the asynchronous replication is also random. In mice, recent work has shown that the random asynchronous replication of distributed autosomal genes is coordinated at the whole chromosome level, indicative of chromosome-pair nonequivalence. Here, we show that autosome-pair nonequivalence is present in human cells, and demonstrate its ability to cross the centromere. Additionally, by examining the replication of these genes in a number of human trisomies, we consistently find one allele replicating early and the other two alleles replicating late, similar to previous observations in X trisomies.

INTRODUCTION

The differential treatment of two sequence-identical alleles is a hallmark of random X inactivation. Established at the time of implantation, X inactivation represents a random choice made at the whole chromosome level, with half of the cells inactivating the paternal X and half the cells inactivating the maternal X (Lyon, 1961). X inactivation extends its influence across the centromere. In addition, X inactivation exhibits the so-called "n-1 rule," as a single X chromosome is chosen to be active regardless of the number of other copies present (Harden, 1961). One of the earliest observable differences between the two X chromosomes is a difference in their respective replication timing (Takagi, 1974). While most human genes are biallelically transcribed and have both alleles replicated synchronously during a specific portion of S-phase (Goren and Cedar, 2003), genes on the inactive X replicate later in S-phase than their active counterparts. This asynchronous replication, initially thought of as unique to X-linked genes, has since emerged as a property shared by all monoallelically expressed genes (Bix and Locksley, 1998; Chess et al., 1994; Kitsberg et al., 1993; Singh et al., 2003). While asynchronous replication and monoallelic expression likely represent different manifestations of a shared epigenetic mark, one important feature of asynchronous replication is that it can be observed in all cell types. For instance, odorant receptors are only expressed in specific, post-mitotic olfactory neurons, yet the DNA encoding these genes replicates asynchronously in fibroblasts, lymphoblasts, and all other cell types examined. Similarly, genes like the X-linked opsins undergo X inactivation early in development such that the inactive allele replicates later in all differentiated cell types.

The chromosome-wide nature of X inactivation was first revealed by early cytological observations and was subsequently confirmed by molecular analyses. In mice, we have studied a number of autosomal loci and shown that their asynchronous replication is also coordinated, rendering the alleles of all the randomly monoallelically expressed genes, scattered across the chosen chromosome, earlier replicating than the alleles on the homologous chromosome (Singh et al., 2003). Here, we have asked whether this process is conserved within the mammalian lineage and extend our analyses to examine autosome-pair nonequivalence in a number of human trisomies.

RESULTS

Asynchronous replication in humans

To test whether asynchronous replication is coordinated in humans, we first demonstrated that the human homologs of several monoallelically expressed mouse genes replicate asynchronously in human cells. Asynchronous replication is established early in development before tissue-specific transcription is established, thus making it possible to study the phenomenon in a number of cell types in which the genes are not expressed, including fibroblasts and lymphoblasts (Chess et al., 1994; Simon et al., 1999; Singh et al., 2003). Loci can be identified as replicating either synchronously or asynchronously using a fluorescence *in situ* hybridization (FISH) assay of replication timing (Selig et al., 1992). FISH analysis of interphase nuclei pulse-labeled with BrdU allows selective examination of cells in S-phase. Using a probe to a particular chromosomal site, some cells display two single hybridization dots indicating that neither allele has replicated (an SS pattern) while cells of a second class display two double dots indicating that both alleles have replicated (a DD pattern). A third class has cells with one single dot and one double dot indicating replication of only one of the two alleles (an SD pattern). Asynchronously replicating genes show the SD pattern in S-phase cells 30-40% of the time; this is higher than what is observed for synchronously replicating genes, which typically present this pattern in only 10-20% of S-phase cells.

The FISH assay we use is an accurate indicator of asynchronous replication; it has been corroborated by direct measurements of asynchronous replication using a number of S-phase fractionation methods (Gribnau et al., 2003; Mostoslavsky et al., 2001). Recently, we confirmed that this was the case for odorant receptor genes (Singh et al.,

2003). Note that while the FISH assay we use detects asynchronous replication, the fraction of cells with a visible doublet signal for a given allele may be influenced by differences in sister chromatid cohesion, especially when different FISH protocols are used (Azuara et al., 2003). These different protocols, known as 3D-FISH, utilize substantially different cell-fixation and denaturation conditions in order to visualize cohesion. By contrast, the FISH conditions we use are designed to minimize the detection of differences in sister chromatid cohesion (see Discussion). Irrespective of the relative contributions of replication asynchrony and sister chromatid cohesion to the doublet FISH signal, it represents an interesting epigenetic mark that distinguishes between the two alleles of monoallelically expressed genes.

Using the FISH assay, we confirmed our expectation that a number of odorant receptor genes, interleukin genes, and the kappa immunoglobulin gene (*IGK*), as well as two X-linked genes, all replicate asynchronously in human cells (Table 1, Fig. 1a). In order to determine whether the asynchronous replication we observed for human odorant receptors was random, we obtained a nonclonal cell line that is heterozygous for a specific deletion on the same arm of chromosome 2 as the odorant receptor, *OR6B3*. This deletion served as a mark for one of the two alleles, as we performed two-color FISH using a BAC mapped within this deletion labeled with FluorX (green) in concert with a Cy3-labeled (red) *OR6B3* probe. We scored 35 cells which displayed a single-double FISH pattern, and observed 17 nuclei in which the deletion was linked to the early allele, and 18 nuclei in which the early allele of *OR6B3* resided on the intact copy of chromosome 2 (Fig. 1b). These results confirmed our expectation that the asynchronous replication of human odorant receptors is random.

Figure 1

Examining asynchronous replication in human cells. (A) The set of genes analyzed in this study. Asynchronously replicating genes are represented in red, and synchronously replicating genes are represented in black. Synchronously replicating genes (black) located in between asynchronous loci demonstrate that the flanking asynchronous genes reside in different replication domains. The location of the BAC, RP11-504L12, is also given. (B) A probe for RP11-504L12, (green) was used to mark one of the two copies of chromosome 2 in lymphoblasts with a chromosome 2 deletion (pter>q34::q36>qter). By analyzing these cells with two-color FISH, we examined whether the asynchronous replication of the nearby *OR6B3* (red) odorant receptor was random. Two neighboring cells have replicated a different allele early. In 35 cells counted, 18 nuclei replicated the intact chromosome 2 early (as did the left cell shown here), whereas 17 first replicated the allele linked to the deletion (as did the right cell), indicating that the asynchronous replication of this odorant receptor is random.

FIGURE 1

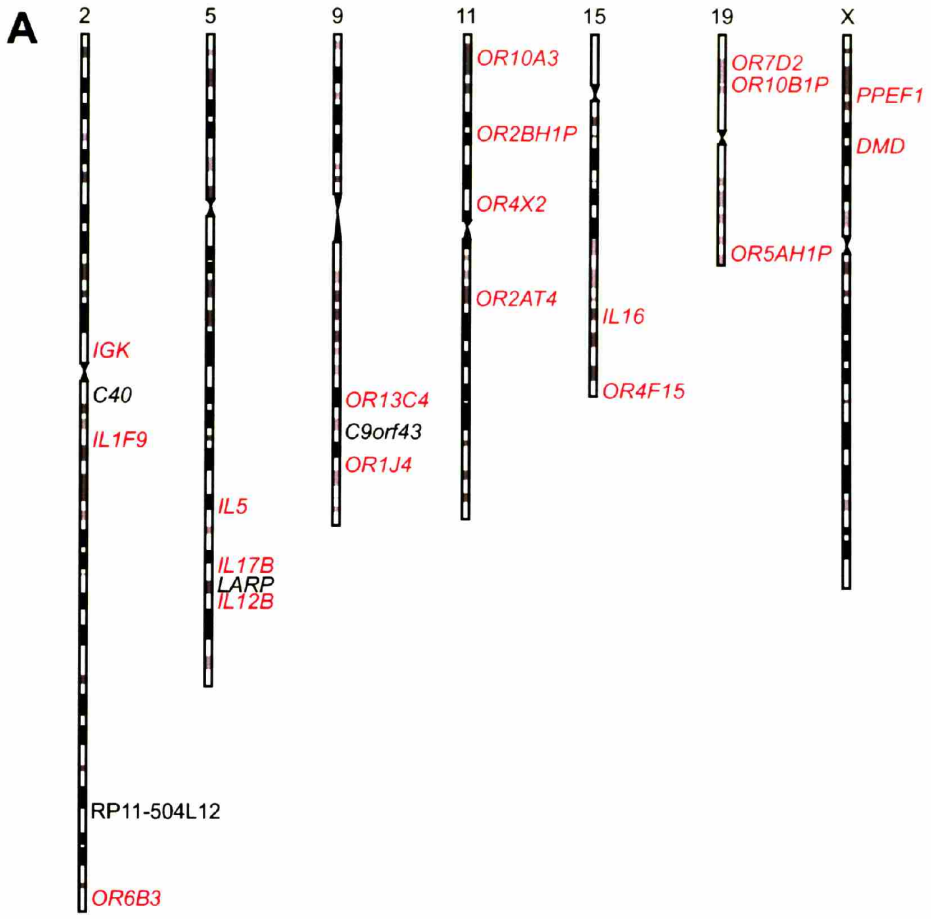

A
2
5
9
11
15
19
X
IGK
C40
IL1F9
RP11-504L12
OR6B3
IL5
IL17B
LARP
IL12B
OR13C4
C9orf43
OR1J4
OR10A3
OR2BH1P
OR4X2
OR2AT4
IL16
OR4F15
OR7D2
OR10B1P
OR5AH1P
PPEF1
DMD
B

Table 1. FISH analysis of a number of human genes.

Probe	SD (%)
IGK	42
IL1F9	41
IL5	42
IL12B	37
IL16	42
IL17B	33
OR2BH1P	36
OR1J4	45
OR2AT4	41
OR4F15	43
OR4X2	41
OR5AH1P	38
OR6B3	44
OR7D2	37
OR10A3	41
OR10B1P	43
OR13C4	37
PPEF1	42
DMD	47
C40	17
LARP	22
MGC17358	17

Loci can be identified as replicating either synchronously or asynchronously using a fluorescence *in situ* hybridization (FISH) assay of replication timing. In the assay, asynchronously replicating genes show the single-dot double-dot (SD) pattern in S-phase cells 30-40% of the time; this is higher than what is observed for synchronously replicating genes, which typically present this pattern in only 10-20% of S-phase cells.

Coordination of asynchronous replication in human cells

Asynchronously replicating genes are scattered throughout the human genome, with synchronously replicating genes, which comprise the bulk of genes, interspersed between them. The random choice of which allele to replicate early could be made at the level of the individual locus, or the individual chromosome (as for X-linked genes). We sought to determine the level of this choice for a number of human loci. The level of coordination of two distant genes on a particular chromosome was examined by using two-color FISH analysis and scoring cells which displayed a single dot-double dot (SD) signal for both genes (Singh et al., 2003). This type of pattern can be found if the two genes replicate in an overlapping portion of S-phase. If the two genes are coordinated, the double dots for both genes should reside on the same chromosome (either maternal or paternal) and thus will be near each other in the nucleus. If the two genes are not coordinated, then the double dots for both genes should be on the same chromosome only 50% of the time. Note that the two-color FISH assay depends on the physical proximity of two linked loci within the nucleus. When probes are greater than 50 Mb apart, the feasibility of the assay begins to diminish, as signal coming from the paternal allele of one gene may be closest to the maternal allele of the other gene. When possible, we examined the coordination of genes which were roughly 10-30 Mb apart. This distance ensures that the two genes are in different replication domains, but are close enough to display nearby FISH signals within each nucleus.

Using this two color approach, we analyzed chromosome-level coordination on four autosomes and the X chromosome for comparison. Two X-linked genes that are located 13.8 Mb apart, dystrophin (*DMD*) and a serine/threonine phosphatase, *PPEF1*,

reveal coordination (31 of 35 cells) (Fig. 2a). The fact that our assay did not show coordination in all 35 cells counted suggests that while the assay is robust ($p<0.001$ for a deviation from 50% in the above example), it does not allow visualization of the coordination in all cells examined. We next analyzed *OR13C4* and *OR1J4*, two odorant receptors located 17.8 Mb apart on chromosome 9, and observed evidence of coordination (30/34 cells, $p<0.001$) (Fig. 2b). Extending these analyses to other chromosomes, coordination was observed between *IL17B* and *IL12B*, located in two different interleukin clusters 9.9 Mb apart on chromosome 5 (30/36 cells, $p<0.001$) (Fig. 2c) as well as between an odorant receptor, *OR4F15*, and an interleukin, *IL16*, 21.5 Mb apart on chromosome 15 (28/35 cells, $p<0.001$) (Fig. 2d). Together with our earlier results in mice (Singh et al., 2003), these analyses suggest that chromosome-pair nonequivalence may be a general feature of mammalian chromosomes.

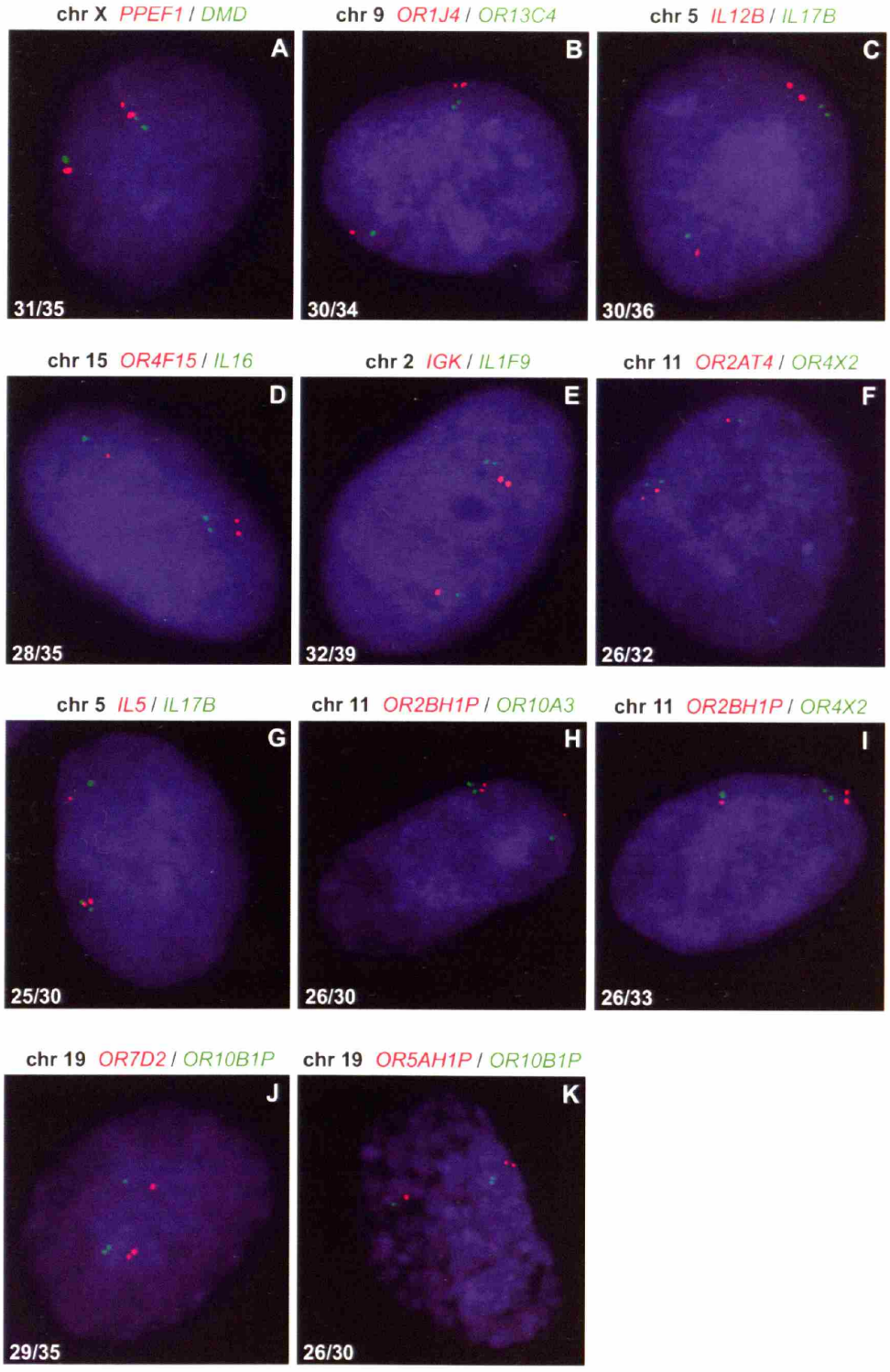
X inactivation is a chromosome-wide process. Not only do genes on both arms of the X chromosome replicate asynchronously, but the direction of this asynchrony is also fixed, such that in any given cell either all the X-linked genes replicate their maternal allele early or all replicate their paternal allele early. While the current mechanistic understanding of X inactivation helps explain why this is the case, this characteristic of X-linked genes can be observed without any assumptions as to how the process occurs. This led us to determine whether the effects of autosomal coordination can also be observed for genes on opposite sides of the centromeric boundary. If such were the case, one would expect the asynchronous replication of two genes on opposite arms of an autosome to be coordinated. While this question could not be asked in the mouse where all chromosomes are telocentric, human chromosomes have two arms of varying lengths.

Figure 2.

Coordination of asynchronous replication for individual human chromosome pairs. Two-color FISH analysis was performed on an apparently normal human 46, XX primary fibroblast population (WI-38). DAPI staining of nuclei (blue) is visible, and individual loci are visualized with 10 kb PCR products labeled with either Cy3 (red) or FluorX (green). The fraction of cells displaying the coordinated pattern are indicated for each probe in the bottom left of each panel. (A) Analysis of the X chromosome as a control. Two X-linked genes, dystrophin (*DMD*, green) and *PPEF1* (red). In 31/35 cells counted, the double-dot signals for each gene were on the same chromosome, indicating coordination of these two distant loci, consistent with uniform late replication of loci on the inactive X chromosome. (B) Two odorant receptors on chromosome 9, *OR13C4* (green) and *OR1J4* (red) show a similar coordinated pattern of replication (30/34 cells counted). (C) Two interleukins on chromosome 5, *IL17B* (green) and *IL12B* (red), (30/36 cells showed the coordinated pattern). (D) An OR and an interleukin on chromosome 15, *OR4F15* (red) and *IL16* (green), (28/35 cells showed the coordinated pattern). (E) *IGK* (red) and *IL1F9* (green), two asynchronous genes that reside on opposite arms of chromosome 2, represent the first demonstration that autosomal coordination can cross a centromere (32/39 cells counted). (F) Likewise, two odorant receptors on opposite arms of chromosome 11, *OR2AT4* (red) and *OR4X2* (green) are coordinated (26/32 cells). (G) In addition to *IL12B*, *IL17B* (green) is also coordinated with *IL5* (red) (25/30 coordinated) indicating that all 3 of these loci are coordinated. (H) On chromosome 11, *OR2BH1P* (red) and *OR10A3* are coordinated (26/30 cells). (I) *OR2BH1P* is also coordinated with *OR4X2* (green) (26/33 cells). (J) On chromosome 19, *OR7D2* is

coordinated with the cluster containing *OR10B1P* (29/35 cells). (K) Since *OR10B1P* (green) is also coordinated with *OR5AH1P* (red) (26/30 cells), all three of these loci, covering most of chromosome 19, are coordinated.

FIGURE 2


chr X PPEF1 / DMD
A
31/35
chr 9 OR1J4 / OR13C4
B
30/34
chr 5 IL12B / IL17B
C
30/36
chr 15 OR4F15 / IL16
D
28/35
chr 2 IGK / IL1F9
E
32/39
chr 11 OR2AT4 / OR4X2
F
26/32
chr 5 IL5 / IL17B
G
25/30
chr 11 OR2BH1P / OR10A3
H
26/30
chr 11 OR2BH1P / OR4X2
I
26/33
chr 19 OR7D2 / OR10B1P
J
29/35
chr 19 OR5AH1P / OR10B1P
K
26/30

Though several human chromosomes possess asynchronously replicating genes on both sides of the centromere, the requirements of our assay constrained our analysis to those loci which are less than 50 Mb apart, but on opposite arms of the chromosome. We identified chromosome 2 as being ideally suited for this analysis, as it contains two genes, *IGK* and the interleukin *IL1F9*, which are located on different arms of chromosome 2, yet only 22.2 Mb apart. The synchronously replicating gene, *C40*, resides between these two genes (Table 1, Fig. 1a), indicating that *IGK* and *IL1F9* are part of different replication domains, rather than belonging to one large domain spanning the centromere. Similar to the analyses with probes on the same side of a centromere, we observed coordination in 32/39 cells ($p<0.001$) (Fig. 2e), demonstrating that coordination extends beyond the centromere and most likely reflects a chromosome-wide choice whose underlying mechanisms are not impeded by centromeric structure. In addition to these two genes on chromosome 2, we also examined two odorant receptors on opposite arms of chromosome 11, *OR4X2* and *OR2AT4*, which are located 26.3 Mb apart. These loci were also coordinated (26/32 cells, $p<0.001$) (Fig. 2f) despite their location on opposite sides of the centromere. While in the case of X inactivation, the spreading of the *XIST* RNA across the centromere is thought to mediate the coordination of silencing and replication timing differences on the two arms, the mechanism allowing the coordination of autosomal genes on opposite sides of the centromere remains to be determined.

The observation of coordination between a number of linked pairs of autosomal genes suggested that the asynchronous replication of those genes was subject to chromosome-wide coordination. However, the possibility still remained that coordination in humans was not chromosome-wide, but rather existed in large subdomains of

chromosomes. In order to confirm that the coordination we observed between pairs of asynchronously replicating genes scattered on human autosomes was indeed chromosome-wide, we next sought to expand our observations to other genes located on the same chromosomes. This approach was based on the understanding that if gene A is coordinated with gene B, and gene B is coordinated with gene C, then gene A, by extension, is coordinated with gene C. We examined *IL5* and *IL17B* on chromosome 5 and observed evidence of coordination (25/30 cells, $p<0.001$) (Fig. 2g). Taken together with our observations of coordination between *IL12B* and *IL17B*, these results indicate that all three of these loci, residing over 26.8 Mb of chromosome 5, are coordinated with each other. Likewise, we extended our analysis of chromosome 11 to include the coordination of *OR10A3* and *OR2BH1P* (26/30 cells, $p<0.001$) (Fig. 2h). *OR2BH1P* and *OR4X2* are also coordinated (26/33 cells, $p<0.001$) (Fig. 2i). Since *OR4X2* and *OR2AT4* are also coordinated with one another, the asynchronous replication of all four of these loci on chromosome 11 is coordinated, covering most of 11p and 66.6 Mb in total. While the distribution of randomly asynchronously replicating genes on particular autosomes limits the contiguous expanses which can be examined by our assay, we identified three asynchronously replicating odorant receptor loci on chromosome 19 for analysis: *OR7D2*, *OR10B1P*, and *OR5AH1P*. These three loci on chromosome 19 extend to both arms and together cover 52.8 Mb or 83% of this chromosome. We observed coordination of *OR7D2* with *OR10B1P* (29/35 cells, $p<0.001$) (Fig. 2j) and *OR10B1P* with *OR5AH1P* (26/30 cells, $p<0.001$) (Fig. 2k). These results provide further support that random asynchronous replication is coordinated across entire chromosomes in humans.

Asynchronous replication in trisomies

Classical studies of X inactivation indicate that regardless of the number of X chromosomes present, in otherwise diploid cells there is always one active X chromosome with every other copy becoming inactivated ("the n-1 rule" of X inactivation) (Harden, 1961). With respect to replication timing, when there are more than two X chromosomes, one X replicates early in S-phase (the active X), while the remaining (inactive) X chromosomes replicate late. We used FISH to determine whether such a replication pattern could be observed for autosomal genes. The number of single dots in any given nucleus corresponds to the number of unreplicated alleles, whereas the number of double dots corresponds to the number of replicated alleles. Thus, if one allele replicates much earlier than the other two, a population of S-phase cells should contain more cells which exhibit a pattern with two single-dots and one double-dot (SSD) than those cells which exhibit one single-dot and two double-dots (SDD).

We studied trisomies of chromosomes 2, 9, 15, and X. Chromosome 21, while also commonly trisomic, does not provide a source of monoallelically expressed genes to use in these analyses. (All odorant receptor genes on chromosome 21 are products of recent duplications with copies present on multiple chromosomes, and no other monoallelically expressed genes are present on chromosome 21.) Since the number of X chromosomes inactivated in a cell is sensitive to the complement of other chromosomes present, we carefully selected trisomic cell lines whose only apparent aberration was the addition of a complete, extra copy of one chromosome. The first cell line we examined was a primary fibroblast line trisomic for chromosome 2 but otherwise diploid. The *IGK*

locus, located at 2p11.2, was analyzed, revealing four patterns of BrdU-positive cells (Fig. 3a, Table 2). Nuclei with either three single-dots (SSS) or three double-dots (DDD), respectively, represented cells in which either none or all of the alleles of *IGK* had replicated. In addition, two intermediate replication states, in which either only one (SSD) or two (SDD) alleles had replicated were also present within the population of cells progressing through S-phase. These cells were highly enriched for the SSD pattern, suggesting that in the trisomic state, one allele of *IGK* replicates early and two replicate late. These results were almost indistinguishable from our analysis of an X-linked gene, dystrophin (*DMD*) in an XXX cell line. As expected, the SSD pattern represented the predominant intermediate in S-phase, indicating that in these cells, one X chromosome replicated early and two alleles replicated late (Fig. 3b, Table 2). A control analysis of a synchronously replicating gene, *C40*, revealed that the two intermediate classes of cells (SSD and SDD) were equally represented in S-phase (Table 2). We next sought to determine how widespread this phenomenon was amongst the randomly asynchronously replicating autosomal genes.

When the analyses of the trisomy 2 cells used to study *IGK* were extended to an additional locus, an odorant receptor located at 2q37.3 (*OR6B3*), we again observed many more SSD than SDD cells, suggesting that this odorant receptor also replicates one allele early and two alleles late (Fig. 3c, Table 2). The SSD pattern also predominated for two chromosome 9 odorant receptor clusters (*OR1J4* and *OR13C4*) in trisomy 9 cells (Fig. 3d,e, Table 2), as well as an odorant receptor (*OR4F15*) in the context of a chromosome 15 trisomy (Fig. 3f, Table 2). Thus, when autosomal asynchronously replicating genes are present in the trisomic state, one allele replicates relatively early in

Figure 3

FISH analysis of trisomic human cell lines. (A-F) Examples of FISH analysis for 6 asynchronously replicating genes, in the context of 4 trisomic human cell lines. When three alleles of an asynchronously replicating gene are present, two intermediate points in S-phase are observable, the SSD pattern, in which one allele has replicated, and the SDD pattern, in which two (of the three) alleles have replicated. (A) The four cells shown represent the four types of cells observed examining the *IGK* gene (red) in a cell line trisomic for chromosome 2. *IGK* replicates one allele early and two late as evidenced by the large excess of SSD when compared with SDD nuclei. (B) The gene dystrophin (green), in an XXX cell line also replicates one allele early and two late, as expected based on the "n-1 rule" of X inactivation. (C) A similar pattern was observed for an odorant receptor, *OR6B3* on the distal end of chromosome 2. Likewise, one allele replicates early and two replicate late for two odorant receptors on chromosome 9, (D) *OR13C4* and (E) *OR1J4*, as well as an odorant receptor on chromosome 15, (F) *OR4F15*. (G) Two-color FISH shows coordination between *IGK* (red) and *IL1F9* (green) in trisomy 2 cells; in cells that were SSD for both genes, 17/22 show coordination, with the double-dot signal for each gene on the same chromosome. Thus, the choice of one early allele is a chromosome-wide decision.

FIGURE 3

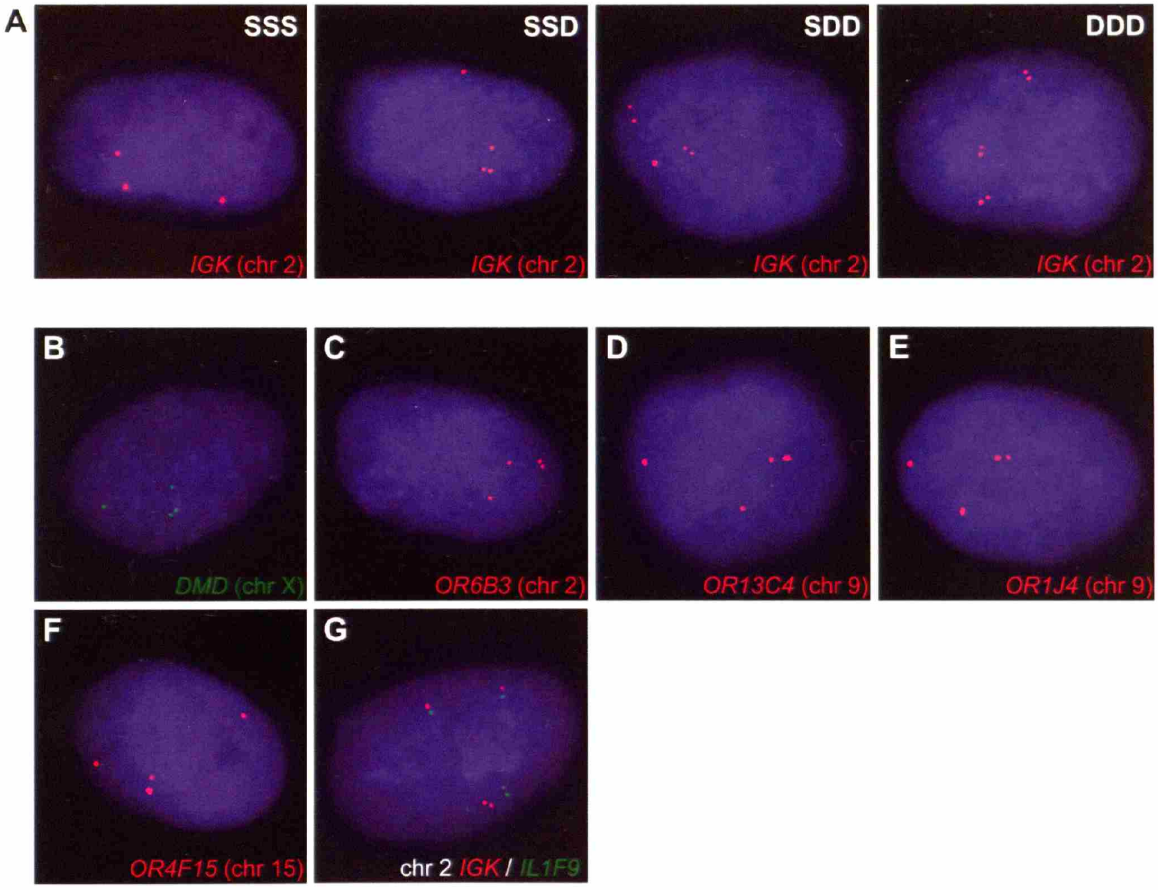

A
SSS
IGK (chr 2)
SSD
IGK (chr 2)
SDD
IGK (chr 2)
DDD
IGK (chr 2)
B
DMD (chr X)
C
OR6B3 (chr 2)
D
OR13C4 (chr 9)
E
OR1J4 (chr 9)
F
OR4F15 (chr 15)
G
chr 2 IGK / IL1F9

Table 2. An autosomal equivalent of the "n-1 rule" of X inactivation.

Probe (trisomy)	SSS (%)	SSD (%)	SDD (%)	DDD (%)
DMD (chr. X)	45	41	11	3
IGK (chr. 2)	36	42	9	13
OR6B3 (chr. 2)	48.5	36	9	6.5
OR13C4 (chr. 9)	44	38	9	9
OR1J4 (chr. 9)	44	38	10	8
OR4F15 (chr. 15)	37	37	10	16
C40 (chr. 2)	42	19	17	22

When three alleles of an asynchronously replicating gene are present, one allele replicates early in S-phase and two replicate late as evidenced by the large excess of SSD when compared with SDD nuclei. The X-linked loci *DMD* and a synchronously replicating gene, *C40*, were also examined for comparison. At least 100 BrdU-positive cells were counted for each percentage given.

S-phase, with the other two alleles replicating late, similar to what is observed for X-linked genes. Note that in the case of the X chromosome, these observations can be extended to XXXX, XXXXX, and XXXXXX individuals, which led to the formulation of the "n-1 rule," yet no analogous autosomal aneuploidies exist which would allow for such an extension to the autosomes.

To address the issue of chromosome-wide coordination in the trisomic state, we used two-color FISH to simultaneously explore the replication status of two discrete trisomic loci. Examining the *IGK* locus and *IL1F9* in the context of the previously described chromosome 2 trisomy, we asked if the same chromosome was early replicating for both genes. In 17/22 SSD cells examined ($p < 0.001$), *IGK* and *IL1F9* replicated the same allele first, suggesting that the choice of which allele replicates early is a well-regulated, coordinated choice. This suggests that the choice of only one early allele in the trisomic state is a chromosome-wide choice, with cells selecting one early chromosome such that asynchronously replicating genes on the remaining homologous chromosomes replicate later in S-phase.

DISCUSSION

Our previous studies of autosomes in mice have shown that random asynchronous replication is coordinated at the level of chromosomes, suggesting chromosome-pair nonequivalence for the autosomes. Genes subject to this coordination include the odorant receptors, immunoglobulins, T-cell receptors, pheromone receptors, and interleukins. The regulation of individual members of these families is critical in specifying the identity of the distinct cells within a tissue type. In this current investigation, we have provided evidence that random asynchronous replication is coordinated in human cells. In addition, by examining human chromosomes, we observe coordination between genes on opposite sides of the centromere, further supporting the idea of a chromosome-wide phenomenon. Additionally, based on the analysis of several different human trisomies, our data demonstrate that autosome-pair nonequivalence must provide a means by which only one copy of each chromosome replicates its asynchronously replicating genes early in S-phase. For the X chromosome, an unknown mechanism prevents stable *XIST* expression (and subsequent inactivation) on only one chromosome, regardless of the number of X chromosomes present.

In this investigation, we have used a fluorescence *in situ* hybridization assay to examine the random asynchronous replication of a number of human genes. In the FISH assay of replication timing, a single dot is interpreted as an unreplicated locus whereas a double dot is interpreted as DNA which has replicated. However, in order for a replicated segment of DNA to appear as a double dot in the nucleus, not only must the locus replicate, but the two pieces of DNA must also separate from one another sufficiently to give two FISH signals. For this reason, it has long been proposed that differences in sister

chromatid cohesion might affect the assay's ability to reliably measure replication timing, despite corroboration of the assay with direct measurements of DNA replication by this lab and others (Gribnau et al., 2003; Mostoslavsky et al., 2001; Singh et al., 2003).

Recent work by Azuara and colleagues has demonstrated that sister chromatid cohesion can be observed through fluorescence *in situ* hybridization methods, known as 3D-FISH (Azuara et al., 2003), however the conditions under which such detection is done should not be confused with the methods utilized here and by others in the field to measure replication timing. Specifically, under 3D-FISH, cells are subjected to fixation conditions (paraformaldehyde) designed to optimize the preservation of nuclear proteins and architecture. The preservation of nuclear structure, while typically a desirable aim, is likely to interfere with the measurement of replication timing due precisely to the architecture it maintains. Such an interpretation is supported by the observation by Azuara and colleagues that the FISH-based assay of replication timing not only gives different results than 3D-FISH, but it also more closely reflects results they obtained from direct measurements of replication timing (Azuara et al., 2003). Thus, the most precise measurements of DNA replication using fluorescence *in situ* hybridization are likely to be made under conditions in which the minimal amount of structure is maintained that might interfere with the separation of replicated sister chromatids.

The coordination of random asynchronous replication along human autosomes suggests the intriguing possibility that chromosome-pair nonequivalence, rather than being limited to X inactivation, could be a fundamental property of mammalian chromosomes. The autosomal genes affected by this phenomenon belong to a number of different gene families, each of which probably makes use of asynchronous replication

(or the underlying epigenetic mark it reflects) in the complex gene regulation that characterizes these families. For instance, in the case of the immunoglobulin genes, we have shown that the early replicating allele is preferentially rearranged (Mostoslavsky et al., 2001). While these genes all depend on monoallelic expression for their proper function, it is difficult to understand the reasons behind any sort of chromosome-wide process related to this expression. Indeed, many of the genes belonging to these families do not share overlapping patterns of expression. Similarly, many of the X-linked genes which are inactivated along one of two chromosomes in females are expressed in different tissues. The functional relevance of this inactivation is that only one of two alleles is expressed, not that the different genes subject to such inactivation are silenced in a chromosome-wide manner. The chromosome-wide nature of X inactivation can be regarded as a consequence of the mechanism behind mammalian dosage compensation rather than a requirement of the latter. Likewise, the chromosome-wide nature of autosomal "inactivation" may merely reflect the mechanisms utilized by this system to arrive at random monoallelic expression rather than some underlying requirement for chromosome-wide regulation.

It is interesting to consider whether X inactivation and autosome-pair nonequivalence might have arisen from a common ancestral process which rendered the two copies of a chromosome pair different from one another. Perhaps X inactivation is an adaptation of general chromosome-pair nonequivalence, with much more extensive events occurring further downstream in the development of an inactive X than in the creation of a coordinated autosome pair. Consistent with this notion is recent work which suggests that the nonrandom distribution of LINE-1 elements hypothesized to play a role

in X inactivation (Lyon, 1998) may also extend to monoallelically expressed genes on the autosomes (Allen et al., 2003). Despite an increasingly robust understanding of the manifestations of X-chromosome inactivation, many of its earliest events remain unknown. It is possible that X inactivation and autosome-pair nonequivalence take advantage of similar mechanisms to achieve similar ends; that a series of epigenetic modifications result in the differential treatment of two chromosomes which at one point in development were equal.

MATERIALS AND METHODS

Cell culture:

The primary human fibroblast cell line, WI-38 (American Type Culture Collection), was used to determine whether loci were either asynchronously replicating or synchronously replicating as well as in subsequent coordination analyses. For the trisomic studies, early passage, primary fibroblasts, GM04626 (47,XXX), GM10401 (47,XX,+2), GM09286 (47,XY,+9), and GM03184 (47,XY,+15) were purchased from the NIGMS Human Genetic Cell Repository (Corriell), as were lymphoblasts, GM10918, with a specific chromosome 2 deletion (46,XX,del(2)(pter>q34::q36>qter)) near *OR6B3*. Cells were maintained under standard conditions, fed 24 hours prior to harvest, and were pulse-labeled with BrdU for 35-45 minutes prior to fixation in 3:1 Methanol:Acetic Acid as previously described (Singh et al., 2003)

FISH:

FISH analysis was performed as previously described (Singh et al., 2003), with an adaptation of using large PCR products as probes. BACs were obtained from BAC PAC Resources and served as templates in the following 10 kb long-range PCRs: the constant region of *IGK* (RP11-344F17), *IL1F9* (RP11-261F13), *OR6B3* (RP11-98P19), *IL17B* (RP11-92I17), *IL12B* (RP11-117N12), *OR13C4* (RP11-317C20), *OR1J4* (RP11-345A24), *IL16* (RP11-35O20), *OR4F15* (RP11-259N2), *IL5* (RP11-17K19), *OR10A3* (RP11-1105A14), *OR2BH1P* (RP11-62M5), *OR4X2* (RP11-111N23), *OR2AT4* (RP11-158C6), *OR7D2* (RP11-1114G15), *OR10B1P* (RP11-1109J16), *OR5AH1P* (RP11-381F14), *PPEF1* (RP11-42E12), *DMD* (RP11-318G17), and *C9ORF43* (RP11-10I9).

Primers supplied by Integrated DNA Technologies (sequences available upon request) were designed to flank each gene's coding region (except the IL5 probe, which used a PCR product from 9 kb upstream of the interleukin's coding region), producing a 9000-11000 bp product using the Advantage 2 PCR system (BD Biosciences Clontech). Products were purified using the Wizard PCR Preps DNA Purification System (Promega). Aliquots of 5 μl (one-tenth of the PCR reaction) were direct-labeled with either Cy3-dCTP or FluorX-dCTP using a Nick Translation kit according to the manufacturer's instructions (Amersham Biosciences). Labeled probes were purified using G-50 Sephadex columns (Roche) and precipitated with 30 mg human cot-1 DNA and 70 mg salmon sperm DNA (Invitrogen), washed in 75% ethanol followed by 100% ethanol and were resuspended in 100 μl hybridization buffer (50% formamide, 10% dextran sulfate, 1X SSC). 10 μl of each probe was prehybridized (90°C for 5 min. followed by 10 minutes at 37°C) and then hybridized overnight with cells dropped on poly-L-lysine slides. Subsequent washes and antibody detection of BrdU were also as previously described. As mentioned in the results, under these conditions, the FISH assay has been shown to corroborate with direct measurements of asynchronous replication using a number of S-phase fractionation methods (Gribnau et al., 2003; Mostoslavsky et al., 2001) and recently we confirmed this was the case for odorant receptor genes in mice (Singh et al., 2003). In this study, we have used the FISH assay because the application of S-phase fractionation methods to the study of randomly asynchronously replicating genes in human cells is complicated by a number of factors. Unlike imprinted genes, the study of randomly asynchronously replicating genes requires the generation of a clonal population of cells, yet human EBV-transformed lymphoblasts are much more difficult to

subclone and subsequently maintain in culture than Abelson Murine Leukemia virus transformed mouse lymphocytes. Moreover, even S-phase fractionation studies of imprinted genes (whose study does no require the generation of clonal cell lines) have not been as robust as similar studies of mouse genes; this is probably due to differences in S-phase fractionation accuracy.

P-values:

P-values were calculated based on a binomial probability distribution.

ACKNOWLEDGEMENTS

The authors thank Itimar Simon for help with the FISH assay. We also thank Julie Claycomb, Paul Garrity, Alexander Gimelbrant, Guilherme Neves, David Page, and Michael Tackett for critical reading of the manuscript; George Bell for assistance with line art production; Howard Cedar, Mark Daly, and Jacob Zucker for insightful discussions; and members of the Chess lab for help and guidance. This work was supported by grants from the U.S. National Institutes of Health. A.W.E. is a Howard Hughes Medical Institute Predoctoral Fellow.

REFERENCES

Allen, E., Horvath, S., Tong, F., Kraft, P., Spiteri, E., Riggs, A. D., and Marahrens, Y. (2003). High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proc Natl Acad Sci U S A *100*, 9940-9945.

Azuara, V., Brown, K. E., Williams, R. R., Webb, N., Dillon, N., Festenstein, R., Buckle, V., Merkenschlager, M., and Fisher, A. G. (2003). Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication. Nat Cell Biol *5*, 668-674.

Bix, M., and Locksley, R. M. (1998). Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. Science *281*, 1352-1354.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834.

Goren, A., and Cedar, H. (2003). Replicating by the clock. Nat Rev Mol Cell Biol *4*, 25-32.

Gribnau, J., Hochedlinger, K., Hata, K., Li, E., and Jaenisch, R. (2003). Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. Genes Dev *17*, 759-773.

Harden, D. G. (1961). Nuclear Sex in Triploid XXY Human Cells. Lancet *2*, 488.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., Nicholls, R. D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. Nature *364*, 459-463.

Lyon, M. F. (1961). Gene Action in the X-chromosome of the Mouse (*Mus musculus L.*). Nature *190*, 372-373.

Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet *80*, 133-137.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221-225.

Selig, S., Okumura, K., Ward, D. C., and Cedar, H. (1992). Delineation of DNA replication time zones by fluorescence in situ hybridization. Embo J *11*, 1217-1225.

Simon, I., Tenzen, T., Reubinoff, B. E., Hillman, D., McCarrey, J. R., and Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. Nature *401*, 929-932.

Singh, N., Ebrahimi, F. A., Gimelbrant, A. A., Ensminger, A. W., Tackett, M. R., Qi, P., Gribnau, J., and Chess, A. (2003). Coordination of the random asynchronous replication of autosomal loci. Nat Genet *33*, 339-341.

Takagi, N. (1974). Differentiation of X chromosomes in early female mouse embryos. Exp Cell Res *86*, 127-135.

Chapter Three

A Search for *Xist*-like nuclear transcripts in the human genome identifies two conserved noncoding RNAs on chromosome 11.

Alexander W. Ensminger[1,2], John Hutchinson[2], Christine Ladd-Acosta[3], Morgan Trimble[1], Todd Golub[3], and Andrew Chess[2]

[1]Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
[2]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114
[3]Howard Hughes Medical Institute, Chevy Chase, Maryland; The Eli and Edythe L. Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, MA; and the Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA

Alexander Ensminger prepared all of the samples which were run on Affymetrix expression arrays by Christine Ladd-Acosta, characterized the nuclear enriched transcripts, analyzed their subnuclear localization using RNA FISH, did all of the quantitative real-time PCR experiments, the RFLP analysis of NEAT1 expression in clonal lymphoblasts, the siRNA transfection experiments, and, together with John Hutchinson and Andrew Chess, performed comparative analysis of the noncoding RNA loci between species.

ABSTRACT

Dosage compensation in mammals depends on the chromosome-wide silencing of one X chromosome in female (XX) cells. A chromosome-wide autosomal process was recently shown to share many similarities with X inactivation. X inactivation represents a random choice, which is propagated along the entire length of the chromosome through a noncoding RNA, *Xist*. *Xist* is a large spliced transcript, polyadenylated and tightly associated with the inactive X in the nucleus of mammalian females. As part of a search for *Xist*-like molecules, we used Affymetrix expression arrays to identify other noncoding RNAs enriched in the nucleus of human cells. This screen identified several probes for *Xist*, as well as two unique noncoding RNAs on chromosome 11. Genomic analysis of these transcripts indicates that they are conserved within the mammalian lineage. These two noncoding RNAs, *NEAT1 and NEAT2* (nuclear enriched autosomal transcripts), are described.

INTRODUCTION

Mammalian dosage compensation results in the chromosome-wide silencing of one copy of the X chromosome in female cells, known as X-chromosome inactivation (Lyon, 1961). As a result of this epigenetic phenomenon, X-linked genes in females are subject to asynchronous DNA replication and monoallelic expression (Lyon, 1961; Priest et al., 1967; Takagi, 1974). X inactivation, however, is not the only example of a chromosome-wide, epigenetic decision in the mammalian cell. We have recently discovered that the asynchronous replication of disparate autosomal loci is coordinated at the level of entire chromosomes, analogous to the asynchronous replication observed for X-linked genes in females (Ensminger and Chess, 2004; Singh et al., 2003). In addition, the asynchronous replication of X-linked genes and autosomal loci both follow an "N-1" rule in human trisomies, where one allele replicates early in S-phase and the other two replicate late (Ensminger and Chess, 2004).

Asynchronous replication is a hallmark of monoallelic expression in mammals (Chess et al., 1994; Kitsberg et al., 1993; Schmidt and Migeon, 1990; Selig et al., 1992). The majority of the human genome replicates synchronously, with both alleles replicating at the same point in S-phase. However, some mammalian loci replicate asynchronously in S-phase with one allele replicating earlier in S-phase than the other. Differences in allelic replication timing often coincide with differential allelic expression (Mostoslavsky et al., 2001; Selig et al., 1992). All known randomly monoallelically expressed genes are also subject to random asynchronous replication. This class of genes includes X-linked genes, immunoglobulins, interleukins, T-cell receptors, odorant receptors, the Toll-like 4 receptor, and p120-catenin (Bix and Locksley, 1998; Chess et al., 1994; Ensminger and

Chess, 2004; Gimelbrant et al., 2005; Mostoslavsky et al., 2001; Schmidt and Migeon, 1990; Singh et al., 2003). While these genes are interspersed across the mammalian genome, the random choice of which allele replicates early and which replicates late is coordinated across entire chromosomes (Ensminger and Chess, 2004; Singh et al., 2003).

The chromosome-wide nature of X inactivation is a direct result of the mechanisms underlying mammalian dosage compensation (Lucchesi et al., 2005). X inactivation is the result of a random choice, made at one central locus, which then propagates across the entire chromosome. The site of random choice is the X-inactivation center (*Xic*), which was defined genetically prior to any molecular understanding of its function (Brown et al., 1991b). Within the *Xic*, two noncoding RNAs, *Tsix* and *Xist* are transcribed from opposite strands at an overlapping locus (Brockdorff et al., 1992; Brown et al., 1991a; Brown et al., 1992; Lee et al., 1999). *Tsix* expression is antagonistic to *Xist* expression, and the interplay between the two transcripts results in monoallelic expression of both, with *Xist* expressed entirely from one allele and *Tsix* expressed from the other (Lee et al., 1999). The role of *Tsix* in human X inactivation is much less established than it is in mouse, with the existence of *Tsix* in humans unconfirmed (Chow et al., 2003; Migeon, 2003; Vasques et al., 2002). Regardless of the details surrounding the stabilization of *Xist* expression during differentiation, it is clear that the propagation of *Xist* transcript along one of the two X chromosome is crucial for the establishment of an inactive X (Lee and Jaenisch, 1997; Panning et al., 1997; Wutz and Jaenisch, 2000). Autosomal asynchronous replication also faces the challenge of creating a chromosome-wide epigenetic mark, yet its mechanism is not yet understood. We sought to determine if RNAs are also involved in

autosomal coordination and embarked on a search for nuclear-enriched, autosomal RNAs that might be candidates for having an *Xist*-like function.

Using a microarray-based approach, we identified 2 evolutionarily conserved noncoding RNA transcripts that, like *Xist*, are enriched in human nuclei. These nuclear enriched autosomal transcripts (NEATs), are located on the same arm of chromosome 11 in humans, less than 70 kb apart. Using a variety of approaches, we describe these two loci, named *NEAT1* and *NEAT2*, and characterize the likelihood that they function in a manner parallel to that of *Xist*.

RESULTS

An array-based approach to identify ubiquitously expressed nuclear RNAs

Xist is a large, spliced, noncoding RNA, which is polyadenylated and stably expressed in female somatic cells (Brockdorff et al., 1992; Brown et al., 1991a; Brown et al., 1992; Hong et al., 2000; Nesterova et al., 2001). In these cells, it is monoallelically expressed and is tightly associated with one of the two X chromosomes (Brockdorff et al., 1992; Brown et al., 1991a; Clemson et al., 1996). If noncoding transcripts also exist that regulate the autosomal process of random asynchronous replication, they may share some of these features. We decided to base our initial screen on one of these features: the localization of *Xist* within the cell (Brockdorff et al., 1992; Brown et al., 1992). We reasoned that the localization of *Xist* transcripts to the nuclei of somatic cells should differentiate it from most human genes. If other transcripts could be identified based on their localization to the nucleus, these could be examined with respect to many of the other features of *Xist*.

In order to examine a large number of human transcripts with respect to their subcellular localization, we utilized commercially available expression arrays from Affymetrix. The array set was designed to query over 30,000 unique human transcripts, based on the Unigene set of expressed sequence tags (ESTs). Typically, these arrays are used to probe for differences in expression profiles between unique populations of cells. Instead, these arrays were used to compare the level of each transcript in nuclear versus cytoplasmic samples. Two female cell types were examined in order to identify broadly expressed nuclear transcripts: the primary human fibroblast cell line, WI-38, and an EBV-transformed lymphoblastoid line, GM10852. These female cell lines had a number

of nuclear enriched transcripts identified by the arrays, including *Xist*. Nuclear enriched probes were aligned to the human genome and qualified using the University of California at Santa Cruz genome browser. Based on these analyses, the top 60 sequences enriched greater than 2-fold in the nuclei of both cell types included *Xist* (6 probes), introns (17), transcripts to protein coding genes (23), repeat elements (10), and two other noncoding RNAs represented twice each (4). Other sequences were enriched in the nuclei of lymphoblasts, or fibroblasts, but not both.

The presence of a number of intronic probes on the Affymetrix arrays was striking, suggesting that a large number of Unigene EST clusters are not unique transcripts but instead represent improperly annotated introns. Indeed, many of these intronic probes had corresponding polyA repeats located in downstream genomic sequence which likely facilitated the annealing of oligo-dT primers at these sites during reverse transcription reactions. Genes do sometimes reside within the introns of other transcripts (Ashburner et al., 1999; Loebel et al., 2005), however, of the over 20 intronic probes we examined, all indicated transcription from the same strand as the flanking pre-mRNA. As there is no reported bias for intra-intronic genes to be transcribed in the same direction as the overlapping transcript, these data suggest that the vast majority of these probes detect intronic RNA rather than novel genes.

Nuclear, non-intronic sequences aligned to either repeat elements, open reading frame (ORF) genes, or non-ORF (noncoding) RNA transcripts. Of the latter set, *Xist* probes were by far the most prevalent. Indeed, *Xist* is the most nuclear-enriched noncoding RNA identified by this screen, as the six probes to it are enriched 7.7 to 13-fold in the nucleus. Two noncoding autosomal transcripts are also enriched in the nuclei

of both cell lines. We named these two loci *NEAT1* and *NEAT2*, to reflect their status as nuclear enriched, autosomal transcripts. Two probes to the first transcript, *NEAT1*, are enriched 11 and 26-fold in nuclei of fibroblasts and 3 and 5.5-fold in lymphoblasts. Two different probes to *NEAT2* also show nuclear enrichment of 3.4 and 7.4-fold in fibroblasts and 2.3 and 2.4-fold enrichment in lymphoblasts. Neither transcript has an open reading frame (ORF) of significant size. The largest ORFs in *NEAT1* are 100 amino acids, 62 aa, and 95 aa. Of all open reading frames, the 62 aa stretch in frame 2 is the only one beginning with a methionine. The largest ORF in *NEAT2* is 50 aa long. While the nucleotide sequence of both transcripts is conserved within the mammalian lineage, this conservation does not extend to any of the open reading frames. Thus, like *Xist*, *NEAT1* and *NEAT2* appear to represent nuclear-enriched, noncoding RNAs present within the nuclei of human cells.

Two noncoding RNAs enriched in the nuclei of human cells

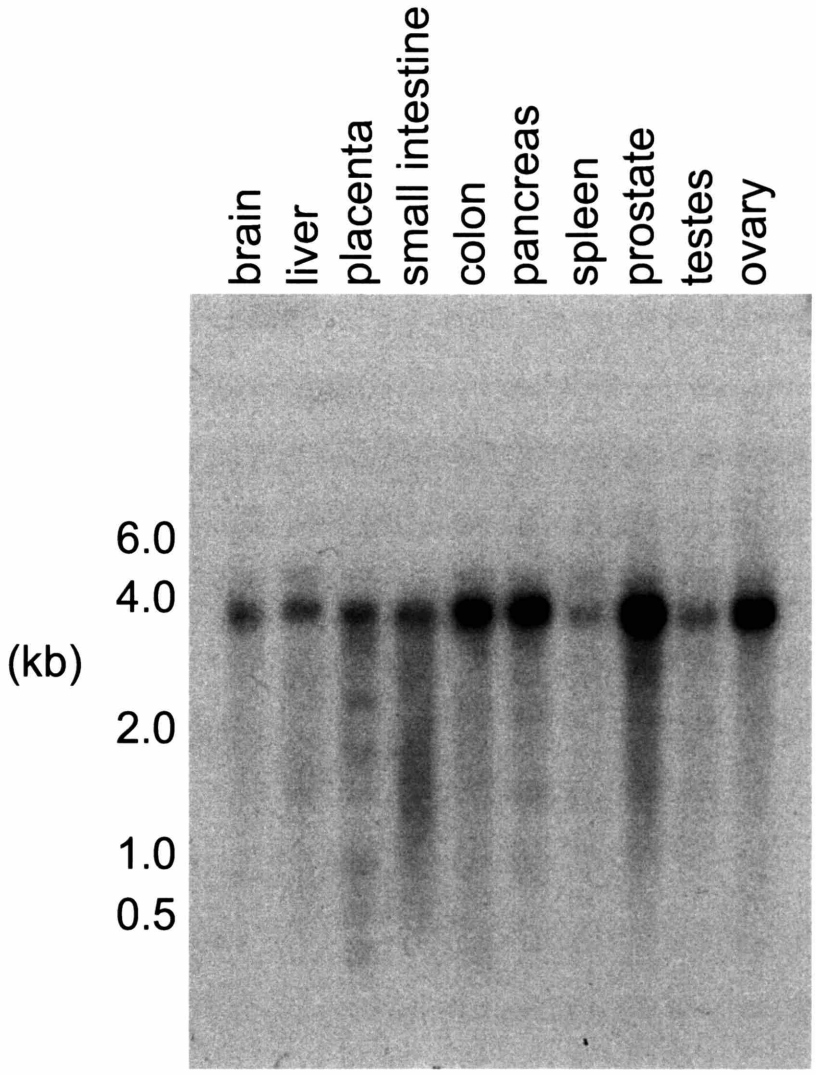
Genomic alignment, complimented by Northern blot and 5'-RACE analysis, defines the two NEAT transcripts as unique loci, separated by less than 60 kb on chromosome 11q13.1 of humans. The genomic proximity of these two transcripts is striking, given that the assembled human genome is roughly 3000 Mb (Lander et al., 2001). *NEAT1* is a large, intronless RNA (Guru et al., 1997). Using 5'-RACE, the transcriptional start of *NEAT1* maps to a discrete point. However, 3'-RACE failed to uncover specific ends of the transcript, due in part to a series of genomic polyA repeats throughout the transcript. Northern analysis suggests a transcript expressed in several tissues, roughly 4 kb in length (Figure 1). While this commercially obtained blot

Figure 1.

Analysis of *hNEAT1* expression in a variety of tissues.

Radiolabeled probe to *hNEAT1* was used to examine the *hNEAT1* transcript in a number of different human cell types. This analysis indicates the presence of a 4 kb transcript at the *hNEAT1* locus. The difficulty in detecting large transcripts on Northern blots may explain the absence of a larger (>17 kb) transcript that has been observed before at this locus (data not shown and Guru et al., 1997). *hNEAT1* is expressed in all cell types, but shows higher levels of expression in colon, pancreas, prostate, and ovary. Data from the Genomics Institute of Novartis Research Foundation (GNF), in which a much large number of human tissues were examined using Affymetrix expression arrays (Su et al., 2002; Su et al., 2004), indicates that probes to *hNEAT1* and *hNEAT2* are expressed at high levels in cells from the immune system ("http://symatlas.gnf.org/SymAtlas/" and data not shown). Discrepancies between Northern analysis and the GNF data set may be explained by the limitation of the Northern blot with respect to the larger transcript or other technical issues.

FIGURE 1


brain
liver
placenta
small intestine
colon
pancreas
spleen
prostate
testes
ovary
6.0
4.0
(kb)
2.0
1.0
0.5

provided the clearest evidence of a widely expressed 4 kb transcript, transcripts above 9 kb in length may not have been transferred efficiently to the membrane. A larger transcript (>17 kb) with an overlapping 5' end is also suggested by other, less precise Northern data (data not shown and Guru et al., 1997). The size of this transcript is more difficult to define, due likely to the difficulty of detecting large RNA transcripts on Northern blots. These results indicate the existence of at least two unique isoforms of *hNEAT1*: a 4 kb transcript and a much larger (> 17 kb) transcript. Both transcripts likely share a transcriptional start site, based on 5'-RACE analysis. *NEAT2* is a spliced transcript of more than 8 kb in length (Ji et al., 2003). Previously identified in a genome-wide screen for genes upregulated in pre-metastatic non-small-cell lung cancer, this transcript was given the name, Metastasis Associated in Lung Adenocarcinoma Transcript 1 (*MALAT-1*) (Ji et al., 2003; Muller-Tidow et al., 2004). Whether this transcript has any functional relevance to metastatic potential, however, awaits further analysis. Thus, we will continue to refer to it as *hNEAT2*, to reflect its enrichment in the nucleus of many different cell types.

RNA FISH of *Xist*, *NEAT1*, and *NEAT2*

The subcellular localization of *Xist* within the nuclei of mammalian cells provided striking clues to its role in X-chromosome inactivation (Brown et al., 1992; Clemson et al., 1996; Panning et al., 1997). Specifically, using RNA fluorescence *in situ* hybridization, *Xist* RNA is observed to paint the inactive X chromosome in mammalian female cells (Clemson et al., 1996). Under conditions designed to detect *Xist* RNA but

reduce the signal from nascent transcription (Figure 2A), we examined the localization of the long isoform of *hNEAT1* and *hNEAT2* in nuclei of human fibroblasts.

Unlike *Xist*, neither *hNEAT1* nor *hNEAT2* appear to localize to a particular region of the nucleus. Instead, both transcripts display dispersed, punctate signals within the nucleus (Figure 2B, C). Recent observations suggest that a small number of cells (around 5%) also show more localized RNA signal for *hNEAT1*, sometimes in a manner similar to *Xist* (data not shown). Further analysis, such as RNA FISH combined with chromosomal painting will be required to determine if this reflects an *Xist*-like role for the transcript. Since the dispersed, punctate signal of expression is observed in most cells, we began by characterizing its localization. In order to determine whether this punctate signal represented the true localization of *hNEAT1* in the nucleus instead of nonspecific background, we performed two-color FISH using probes to different regions of *hNEAT1*. These two unique probes showed a striking overlap, suggesting that the punctate RNA FISH signal is specific to *hNEAT1* RNA (Figure 2D). Using two-color RNA FISH, we saw no overlap of *hNEAT1* and *hNEAT2* signal, suggesting that these noncoding RNAs reside in different locations within the nucleus (data not shown).

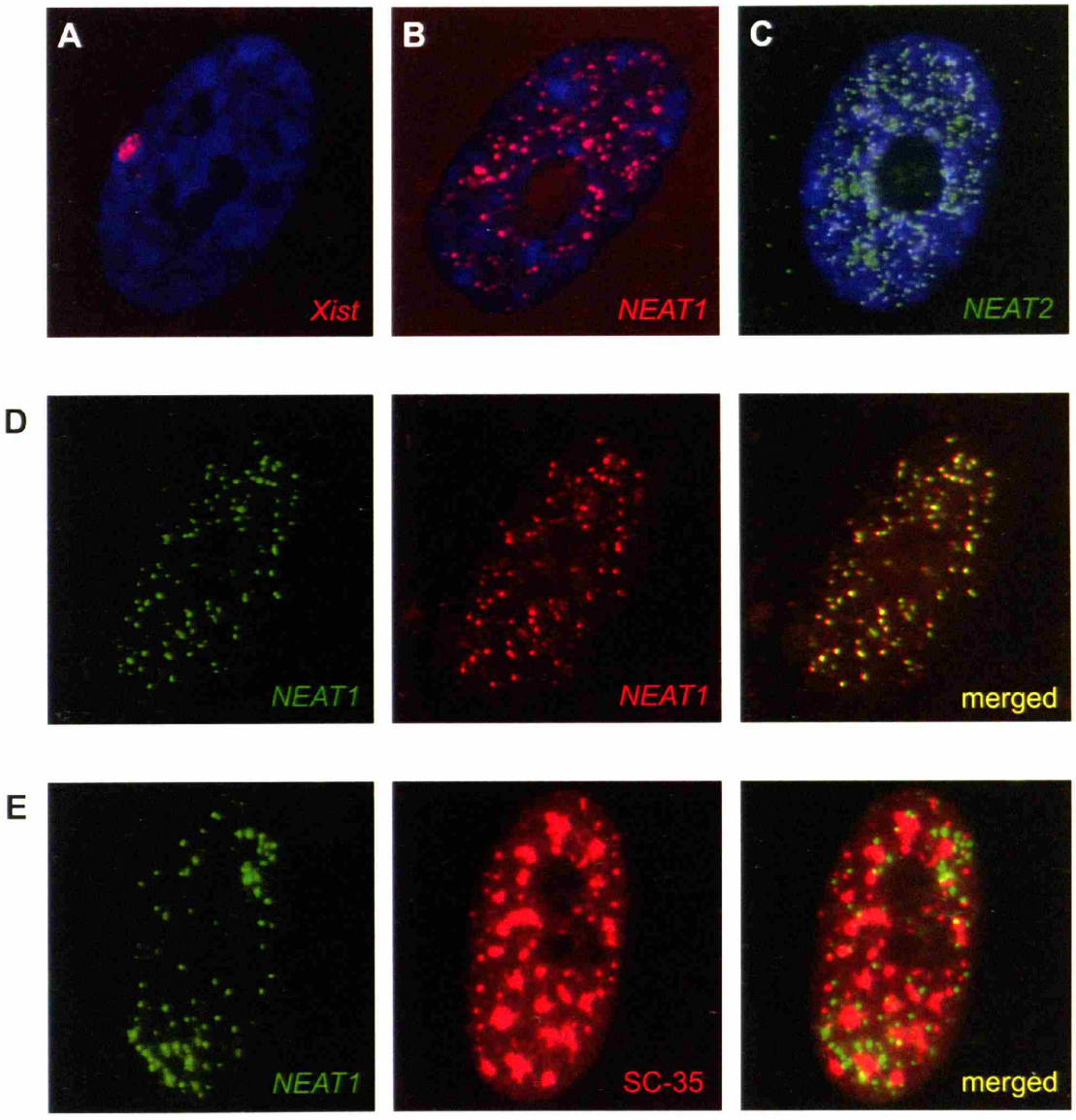
The splicing factor SC-35 localizes to several splicing domains within the nucleus, where structural RNAs are thought to reside (Clemson et al., 1996; Huang et al., 1994). The absence of *Xist* from these domains was considered early evidence that the RNA did not play a structural role in nuclear architecture (Clemson et al., 1996). Like *Xist, hNEAT1* does not localize to SC-35 domains (Figure 2E).

It is important to note that these RNA FISH analyses were performed prior to the identification of a 4 kb transcript at the *hNEAT1* locus. After characterization of the 4 kb

Figure 2.

RNA FISH analysis of *hNEAT1* and *hNEAT2*. **(A)** Probe to the *hXist* locus in female human fibroblasts shows the canonical pattern of *Xist* localization to a single domain (the inactive X) in the nucleus. In contrast, the long isoform of *hNEAT1* **(B)** as well as *hNEAT2* **(C)** are present at many spots within the nucleus. **(D)** RNA FISH to *hNEAT1* using distinct, non-overlapping probes demonstrates that this signal is likely not due to noise, but rather represents the true localization of the *hNEAT1* transcript in the nucleus. **(E)** *hNEAT1*, unlike structural RNAs, does not appear to localize to SC-35 splicing domains.

FIGURE 2


A
Xist
B
NEAT1
C
NEAT2
D
NEAT1
NEAT1
merged
E
NEAT1
SC-35
merged

hNEAT1 transcript by 5' RACE and Northern analysis, it was discovered that the two RNA FISH probes used to examine *hNEAT1* had either no, or very little overlap, to the smaller (4 kb) human transcript. Thus, if the localization of the small isoform of *hNEAT1* differs from that of the larger form, probes designed to the 4 kb transcript might display an entirely different pattern in RNA FISH. With this in mind, RNA FISH will be performed with new probes in order to determine if both isoforms of *hNEAT1* share an overlapping, punctate distribution within the nucleus.

Examining transcription around *NEAT1* and *NEAT2*

Xist transcription is regulated by the presence of antisense transcription on the opposite strand of the locus (Lee et al., 1999). This antisense transcript, *Tsix*, is expressed early in development, and its transcription is antagonistic to Xist expression from the same allele. In order to investigate whether antisense transcription occurs at either NEAT locus, we used real-time PCR analysis to determine the relative transcription levels of both the sense and antisense strands of each locus.

Reverse transcription of RNA using either olido-dT or random primers, followed by PCR, does not distinguish between transcription of different strands. In order to quantify the level of transcription from each strand of NEAT1 and NEAT2, we used gene-specific primers to reverse transcribe RNA from a number of cell types. Using PCR primers located less than 100 bp from each RT primer, we used quantitative real-time PCR analysis to assay strand-specific transcription in two regions in the *NEAT1* transcript and a central position within *NEAT2*. No antisense transcription was detected above the level of control RT reactions (in which no primer was added) for either *hNEAT1* or

hNEAT2 in human female lymphoblasts or HeLa cells. Based on comparison to a standard curve of random primed cDNA, the difference at the 3' end of the *hNEAT1* 4 kb transcript between sense and antisense transcript is at least 222.4 (+/- 22.4)-fold in these lymphoblasts. Likewise, sense transcription predominates near the presumptive 3' end of the larger *hNEAT1* transcript (82.7 +/- 25.5) and at the *hNEAT2* locus (781.4 +/- 46.7). Similar results were obtained for HeLa cells (data not shown). Differences in relative numbers likely reflect changes in the overall expression level of these transcripts, as the - primer control remained constant for all 3 assays, whereas the total amount of sense transcript varied. Because transcription of *Tsix* across the *Xist* locus is developmentally regulated (Lee et al., 1999), we examined human embryonic stem cells for antisense transcription overlapping *hNEAT1* and *hNEAT2*. While both *hNEAT1* and *hNEAT2* were expressed in the undifferentiated female human embryonic stem cell line H9 (NIH code WA09), no evidence of antisense transcription was detected at either locus. For the *hNEAT1* assay located near the 3' end of the 4 kb transcript, sense transcript levels were 39.9 (+/- 5.5)-times greater than the level seen in no RT primer control and antisense reactions. Likewise, sense transcription was 21.4 (+/- 2.9)-fold higher near the putative 3' end of the large (>17 kb) *hNEAT1* transcript. The *hNEAT2* locus is also transcribed unidirectionally in human ES cells, with 194.8 (+/- 16.5)-fold higher levels of sense transcripts that antisense. Taken together, these results suggest that antisense transcription does not occur at detectable levels within either the *hNEAT1* or *hNEAT2* locus, in either undifferentiated or differentiated cells.

Functional analysis of *hNEAT1* and *hNEAT2* using RNA interference

In order to gain insight into the function of *hNEAT1* and *hNEAT2*, we decided to use RNA interference to reduce the level of expression of each noncoding RNA in transfected cells. As both of these transcripts are predominantly nuclear, it was unclear whether RNA interference would succeed in targeting either transcript. This is because the RNA-induced silencing complex (RISC), responsible to cleaving mRNAs in siRNA-mediated silencing, was thought to primarily reside in the cytoplasm. However, recent evidence demonstrated that some nuclear RNAs could indeed be targets for RISC degradation (Robb et al., 2005). siRNA constructs were designed for both *hNEAT1* and *hNEAT2* (Kim et al., 2005; Rose et al., 2005) and were individually transfected into HeLa cells.

Quantitative real-time RT-PCR analysis of RNA from transfected samples was performed, normalized using primers to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as an endogenous control. The level of *NEAT1* and *NEAT2* transcription was significantly reduced, relative to non-specific control siRNA (Figure 3). Such a reduction in expression was seen as early as 24 hours post-transfection and was comparable to experiments involving the targeting of protein coding genes (data not shown).

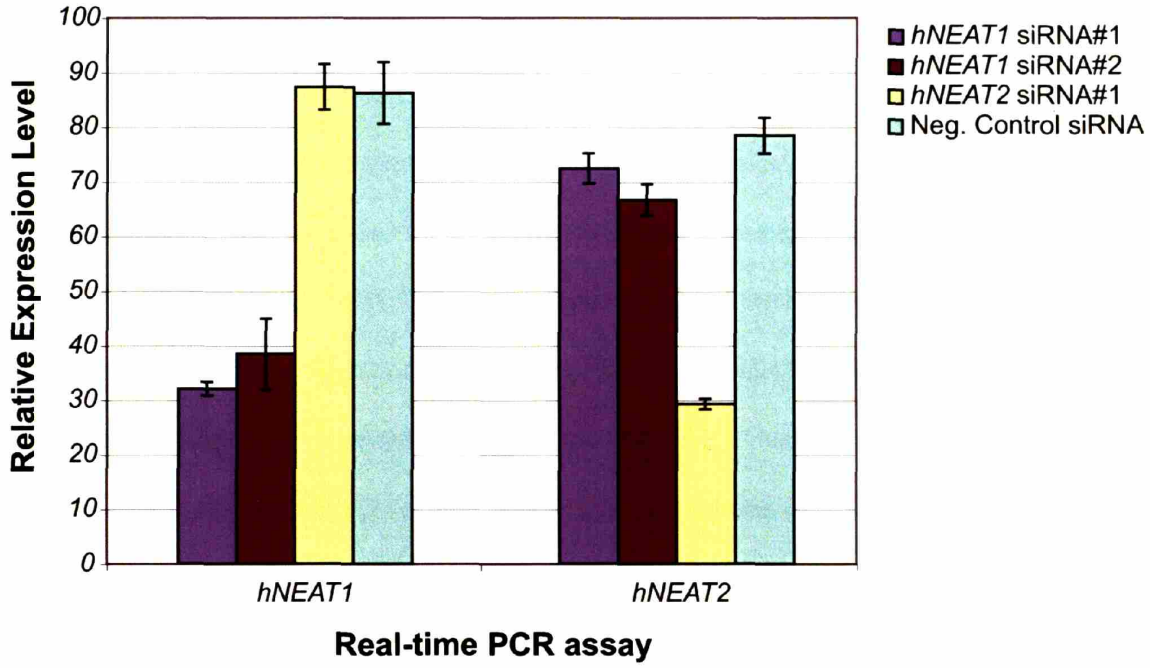
Comparative genome analysis of *NEAT1* and *NEAT2*

Through comparative genomic analysis, we sought to identify the mouse homolog of *hNEAT1* and *hNEAT2*. Noncoding RNAs are typically less conserved between species than protein coding sequences, likely due to less stringent requirements for their sequence content (Jareborg et al., 1999; Nesterova et al., 2001). In particular, where insertions and

Figure 3.

Specific reduction of *hNEAT1* and *hNEAT2* by RNA interference in HeLa cells. The levels of nuclear, noncoding transcripts, *hNEAT1* and *hNEAT2* can be reduced through RNA interference (RNAi). Specific small-interfering RNA (siRNA) to each gene specifically reduces the level of transcript within HeLa cells 48 hours after transfection with oligofectamine.

FIGURE 3


100
90
80
70
60
50
40
30
20
10
0
Relative Expression Level
hNEAT1
hNEAT2
Real-time PCR assay
hNEAT1 siRNA#1
hNEAT1 siRNA#2
hNEAT2 siRNA#1
Neg. Control siRNA

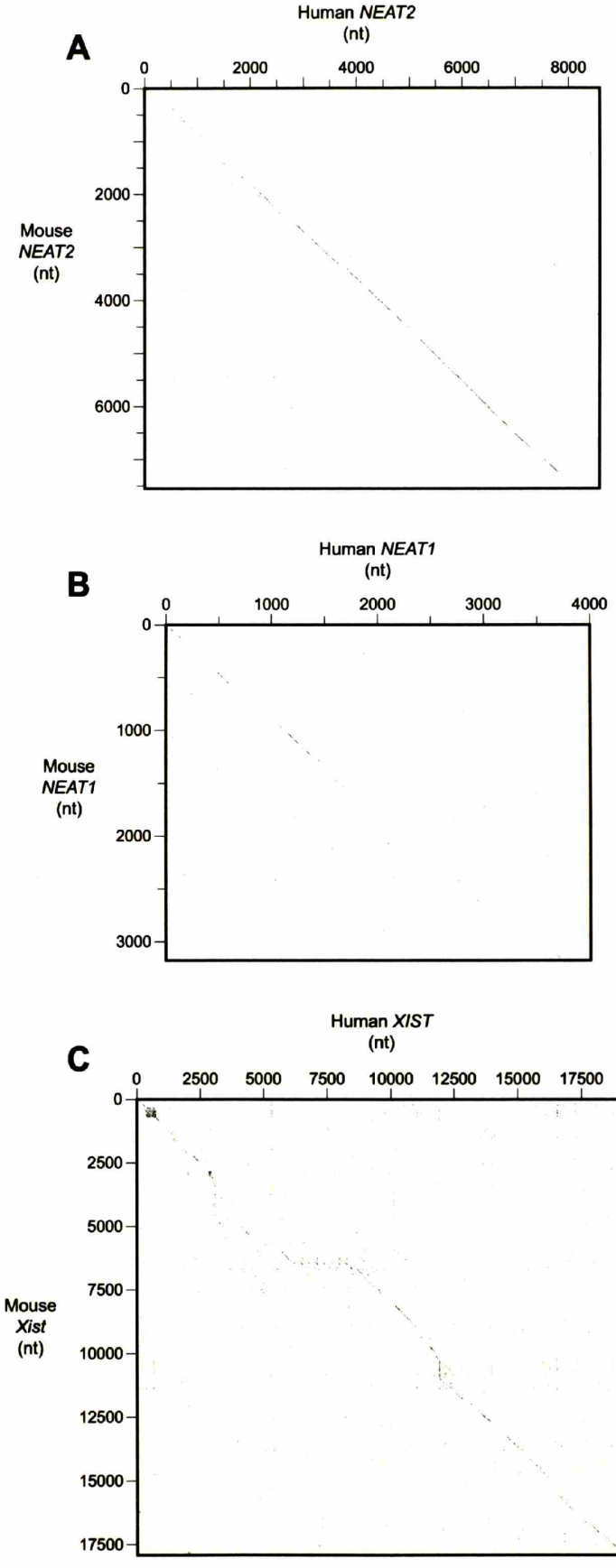
deletions often disrupt open reading frames by frame-shift mutations, noncoding RNA function is likely more tolerant of such changes, provided that they do not interfere with secondary structure or function. *hNEAT2* is quite conserved for a noncoding RNA (Figure 4A), and its mouse homolog was previously discovered in the initial description of the transcript (Ji et al., 2003). Note that unlike *Xist* (Figure 4C), *NEAT2* conservation does not deviate much from the diagonal. *Xist* contains a series of expanded repeats in human and in mouse, indicated by vertical or horizontal stretches on the dot plot (Brown et al., 1992). In contrast, *NEAT2* does not appear to contain such repeats, nor are there many other insertions or deletions between *hNEAT2* and *mNEAT2*, perhaps suggesting that the transcript is less tolerant of such changes than *Xist*. Comparative genomic analysis indicates that *NEAT2* is conserved within the genomes of *Mus musculus* (mouse), *Bos Taurus* (cow), *Canis familiaris* (dog), *Homo sapiens* (human), *Rattus norvegicus* (rat), and *Pan troglodytes* (chimp). The presence of a *NEAT2* homolog in the non-eutherian opossum, *Monodelphis domestica*, together with the absence of the transcript in non-mammalian species such as zebrafish and chickens suggests that *NEAT2* noncoding RNA is specific to the mammalian lineage.

Conservation of *hNEAT1* between the human and mouse lineage is less than that of *hNEAT2*. The presence of two highly conserved sequences within *NEAT1*, along with synteny between the human and mouse genomes, facilitated the discovery of the mouse homolog of *NEAT1* (Figure 4B). Using 5' and 3' RACE we have delineated the 3.1 kb mouse transcript on mouse chromosome 19, located approximated 50 kb from *mNEAT2*. As is the case with *hNEAT1*, *mNEAT1* has no significant open reading frames (with the largest predicted ORFs as 96 aa, 76 aa, and 90 aa in each of the three frames).

Figure 4.

Comparison of *NEAT1* and *NEAT2* between the human and mouse genomes. (A) Using the dot plot program, jdotter, the genomic region spanned by *NEAT2* in humans was compared to its homolog on mouse chromosome 19. The dot plot shows strong conservation across the length of the transcript. **(B)** The mouse homolog of *NEAT1* also resides within a region of mouse chromosome 19 syntenous to the region of human chromosome 11 containing *hNEAT1*. The mouse transcript was delineated using 5' and 3' RACE. The dot plot details the first 4kb of the human *NEAT1* transcript as compared to the 3177 nt mouse *NEAT1* transcript. Two islands of relatively high identity were identified by this comparison. The first region stretches from nucleotide position 483-597 of hNEAT1 and 450-565 of *mNEAT1*. This region shows 87% identity between the two species. A second region spans positions 1113-1368 of *hNEAT1* and 1013-1238 of *mNEAT1*. This region is 80% identical between the two species. **(C)** A dot plot between the primary human and mouse *Xist* transcripts, for comparison. For more discussion of *Xist* conservation, see (Brockdorff et al., 1992; Brown et al., 1992; Hong et al., 2000). (For all three dot plots, a sliding window size of 25 nt was used. The Greyramp tool was set between 50 and 150 for *Xist* and *NEAT2* and between 0 and 100 for *NEAT1* in order to highlight regions of increased conservation).

FIGURE 4


A
Human NEAT2
(nt)
0
2000
4000
6000
8000
Mouse
NEAT2
(nt)
B
Human NEAT1
(nt)
0
1000
2000
3000
4000
Mouse
NEAT1
(nt)
C
Human XIST
(nt)
0
2500
5000
7500
10000
12500
15000
17500
Mouse
Xist
(nt)

Furthermore, there is no conservation between reading frames of either *NEAT1* transcript. Unlike *hNEAT1*, however, the 3.1 kb transcript appears to be the only major transcript from the region around *mNEAT1*.

Islands of high conservation within the NEAT1 transcript allowed us to examine whether *NEAT1* homologs are present in the genomes of other mammals. This noncoding transcript is present in all other eutherian mammals examined. Strikingly, the conserved region of *hNEAT1* spanning positions 1113-1368 is also the most conserved region between humans and *NEAT1* of dogs, rats, and cows (data not shown). Interestingly, *NEAT1* is present twice in the dog genome, with two paralogs of *NEAT1* present on chromosome 18 and chromosome 30, respectively. In contrast, *NEAT2* is present only once in the dog genome, near *NEAT1* on chromosome 18. As there is no synteny between these two chromosomes in the surrounding sequence, the simple explanation of large-scale duplications within the dog lineage does not explain this event. No homologs of *NEAT1* were identified in either the chicken or opossum genomes. While close examination of the opossum genome surrounding *NEAT2* suggests that gaps in the assembly may explain the absence of a *NEAT1* transcript, the intriguing possibility exists that *NEAT1* may be specific to the lineage of eutherian mammals.

The identification of mouse homologs of *NEAT1* and *NEAT2* allowed us to examine the publicly available data from the Mouse Transcriptome Project (Edgar et al., 2002). This dataset examined expression levels of genes in different tissues using technology different that array-based approaches to expression profiling. Instead, it relied on the sequencing of over 2 million short expressed sequence tags (ESTs) for each sample analyzed (Brenner et al., 2000a; Brenner et al., 2000b). The relative abundance

of a particular tag is then determined per million tags sequenced. In addition to samples of different tissue types, EST populations of nuclear and post-nuclear subcellular fractionations were also examined in the mouse cell line BLK CL.4 (Patek et al., 1978) and mouse liver tissue. In BLK CL.4 cell lines, *mNEAT1* was enriched in nuclear fractions 42.8-fold and 2 tags for *mNEAT2* were enriched 80.8-fold and 90.4-fold. In liver tissue, *mNEAT1* was also nuclear-enriched (19.5-fold), as was one tag for *mNEAT2* (17.4-fold). This level of enrichment is significant, placing both *mNEAT1* and *mNEAT2* near the top of the list of nuclear-enriched, highly expressed transcripts for both cell types. Strikingly, using a different technology for expression analysis, in different tissues from a different organism, we have observed nuclear enrichment of *NEAT1* and *NEAT2*. Taken together, the level of conservation of sequence and subcellular localization in both human and mouse, suggest an important role for *NEAT1* and *NEAT2* within the nuclei of mammalian cells.

DISCUSSION

The epigenetic regulation of disparate loci in a coordinated fashion would appear to be a challenging hurdle for the developing mammalian cell. The coordinated asynchronous replication of a large number of autosomal genes, located in isolated replication domains across entire chromosomes, shares many similarities to X inactivation and mammalian dosage compensation (Ensminger and Chess, 2004; Singh et al., 2003). These phenomenological similarities may extend to the mechanisms by which such chromosome-wide coordination is achieved. In X inactivation, a monoallelic signal is propagated along the length of one X chromosome, through the spreading of a noncoding RNA, *Xist*. While the mechanisms by which *Xist* spreads in *cis* across the chromosome are not well understood (Hall et al., 2002; Lyon, 1998), it is clearly an elegant solution to a complex regulatory problem. One potential mechanism for autosomal coordination is that similar RNA species exist on each autosome, associating with the asynchronously replicating domains on one homolog but not the other.

We have examined the human nucleus for the presence of other noncoding RNAs that might function in a manner similar to *Xist*. Using Affymetrix arrays to examine nuclear and cytoplasmic RNA fractions from the same cells, it is clear that many probes on the U133 chipset target nuclear enriched sequences. However, the majority of these sequences are not derived from noncoding loci. In fact, some target sequences correspond to known protein coding genes, such as microtubule-actin crosslinking factor 1 (MACF1) and paired immunoglobulin-like type 2 receptor beta (PILRB). The enrichment of these genes within the nucleus may hold clues to RNA metabolism. Their enrichment may be indicative of high-level constitutive transcription together with

relatively rapid turnover of mRNA in the cytoplasm. Perhaps the regulation of these genes is post-transcriptional, with other factors (such as microRNAs) regulating RNA stability in the cytoplasm.

Of the noncoding sequences identified in this screen, most represent intronic sequences. This is perhaps surprising, since probes on the array were designed to Unigene clusters of ESTs. All of the nuclear enriched probes identified in our screen targeted the same strand as the transcription of the pre-mRNA they were contained in. This does not exclude the possibility that these sequences are sometimes transcribed independent of the larger pre-mRNA (Ashburner et al., 1999; Loebel et al., 2005). However, unless there is an uncharacterized bias towards parallel transcription between flanking pre-mRNA and intronic loci, it suggests that our results may indicate mispriming of olido-dT primers within the intronic sequence. Indeed, most of these introns contain polyA repeat regions downstream of the target sequence, allowing for their amplification during reverse transcription and EST library construction. Thus, these repeats likely explain the presence of such ESTs within the Unigene set used to design the Affymetrix U133 arrays. By examining nuclear enriched RNA, we have identified the presence of intronic probes within the chipset. These probes can now be compared to others contained within the corresponding mature mRNA in existing collections of array data in order to examine the metabolism of specific intronic sequences relative to their mature mRNA in a number of cell types and conditions.

A small transcript overlapping the *hNEAT1* locus has previously been identified in a screen for cDNA enriched in trophoblasts (Geirsson et al., 2004; Geirsson et al., 2003a; Geirsson et al., 2003b). Our studies show that trophoblast noncoding RNA

(TncRNA) maps to the 3' end of the 4 kb *hNEAT1* transcript. Overexpression of TncRNA has been shown to silence the expression of the MHC class II transactivator (CIITA). This transcript is not present in either fibroblasts or lymphoblasts. Our comparative genomic analysis shows no homology within the region of *NEAT1* overlapping with the TncRNA transcript, in stark contrast to the homology observed in other parts of the NEAT1 locus. As all of the data concerning TncRNA comes from overexpression of one cDNA clone, more work is required to ascertain the true function of this truncated transcript and its relationship to full-length *hNEAT1*.

NEAT1 and *NEAT2* are conserved between humans and mice. They are among the most enriched nuclear species in a variety of cell types from both species. This enrichment in the nucleus may indicate an epigenetic or otherwise transcriptional regulatory role for both RNAs. Perhaps the dispersed, punctate RNA FISH pattern observed for each transcript may represent loci under the influence of these noncoding RNAs. If so, the subnuclear localization of each noncoding transcript holds clues as to the function of each transcript within the nucleus. In addition, the ability to disrupt both *NEAT1* and *NEAT2* through specific siRNA constructs should provide further insights into the function of these noncoding RNAs. Expression array analysis of cells exposed to either *hNEAT1* or *hNEAT2* siRNA (compared to negative control transfections) may identify specific pathways that are regulated by the presence of these two noncoding transcripts in the nucleus.

MATERIALS AND METHODS

Subcellular fractionation of RNA and array analysis

Apparently normal human cells from the fibroblastoid and lymphoblastoid lineages were grown under typical conditions. WI-38 primary human fibroblasts were obtained from the American Type Culture Collection (ATCC). Human EBV-transformed human lymphoblasts, GM10852, from an apparently normal individual were obtained from the Coriell Cell Repositories (CCR). Nuclei were purified in triplicate from each cell line, using the Nuclei PURE nuclei isolation kit from Sigma Aldrich. RNA from these nuclei, as well as cytoplasmic fractions, was purified using Qiagen's RNeasy Maxi column-based system, per the manufacturer's instructions. RNA representing equivalent numbers of cell equivalents was run on U133A and U133B Affymetrix expression arrays.

Northern blot analysis

A commercially obtained Northern blot (Ambion's FirstChoice® Human Blot 2) was examined for the presence of *hNEAT1*, according to the manufacturer's instructions. Briefly, probe for human *NEAT1* Northern analysis was directed against human chr11:64946880 - 64947322 (May 2004 build) and was amplified from genomic DNA with primers hNEAT1NP1F (5'-TAGTTGTGGGGGAGGAAGTG-3') and hNEAT1NP1R (5'-TGGCATGGACAAGTTGAAGA-3'). PCR product was TOPO cloned into pCR4TOPO (Invitrogen) and the construct linearized with NotI. Probe was labeled with [α-32P]UTP by T3 RNA polymerase with the Ambion Strip-EZ RNA probe synthesis kit according the manufacturer's instructions. Labeled probe was hybridized to the blot using the Ambion NorthernMax formaldehyde-based system for Northern blots according to the manufacturer's instructions.

RNA FISH

RNA FISH was performed on WI-38 primary human fibroblasts under conditions designed to minimize the signal produced from nascent mRNA transcription, as previously described (Clemson et al., 1996; Panning et al., 1997). PCR products off of genomic DNA were used as template for either biotin using the Bioprime DNA labeling kit (Invitrogen) or digoxigenin using nick translation (Roche) labeling reactions. Primer sequences: hXistE6F: 5'- GGTCACATGCTGTGTGCTTTTTGTCCT-3'; hXistE6R: 5' CTGTAGGCCAGGTCAAGGTGGGTCTAA-3'; hNEAT1F: 5' CTAAAAAGGGAAGGGGATGGGGATTGT-3'; hNEAT1R: 5'-CATTTACCCGCATTTCACAGACACAGG-3'; hNEAT1#2F: 5'-TCGTGCAGCTGTGAGCATCTCTGTAAA-3'; hNEAT1#2R: 5'-ACGCCCTATGACCCAGGAATTTCACTT-3'; hNEAT2F: 5'-GGAAGACAGAAGTACGGGAAGGCGAAG-3'; hNEAT2R: 5'-CATCACTGAAGCCCACAGGAACAAGTC-3'.

Design of small interfering RNAs targeting hNEAT1 and hNEAT2

Dicer-substrate siRNAs were designed to produce only one predominant 21mer product (Rose et al., 2005), using Integrated DNA Technologies RNAi design tool. siRNAs were purified using RNase-free HPLC by the manufacturer and annealed prior to use, as per the manufacturer's instructions. The sequences for each siRNA are as follows: hNEAT1 site 1 Sense: 5' - [Phosphate]rCrUrG rGrUrA rUrGrU rUrGrC rUrCrU rGrUrA rUrGrG rUrAA G - 3'; hNEAT1 site 1 Antisense: 5' - rCrUrU rArCrC rArUrA rCrArG rArGrC

rArArC rArUrA rCrCrA rGrUrA - 3'; hNEAT1 site 2 Sense: 5' - [Phosphate]rGrUrG rArGrA rArGrU rUrGrC rUrUrA rGrArA rArCrU rUrUC C - 3'; hNEAT1 site 2 Antisense: 5' - rGrGrA rArArG rUrUrU rCrUrA rArGrC rArArC rUrUrC rUrCrA rCrUrU - 3'; hNEAT2 site 1 Sense: 5' - [Phosphate]rCrArA rGrUrA rArCrU rCrCrC rArArU rGrArU rUrUrA rGrUT T - 3'; hNEAT2 site 1 Antisense: 5' - rArArA rCrUrA rArArU rCrArU rUrGrG rGrArG rUrUrA rCrUrU rGrCrC - 3'; hNEAT2 site 2 Sense: 5' - rCrArG rGrArA rGrGrA rGrCrG rArGrU rGrCrA rArUrU rUrGG T - 3'; hNEAT2 site 2 Antisense: 5'-rArCrC rArArA rUrUrG rCrArC rUrCrG rCrUrC rCrUrU rCrCrU rGrGrA-3'; EGFP Sense: 5' - [Phosphate]rArGrC rUrGrA rCrCrC rUrGrA rArGrU rUrCrA rUrCrU rGrCA C - 3'; EGFP: Antisense: 5' - rGrUrG rCrArG rArUrG rArArC rUrUrC rArGrG rGrUrC rArGrC rUrUrG - 3'. Non-specific control #2 (21mer) siRNA from Ambion was also used in some experiments.

Transfection of siRNA into HeLa cells

Oligofectamine (Invitrogen) was used to transfect HeLa cells with siRNA duplexes, per the manufacturer's instructions. Briefly, $3x10^4$ cells were seeded into each well of a 24 well plate, with 0.5 ml of antibiotic-free media (DME, 10% uninactivated FBS) in each well. The next day, when cells were roughly 50% confluent, 60 pmol (3 μl of 20 μM) siRNA was mixed with 50 μl room temperature OptiMEM I reduced serum media (Invitrogen) and 3 μl Oligofectamine reagent was mixed with 12 μl OptiMEM I reduced serum media. After a 5 minute incubation at room temperature, the diluted siRNA and Oligofectamine were mixed and allowed to sit for an additional 20 minutes to allow for complex formation. Afterwards, this mixture was added to one well of the seeded 24

well plate, the plate was rocked gently, and cells were harvested for analysis either 24, 48, 72, or 96 hours post-transfection. For 72 and 96 hour time points, cells were split into 6 well dishes to prevent overconfluency. RNA was harvested directly from plates using TRIzol reagent (Invitrogen). DNA was removed from RNA samples using Ambion's DNAfree system. RNA was subjected to further clean-up using Qiagen RNeasy mini columns, per the manufacturer's instructions. Real-time quantitative PCR analysis was performed using "hNEAT1 site 1" and "hNEAT2" primer pairs (see below).

Quantitative Real-time PCR analysis

All real-time PCR analysis was performed using an Applied Biosystems 7500 machine, using SYBR green master mix (ABI). A ten-fold dilution series of random primed cDNA was used to quantify differences between sense and antisense transcription. For siRNA studies, cDNA from untreated cells was used for the standard curve. Primers were designed to produce small (60-100 nt) products off of cDNA (see below). Reverse transcription of RNA into cDNA proceeded using Superscript III Reverse Transcriptase (Invitrogen) and random decamer primers (Ambion). Post-run analyses were performed using Applied Biosystem's Sequence Detection Software (version 1.3).

Oligonucleotides

Oligonucleotide primers were designed using Primer3 and ordered from Integrated DNA Technologies (IDT). Strand-specific RT primers and PCR pairs: "hNEAT1 site 1 RT forward": 5'-AACCAATGACTTGGGGATGA-3'; "hNEAT1 site 1 RT reverse": 5'-TTGTGCTGTAAAGGGGAAGAA-3'; "hNEAT1 site 1 PCR forward": 5'-

TCGGGTATGCTGTTGTGAAA-3'; "hNEAT1 site 1 PCR reverse": 5'-TGACGTAACAGAATTAGTTCTTACCA-3'; "hNEAT1 site 2 RT forward": 5'-AATTCATGCTTTTGAAATGTTCT-3'; "hNEAT1 site 2 RT reverse": 5'-TTCACACAGACAGAGAGCACA-3'; "hNEAT1 site 2 PCR forward": 5'-TGCAACCATCGACACTATCC-3'; "hNEAT1 site 2 PCR reverse": 5'-GAGCGTCTGTTTGGGATGAC-3'; "hNEAT2 RT forward": 5'GGGATCAAGTGGATTGAGGA-3'; "hNEAT2 RT reverse": 5'-AAGCACTTATCCCTAACATGCAA-3'; "hNEAT2 PCR forward": 5'-TCGTTTGCCTCAGACAGGTA-3'; "hNEAT2 PCR reverse": 5'-GCTCCCAGATGAAATGAAGC-3'; For siRNA studies, "hNEAT1 site 1" and "hNEAT2 PCR" primers were used, "GAPDH forward": 5'-GATCATCAGCAATGCCTCCT-3'; "GAPDH reverse": 5'-TGTGGTCATGAGTCCTTCCA-3'.

REFERENCES

Ashburner, M., Misra, S., Roote, J., Lewis, S. E., Blazej, R., Davis, T., Doyle, C., Galle, R., George, R., Harris, N., *et al.* (1999). An exploration of the sequence of a 2.9-Mb region of the genome of Drosophila melanogaster: the Adh region. Genetics *153*, 179-219.

Bix, M., and Locksley, R. M. (1998). Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. Science *281*, 1352-1354.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al.* (2000a). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol *18*, 630-634.

Brenner, S., Williams, S. R., Vermaas, E. H., Storck, T., Moon, K., McCollum, C., Mao, J. I., Luo, S., Kirchner, J. J., Eletr, S., *et al.* (2000b). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. Proc Natl Acad Sci U S A *97*, 1665-1670.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., and Willard, H. F. (1991a). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., and Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527-542.

Brown, C. J., Lafreniere, R. G., Powers, V. E., Sebastio, G., Ballabio, A., Pettigrew, A. L., Ledbetter, D. H., Levy, E., Craig, I. W., and Willard, H. F. (1991b). Localization of the X inactivation centre on the human X chromosome in Xq13. Nature *349*, 82-84.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834.

Chow, J. C., Hall, L. L., Clemson, C. M., Lawrence, J. B., and Brown, C. J. (2003). Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics *82*, 309-322.

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol *132*, 259-275.

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

Ensminger, A. W., and Chess, A. (2004). Coordinated replication timing of monoallelically expressed genes along human autosomes. Hum Mol Genet *13*, 651-658.

Geirsson, A., Bothwell, A. L., and Hammond, G. L. (2004). Inhibition of alloresponse by a human trophoblast non-coding RNA suppressing class II transactivator promoter III and major histocompatibility class II expression in murine B-lymphocytes. J Heart Lung Transplant *23*, 1077-1081.

Geirsson, A., Lynch, R. J., Paliwal, I., Bothwell, A. L., and Hammond, G. L. (2003a). Human trophoblast noncoding RNA suppresses CIITA promoter III activity in murine B-lymphocytes. Biochem Biophys Res Commun *301*, 718-724.

Geirsson, A., Paliwal, I., Lynch, R. J., Bothwell, A. L., and Hammond, G. L. (2003b). Class II transactivator promoter activity is suppressed through regulation by a trophoblast noncoding RNA. Transplantation *76*, 387-394.

Gimelbrant, A. A., Ensminger, A. W., Qi, P., Zucker, J., and Chess, A. (2005). Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. J Biol Chem *280*, 1354-1359.

Guru, S. C., Agarwal, S. K., Manickam, P., Olufemi, S. E., Crabtree, J. S., Weisemann, J. M., Kester, M. B., Kim, Y. S., Wang, Y., Emmert-Buck, M. R., *et al.* (1997). A transcript map for the 2.8-Mb region containing the multiple endocrine neoplasia type 1 locus. Genome Res *7*, 725-735.

Hall, L. L., Clemson, C. M., Byron, M., Wydner, K., and Lawrence, J. B. (2002). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum Mol Genet *11*, 3157-3165.

Hong, Y. K., Ontiveros, S. D., and Strauss, W. M. (2000). A revision of the human XIST gene organization and structural comparison with mouse Xist. Mamm Genome *11*, 220-224.

Huang, S., Deerinck, T. J., Ellisman, M. H., and Spector, D. L. (1994). In vivo analysis of the stability and transport of nuclear poly(A)+ RNA. J Cell Biol *126*, 877-899.

Jareborg, N., Birney, E., and Durbin, R. (1999). Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. Genome Res *9*, 815-824.

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P. M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., *et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene *22*, 8031-8041.

Kim, D. H., Behlke, M. A., Rose, S. D., Chang, M. S., Choi, S., and Rossi, J. J. (2005). Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. Nat Biotechnol *23*, 222-226.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., Nicholls, R. D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. Nature *364*, 459-463.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet *21*, 400-404.

Lee, J. T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275-279.

Loebel, D. A., Tsoi, B., Wong, N., and Tam, P. P. (2005). A conserved noncoding intronic transcript at the mouse Dnm3 locus. Genomics *85*, 782-789.

Lucchesi, J. C., Kelly, W. G., and Panning, B. (2005). Chromatin Remodeling in Dosage Compensation. Annu Rev Genet.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

Lyon, M. F. (1998). X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet *80*, 133-137.

Migeon, B. R. (2003). Is Tsix repression of Xist specific to mouse? Nat Genet *33*, 337; author reply 337-338.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221-225.

Muller-Tidow, C., Diederichs, S., Thomas, M., and Serve, H. (2004). Genome-wide screening for prognosis-predicting genes in early-stage non-small-cell lung cancer. Lung Cancer *45 Suppl 2*, S145-150.

Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N., and Zakian, S. M. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res *11*, 833-849.

Panning, B., Dausman, J., and Jaenisch, R. (1997). X chromosome inactivation is mediated by Xist RNA stabilization. Cell *90*, 907-916.

Patek, P. Q., Collins, J. L., and Cohn, M. (1978). Transformed cell lines susceptible or resistant to in vivo surveillance against tumorigenesis. Nature *276*, 510-511.

Priest, J. H., Heady, J. E., and Priest, R. E. (1967). Delayed onset of replication of human X chromosomes. J Cell Biol *35*, 483-487.

Robb, G. B., Brown, K. M., Khurana, J., and Rana, T. M. (2005). Specific and potent RNAi in the nucleus of human cells. Nat Struct Mol Biol *12*, 133-137.

Rose, S. D., Kim, D. H., Amarzguioui, M., Heidel, J. D., Collingwood, M. A., Davis, M. E., Rossi, J. J., and Behlke, M. A. (2005). Functional polarity is introduced by Dicer processing of short substrate RNAs. Nucleic Acids Res *33*, 4140-4156.

Schmidt, M., and Migeon, B. R. (1990). Asynchronous replication of homologous loci on human active and inactive X chromosomes. Proc Natl Acad Sci U S A *87*, 3685-3689.

Selig, S., Okumura, K., Ward, D. C., and Cedar, H. (1992). Delineation of DNA replication time zones by fluorescence in situ hybridization. Embo J *11*, 1217-1225.

Singh, N., Ebrahimi, F. A., Gimelbrant, A. A., Ensminger, A. W., Tackett, M. R., Qi, P., Gribnau, J., and Chess, A. (2003). Coordination of the random asynchronous replication of autosomal loci. Nat Genet *33*, 339-341.

Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapinoso, L. M., Moqrich, A., *et al.* (2002). Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A *99*, 4465-4470.

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., *et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A *101*, 6062-6067.

Takagi, N. (1974). Differentiation of X chromosomes in early female mouse embryos. Exp Cell Res *86*, 127-135.

Vasques, L. R., Klockner, M. N., and Pereira, L. V. (2002). X chromosome inactivation: how human are mice? Cytogenet Genome Res *99*, 30-35.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Conclusions and Afterword

CONCLUSIONS

The work presented in this thesis advances the field of epigenetic regulation of monoallelic expression and random asynchronous replication in a number of ways. First, the studies have shown the surprising result that the random asynchronous replication of autosomal genes is coordinated at the level of whole chromosomes in both humans and mice. This coordination is similar to that seen for X-linked genes in mammalian female cells subject to X-chromosome inactivation. By studying autosomal coordination in human cells, we have established that this process, like X inactivation, is a true chromosome-wide process, capable of crossing the centromeric boundary. Both X inactivation and autosomal coordination are examples of chromosome-wide modifications that render the two genetically similar homologs of each chromosome epigenetically different.

Second, this work has examined the relative replication timing of individual trisomic loci. We have shown that the "N-1" rule of X-chromosome inactivation can be observed using the FISH assay of replication timing. The "N-1" rule was defined cytologically and refers to the number of inactive X chromosomes present in an otherwise diploid cell. With respect to asynchronous replication of X-linked genes, one allele replicates early and the remaining alleles replicate late. When applied to the study of asynchronously replicating loci on autosomal trisomies, a similar rule applies, with one allele replicating early and the remaining two alleles replicating late.

Third, this work has explored the role of noncoding RNA transcripts in the regulation of autosomal coordination and other epigenetic phenomena. Noncoding RNAs may propagate the chromosome-wide decision of autosomal coordination from a central

autosomal coordinating center, in a manner analogous to the transcription of *Xist* from the X-inactivation center (*Xic*). Two noncoding RNAs on human chromosome 11 were identified in an expression array based screen for nuclear enriched transcripts. These noncoding RNAs, *NEAT1* and *NEAT2* are conserved within the mammalian lineage. Their localization within the nucleus of mammalian cells may suggest a role in transcriptional regulation, RNA metabolism, or the establishment of subnuclear structures. Chromosome 11 contains over 40% of the odorant receptor loci in humans, and these two noncoding RNAs continue to be examined with respect to a role in autosomal coordination.

With the initial description of this chromosome-wide regulation of autosomal loci, several questions remain. Identification of *Xist*-like transcripts involved in autosomal coordination would certainly advance our understanding of the mechanisms underlying this process. Future directions stemming from this work, many of which do not require the existence of such transcripts, are outlined below.

Potential mechanisms of coordination

With the recent discovery of autosomal coordination, one of the first questions that comes to mind is how can large interspersed replication domains on the same chromosome coordinate whether to replicate early or replicate late. While the mechanisms of coordination remain to be elucidated, there are a number of possible models for its establishment and/or maintenance.

Perhaps the simplest model comes by way of analogy to X inactivation in female mammalian cells. X-linked genes replicate asynchronously in these cells because of

chromosome-wide changes to the underlying chromatin structure of the inactivated X (Priest et al., 1967; Schmidt and Migeon, 1990; Takagi, 1974). The establishment of X inactivation, however, depends on the spreading, in *cis*, of the noncoding RNA, *Xist*, across only one of the two copies of the X chromosome (Brockdorff et al., 1992; Brown et al., 1992; Panning et al., 1997; Wutz and Jaenisch, 2000). Established around the time of blastula implantation, this random choice also involves another noncoding RNA transcribed antisense to the major *Xist* transcript, *Tsix* (Lee et al., 1999). The expression of these two overlapping transcripts is mutually exclusive: if *Xist* is expressed from one allele, *Tsix* is expressed from the other. While the dichotomy between *Tsix* and *Xist* expression is thought to play an important role in the stochastic nature of X-chromosome inactivation, the existence or role of a human *Tsix* transcript remains contentiously debated (Chow et al., 2003; Migeon, 2003; Vasques et al., 2002).

Thus, one model for autosomal coordination is that, like X inactivation, noncoding transcripts play an important role in propagating a random choice, made at a discrete part of a chromosome to the far reaches of that homolog. In this way, monoallelic expression at one locus would affect the replication timing of all loci on the same chromosome. Perhaps early steps along the path to an inactive X chromosome are conserved on the autosomes, while the more drastic downstream chromatin changes are reserved for the X chromosome. This model has many predictions, the chief among them that each autosome must have a locus, analogous to the *Xic*, where a noncoding transcript is stably transcribed from one of two alleles at some point in development. Unfortunately, technical limitations or biological features, such as the relative dispensability of *Xist* expression later in development, might hinder efforts to find these

transcripts. For instance, the persistence of *Xist* expression in somatic tissues is not required for the maintenance of X inactivation. If the autosomal *Xist*-like transcripts are equally dispensable for the maintenance of coordination, they may only be observable during specific windows of development. Failure to find such transcripts in specific cell types could therefore not be taken as evidence against their existence.

Another possible mechanism for autosomal coordination is that spatial features within the nucleus are involved in determining which homolog's affected genes will replicate early and which will replicate late. Perhaps there is one nuclear neighborhood where chromosomes are marked as early and one where they are marked as late. We have not observed any spatial association of early replicating genes on different chromosomes within the nucleus using our FISH assays, but these assays are deliberately designed to not maintain nuclear structure. It is possible that at some point in development, homologs are marked as early or late due in part to their relative locations within the nucleus. Indeed, nuclear localization may play a role in the establishment of expression in nonrandom imprinted loci (Gribnau et al., 2003) as well as in the random allelic exclusion of immunoglobulins (Kosak et al., 2002; Skok et al., 2001). Perhaps similar events are involved in establishing random chromosome-wide epigenetic marks in the nucleus (marks that are presumably involved in autosomal coordination).

Asynchronous replication as a potential means to maintain genomic stability

A number of factors work to prevent genomic instability in mitotically dividing mammalian cells. Chief among these are the many checkpoints that regulate progression from G1 to S, intra-S-phase, G2 to M, and metaphase to anaphase during the cell cycle

(Zhou and Elledge, 2000). One source of genomic instability is the presence of repetitive elements in mammalian genomes, in particular microsatellite repeats (Cleary and Pearson, 2005; Pearson et al., 2005). The expansion of trinucleotide repeats contribute to a number of human diseases, including Huntington disease (MacDonald et al., 2003; Paulson and Fischbeck, 1996). The expansion of these repeats is thought to arise from various errors in DNA metabolism, such as slippage of the DNA polymerase and unequal crossing over during mitosis (Brown and Brown, 2004).

Olfactory receptors, interleukins, natural killer cell receptors, and a number of other monoallelically expressed genes exist in large, repetitive chromosomal domains. The large number of olfactory receptors in the mammalian genome is a consequence of local repeat expansion (Young et al., 2002). Clusters of olfactory receptors in humans range from a solitary gene in a few loci to the two clusters on chromosome 11 which each contain over 100 ORs (Glusman et al., 2001). Even genes such as the immunoglobulins, which are not clustered in the typical sense, exist as highly repetitive sequences, with unrearranged loci containing hundreds of very similar segments. Could the highly repetitive nature of these loci, left unregulated, lead to genomic instability, either through unequal mitotic crossing over or other means? In dealing with such repeats, cellular mechanisms must balance genomic stability with the evolutionary advantages afforded by gene diversification. Is asynchronous replication, and its chromosome-wide coordination, a response to this problem?

If asynchronous replication is a response to genomic repeats, perhaps the repetitive nature of a sequence might be of predictive value in determining whether it replicates asynchronously in S-phase or not. One way to examine this question would be

to choose a number of repetitive sequences in either the human or mouse genomes and use the FISH assay to determine whether they replicate synchronously or asynchronously. (Because direct probing of repetitive sequences is not compatible with the FISH assay, unique sequences within these repeats would have to be probed.) An unbiased sampling of several clusters of differing size and repeat length might provide evidence for a link between asynchronous replication and genomic repeats, if the number of asynchronously replicating probes is highly enriched relative to the genome as a whole. Another approach might be to introduce ectopic clusters into the genome and determine whether these sequences replicate asynchronously. One source of such clusters might be well-characterized mouse transgene arrays, for which the number of insertions, as well as the location of the insertion site, were both known. First, it would be determined whether the transgene contained an otherwise synchronously replicating gene and the presumptive insertion site was also synchronously replicating. If transgenic arrays of certain lengths for several of these cases replicate asynchronously, it would provide evidence for a causal relationship between genomic repetition and asynchronous replication. Repetition may contribute to asynchronous replication but may not be sufficient. If so, multiple insertion sites and transgenes may need to be examined in order to detect any causal link between asynchronous replication and repetitive sequence.

Examining the evolutionary relationship between X-chromosome inactivation and autosomal coordination

Dosage compensation is a common regulatory challenge confronting many species within the metazoan lineage (Lucchesi et al., 2005; Parkhurst and Meneely,

1994). Despite facing a similar challenge, different branches of this lineage have utilized different approaches to normalize gene expression between their heterogametic and homogametic sexes. It is likely that these different dosage compensation strategies represent independent adaptations to a common evolutionary challenge. Within the mammalian lineage, the strategy adopted for dosage compensation utilizes a chromosome-wide silencing mechanism. Evidence from other organisms suggests that dosage compensation does not necessarily require the differential treatment of the two X chromosomes in female cells. The observation of autosomal coordination, and the differential treatment of autosome-pairs, raises a number of interesting questions with respect to the evolution of mammalian dosage compensation. Some have suggested that these processes first established themselves on the X chromosome, from which they spread to select autosomal loci (Lee, 2003). Is this the case, or did the pre-existence of autosomal mechanisms get co-opted in the mammalian lineage for the purposes of X-chromosome inactivation? Perhaps chromosome-wide processes shaped the regulation of the predecessors to the modern day X and Y chromosomes while they were still an autosomal pair.

Most discussions of the evolutionary history of X-chromosome inactivation focus on the fact that in some parts of the mammalian lineage, X inactivation is random, whereas in others, it is imprinted (Huynh and Lee, 2005; Lee, 2003; Reik and Lewis, 2005). Specifically, X inactivation takes the form of a non-random, imprinted silencing of the paternal X chromosome in metatherian (marsupial) mammals such as the kangaroo and wallaby (Cooper et al., 1971; Johnston et al., 1975). X inactivation is also non-random in the extraembryonic tissues of mice, in which the paternal X is also always

silenced (Harper et al., 1982; Monk and Harper, 1979; Takagi and Sasaki, 1975). The nature of extraembryonic X inactivation in humans remains subject to debate, with evidence for (Goto et al., 1997; Harrison, 1989; Ropers et al., 1978) and against (Migeon and Do, 1979; Migeon et al., 1985) such imprinting in humans. These observations of imprinted X inactivation in some tissues might indicate that ancestral X inactivation more closely resembles imprinting than random autosomal monoallelic expression and asynchronous replication.

It is also possible that processes such as coordinated random asynchronous replication existed in the ancestral lineage prior to the establishment of X inactivation. A pre-existing mechanism for the differential treatment of two (nearly) sequence identical homologous chromosomes, could easily be adapted for the purposes of dosage compensation.

One means by which to explore the potential evolutionary relationship between X inactivation and autosomal coordination would be to determine if asynchronous replication and coordination was present prior to the establishment of X inactivation. A number of species should be examined for random asynchronous replication. The first species that should be examined is the chicken. Chickens have a surprising number of olfactory receptors. Recent evidence suggests that chicken orthologs of known imprinted mammalian genes replicate asynchronously in S-phase, especially when present on macrochromosomes (Dunzinger et al., 2005). Perhaps random asynchronous replication also occurs in chicken cells. Chromosome 5 has a number of olfactory receptors on it, with COR2 and COR4 located approximately 2 Mb from COR6 and COR3b. Probes to each of these COR clusters, along with probes to sequences between the two, should be

used in the FISH assay to look for asynchronous replication and coordination. Also, large tracts of chicken chromosome 4 show homology to the mammalian X chromosome, as the X and Y chromosomes were part of an ordinary autosome pair prior to the divergence of mammals and birds (Hillier et al., 2004). Perhaps chromosome-wide processes were already shaping that ancestral autosome pair. To that end, several probes within this region, especially in and around the neighborhood of the future Xic, should also be examined for asynchronous replication.

In addition to the avian lineage, diverse mammalian species should be examined for information about the evolutionary history of X inactivation and autosomal coordination. In particular, cell lines from marsupials, which undergo imprinted X inactivation, should be examined to determine whether the orthologs of genes which replicate asynchronously in humans and mice, also replicate asynchronously in this lineage. Analyses of this type are currently complicated by the dearth of genomic sequence for these species, but this limitation should be temporary in nature.

Not much is known about either X inactivation or autosomal coordination in monotremes, such as the platypus. Only recently was it discovered that platypus sex chromosomes are surprisingly different from those present in eutherian mammals (Grutzner et al., 2004; Waters et al., 2005). Specifically, the platypus has 5 copies of each sex chromosome that align as chains during meiosis. In female platypus cells, with two chains of 5 X chromosomes each, does X inactivation occur? If so, what does it look like? Do the 5 X chromosomes from one of the chains show any of the hallmarks of chromosome-wide silencing? If the level of conservation at the Xist locus between mice and humans is any indication, it is highly unlikely that answers will be found relying

entirely on comparative genomic analysis. Instead, many of the earliest observations about X inactivation such as asynchronous replication and chromatin condensation should be examined cytologically in these cells. The expression of X-linked genes should be examined in whole tissues derived from monotremes. If X-linked genes are expressed from only one allele in these whole tissues, it would strongly suggest imprinted X inactivation, with all the cells making the same choice with regard to expression. If not, analyses of single cells or clonal cell lines could determine if random monoallelic expression underlies dosage compensation in these animals. Biallelic expression of the vast majority of X-linked genes might suggest that other mechanisms equilibrate the levels of X-linked expression between the two sexes.

Generating genetically tractable systems for the study of monoallelic expression

The identification of monoallelically expressed genes that can be examined in tissue culture cells will likely facilitate a number of important investigations. Mechanistically, little is understood concerning either the establishment or maintenance of random monoallelic expression on the autosomes. Many of the epigenetic factors involved in X-chromosome inactivation or other heterochromatic states may mark one of the two alleles for presumptive silencing. Most of the current methods for characterizing monoallelic expression do not seem well adapted to high-throughput screens designed to identify modulators of expression. These methods typically involve reverse-transcription of RNA into cDNA, amplification of cDNA by PCR, and then analysis of relative amounts of each allele. This analysis depends on the presence of single nucleotide polymorphisms (SNPs) and modified genotyping assays.

While full-scale forward screens for epigenetic modulators may not be very tractable with these methods, a number of previously identified genes should be examined for a role in random monoallelic expression. This could be done by RNA-interference or chemical inhibition. DNA methylation could be disrupted by 5-azadeoxycytidine, which has been shown to reverse X-chromosome inactivation (Hansen et al., 1996). Candidate genes might be disrupted by RNA-interference, in preparation for large-scale RNAi library screening.

While candidate gene approaches might prove useful in elucidating the mechanisms of monoallelic expression, forward genetic screens would likely be more powerful tools. Unfortunately, the currently available readouts of monoallelic expression are too labor intensive and thus prove inadequate for such large-scale screens. The discovery of genes expressed monoallelically in tissue culture raises the intriguing possibility of creating transgenic cell lines designed to monitor monoallelic expression. This might be done by modifying both alleles of a particular gene such that each allele carries a unique mark. Specifically, each allele could carry the endogenous open reading frame followed by an internal ribosomal entry site (IRES) and a marker. If one allele's transcription produced green fluorescent protein (GFP) and the other allele produced red fluorescent protein (RFP), monoallelic expression would be either green or red whereas biallelic expression would be produce yellow cells. Alternatively, the expression of fluorophores could be indirectly linked, via the expression on each allele of a different exogenous transcription factor that could activate a number of reporter constructs. Modeled after the Gal4/UAS system used in flies, this two-component system might

allow for more flexibility with respect to forward and reverse genetic screens and selections.

One of the advantages of using fluorescence to mark each allele is that it provides the ability to discern small changes in expression levels by the relative intensities of each fluorophore in individual cells. Studies that have examined the reversibility of X inactivation in mature cells have shown it to be a remarkably stable event. For instance, when *Xist* expression is genetically ablated in differentiated cells, reversion to biallelic expression is a very rare event (Wutz and Jaenisch, 2000). If autosomal monoallelic expression was even a fraction as stable as X inactivation, the ability to examine individual cells for subtle changes in gene expression will be crucial to identifying factors required for maintaining this epigenetic state.

Examining established epigenetic regulators for their role in asynchronous replication, coordination, and monoallelic expression

A number of factors are known to play a role in either establishing or maintaining X inactivation and other constitutive heterochromatin. Perhaps some of these same factors are involved in autosomal coordination and/or monoallelic expression. The development of lentiviral shRNA libraries raises the possibility of forward genetic screens for modulators of both asynchronous replication and monoallelic expression. However, the existing assays for both of these features do not, as of yet, facilitate such studies. Instead, starting with known effectors of X inactivation, a set of candidate genes, should be examined using a combination of RNA-interference, the FISH assay of replication timing, and PCR-based analysis of monoallelic expression.

Based on their implication in X inactivation and other epigenetic processes (Lucchesi et al., 2005), genes which should be disrupted by RNAi include: DNA methyltransferases (in particular Dnmt3b), the three isoforms of HP1 (Chadwick and Willard, 2003), the histone variant macroH2A (Costanzi and Pehrson, 1998), BRCA1 (which associates with the inactive X) (Ganesan et al., 2002), SWI/SNF proteins (Angelov et al., 2003), and members of both the PRC1 and PRC2 polycomb complexes (Levine et al., 2004). One complication of these analyses is the potential for functional redundancy, which may necessitate the simultaneous disruption of multiple genes in order to observe any effect on asynchronous replication or monoallelic expression.

Other approaches for identifying *Xist*-like molecules in the mammalian nucleus

Xist is a large, nuclear transcript with no significant open reading frames. It associates with the inactive X chromosome and a number of proteinaceous epigenetic modulators. Perhaps because of this association, *Xist* RNA is relatively insensitive to RNase-treatment *in vivo*. In addition, *Xist* is monoallelically expressed (Brown et al., 1991) and therefore the *Xic* replicates asynchronously in S-phase (Gribnau et al., 2005). Thus, the nuclear enrichment of *Xist* is not the only feature that can be harnessed by those looking for *Xist*-like epigenetic modulators.

One characteristic of *Xist* that has already been explored is its relative insensitivity to RNase-treatment. Evidence from RNA FISH studies, which typically use RNase A-treatment as a control for DNA cross-hybridization, suggested that unlike other RNA transcripts, *Xist* signal could not be disrupted by such treatment (J. Gribnau, personal communication). Instead, RNA FISH controls for *Xist* depended on post-

hybridization RNase H-treatment (Clemson et al., 1996). As RNase H digests the RNA in RNA-DNA hybrids, this treatment presumably works by removing the small amount of *Xist* RNA that anneal to the FISH probe. In Appendix 2, we outline a procedure for treatment of chromatin with a cocktail of RNase A and T1. The results of this procedure are a differential pattern of relative RNase-protection across the *Xist* and *NEAT2* transcripts. One might expect other noncoding RNAs to be enriched within the RNase-protected fraction of the transcriptome. By modulating the amount and types of RNase used, it may be possible to reduce the background level of protection to allow for the cloning of small, protected sequences. Through alignment to the genome, the primary transcripts from which these sequences originated could be easily discerned.

The monoallelic expression of *Xist* is crucial for its function. Thus, a concerted effort to identify other monoallelically expressed, noncoding transcripts might identify other RNA species serving similar roles in the nucleus. This could be done through a number of means. One method would be through the use of SNP-based approaches to assess intra-allelic expression differences for a number of genes. Such approaches will likely be useful for identifying many monoallelically expressed genes. However, many other loci may be overlooked due to the fact that even high-density SNP arrays do not currently contain enough features to provide adequate genomic coverage of the transcriptome.

Another method for looking for monoallelically expressed genes, including noncoding RNAs, would be to use rodent/human somatic cell hybrid lines. Typically, these lines are hemizygous for one human chromosome in an otherwise diploid rodent karyotype. Recently, such an approach was used to determine which X-linked genes

escape X inactivation (Carrel and Willard, 2005). In the case of the X chromosome, hybrid cell lines maintained the epigenetic state of the human X chromosome, whether it was inactive or active. Autosomal hybrid lines that maintain the replication status (early or late) of randomly asynchronously replicating genes have been created (H. Cedar, personal communication). We propose to compare the expression of human genes in cells that contain the "early" replicating copy of a given autosome to those which contain only the "late" replicating copy of that autosome. This could be done either through expression arrays or tiling arrays. Monoallelically expressed genes would be identified based on their differential expression between cells with the early replicating copy of an autosome versus those that contain the late replicating copy of the same autosome. Autosomal analogs of *Xist* would likely be expressed only in cells that contain the late replicating copy of their chromosome.

Immunoprecipitation experiments have demonstrated that *Xist* RNA associates with a number of epigenetic factors, including the histone variant, MacroH2A (Gilbert et al., 2000), BRCA1 (Ganesan et al., 2002), and unacetylated (but not acetylated) histones H3 and H4 (Ganesan et al., 2002; Gilbert et al., 2000). Autosomal analogs of Xist with similar epigenetic functions might also be associated with BRCA1, macroH2A, H3 and H4. In order to search for these RNA species, one could examine immunoprecipitated RNA with either Affymetrix expression arrays or genomic tiling arrays. Autosomally linked transcripts that were enriched in these fractions would be strong candidates for epigenetic modulators.

SUMMARY

In summary, we have observed many similarities between X-chromosome inactivation and the coordination of random asynchronous replication on mammalian autosomes. Chromosome-wide differential treatment in mammals is not limited to the X chromosome in females. Rather, chromosome-wide, stochastic choices seem to be a hallmark of mammalian chromosomes. These similarities may suggest a conserved evolutionary history between the two processes. The chromosome-wide nature of both processes may underlie some feature of mammalian chromosomes which has been co-opted for the purposes of establishing differential treatment between nearly sequence-identical alleles. Further similarities between the processes may or may not emerge. These similarities may extend to mechanistic features, such as the use of noncoding RNAs to propagate stochastic choice across long distances along the chromosome. Surely differences will emerge between the two processes. However, the development of forward genetic tools, such as RNA-interference libraries should allow for the dissection of autosomal coordination even in the absence of further parallels with X inactivation.

REFERENCES

Angelov, D., Molla, A., Perche, P. Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Mol Cell *11*, 1033-1041.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R., and Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., and Willard, H. F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527-542.

Brown, L. Y., and Brown, S. A. (2004). Alanine tracts: the expanding story of human illness and trinucleotide repeats. Trends Genet *20*, 51-58.

Carrel, L., and Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400-404.

Chadwick, B. P., and Willard, H. F. (2003). Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Hum Mol Genet *12*, 2167-2178.

Chow, J. C., Hall, L. L., Clemson, C. M., Lawrence, J. B., and Brown, C. J. (2003). Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics *82*, 309-322.

Cleary, J. D., and Pearson, C. E. (2005). Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. Trends Genet *21*, 272-280.

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol *132*, 259-275.

Cooper, D. W., VandeBerg, J. L., Sharman, G. B., and Poole, W. E. (1971). Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. Nat New Biol *230*, 155-157.

Costanzi, C., and Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Dunzinger, U., Nanda, I., Schmid, M., Haaf, T., and Zechner, U. (2005). Chicken orthologues of mammalian imprinted genes are clustered on macrochromosomes and replicate asynchronously. Trends Genet *21*, 488-492.

Ganesan, S., Silver, D. P., Greenberg, R. A., Avni, D., Drapkin, R., Miron, A., Mok, S. C., Randrianarison, V., Brodie, S., Salstrom, J., *et al.* (2002). BRCA1 supports XIST RNA concentration on the inactive X chromosome. Cell *111*, 393-405.

Gilbert, S. L., Pehrson, J. R., and Sharp, P. A. (2000). XIST RNA associates with specific regions of the inactive X chromatin. J Biol Chem *275*, 36491-36494.

Glusman, G., Yanai, I., Rubin, I., and Lancet, D. (2001). The complete human olfactory subgenome. Genome Res *11*, 685-702.

Goto, T., Wright, E., and Monk, M. (1997). Paternal X-chromosome inactivation in human trophoblastic cells. Mol Hum Reprod *3*, 77-80.

Gribnau, J., Hochedlinger, K., Hata, K., Li, E., and Jaenisch, R. (2003). Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. Genes Dev *17*, 759-773.

Gribnau, J., Luikenhuis, S., Hochedlinger, K., Monkhorst, K., and Jaenisch, R. (2005). X chromosome choice occurs independently of asynchronous replication timing. J Cell Biol *168*, 365-373.

Grutzner, F., Rens, W., Tsend-Ayush, E., El-Mogharbel, N., O'Brien, P. C., Jones, R. C., Ferguson-Smith, M. A., and Marshall Graves, J. A. (2004). In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. Nature *432*, 913-917.

Hansen, R. S., Canfield, T. K., Fjeld, A. D., and Gartler, S. M. (1996). Role of late replication timing in the silencing of X-linked genes. Hum Mol Genet *5*, 1345-1353.

Harper, M. I., Fosten, M., and Monk, M. (1982). Preferential paternal X inactivation in extraembryonic tissues of early mouse embryos. J Embryol Exp Morphol *67*, 127-135.

Harrison, K. B. (1989). X-chromosome inactivation in the human cytotrophoblast. Cytogenet Cell Genet *52*, 37-41.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., Bork, P., Burt, D. W., Groenen, M. A., Delany, M. E., *et al.* (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature *432*, 695-716.

Huynh, K. D., and Lee, J. T. (2005). X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. Nat Rev Genet *6*, 410-418.

Johnston, P. G., VandeBerg, J. L., and Sharman, G. B. (1975). Inheritance of erythrocyte glucose 6-phosphate dehydrogenase in the red-necked wallaby, Macropus rufogriseus (Desmarest), consistent with paternal X inactivation. Biochem Genet *13*, 235-242.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science *296*, 158-162.

Lee, J. T. (2003). Molecular links between X-inactivation and autosomal imprinting: X-inactivation as a driving force for the evolution of imprinting? Curr Biol *13*, R242-254.

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet *21*, 400-404.

Levine, S. S., King, I. F., and Kingston, R. E. (2004). Division of labor in polycomb group repression. Trends Biochem Sci *29*, 478-485.

Lucchesi, J. C., Kelly, W. G., and Panning, B. (2005). Chromatin Remodeling in Dosage Compensation. Annu Rev Genet.

MacDonald, M. E., Gines, S., Gusella, J. F., and Wheeler, V. C. (2003). Huntington's disease. Neuromolecular Med *4*, 7-20.

Migeon, B. R. (2003). Is Tsix repression of Xist specific to mouse? Nat Genet *33*, 337; author reply 337-338.

Migeon, B. R., and Do, T. T. (1979). In search of non-random X inactivation: studies of fetal membranes heterozygous for glucose-6-phosphate dehydrogenase. Am J Hum Genet *31*, 581-585.

Migeon, B. R., Wolf, S. F., Axelman, J., Kaslow, D. C., and Schmidt, M. (1985). Incomplete X chromosome dosage compensation in chorionic villi of human placenta. Proc Natl Acad Sci U S A *82*, 3390-3394.

Monk, M., and Harper, M. I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. Nature *281*, 311-313.

Panning, B., Dausman, J., and Jaenisch, R. (1997). X chromosome inactivation is mediated by Xist RNA stabilization. Cell *90*, 907-916.

Parkhurst, S. M., and Meneely, P. M. (1994). Sex determination and dosage compensation: lessons from flies and worms. Science *264*, 924-932.

Paulson, H. L., and Fischbeck, K. H. (1996). Trinucleotide repeats in neurogenetic disorders. Annu Rev Neurosci *19*, 79-107.

Pearson, C. E., Edamura, K. N., and Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet *6*, 729-742.

Priest, J. H., Heady, J. E., and Priest, R. E. (1967). Delayed onset of replication of human X chromosomes. J Cell Biol *35*, 483-487.

Reik, W., and Lewis, A. (2005). Co-evolution of X-chromosome inactivation and imprinting in mammals. Nat Rev Genet *6*, 403-410.

Ropers, H. H., Wolff, G., and Hitzeroth, H. W. (1978). Preferential X inactivation in human placenta membranes: is the paternal X inactive in early embryonic development of female mammals? Hum Genet *43*, 265-273.

Schmidt, M., and Migeon, B. R. (1990). Asynchronous replication of homologous loci on human active and inactive X chromosomes. Proc Natl Acad Sci U S A *87*, 3685-3689.

Skok, J. A., Brown, K. E., Azuara, V., Caparros, M. L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R. P., Merkenschlager, M., and Fisher, A. G. (2001). Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. Nat Immunol *2*, 848-854.

Takagi, N. (1974). Differentiation of X chromosomes in early female mouse embryos. Exp Cell Res *86*, 127-135.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature *256*, 640-642.

Vasques, L. R., Klockner, M. N., and Pereira, L. V. (2002). X chromosome inactivation: how human are mice? Cytogenet Genome Res *99*, 30-35.

Waters, P. D., Delbridge, M. L., Deakin, J. E., El-Mogharbel, N., Kirby, P. J., Carvalho-Silva, D. R., and Graves, J. A. (2005). Autosomal location of genes from the conserved mammalian X in the platypus (Ornithorhynchus anatinus): implications for mammalian sex chromosome evolution. Chromosome Res *13*, 401-410.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L., and Trask, B. J. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. Hum Mol Genet *11*, 535-546.

Zhou, B. B., and Elledge, S. J. (2000). The DNA damage response: putting checkpoints in perspective. Nature *408*, 433-439.

Appendix One

Coordination of the Random Asynchronous Replication of Autosomal Loci

Nandita Singh[*‡], Farah A.W. Ebrahimi[*†‡], Alexander A. Gimelbrant[*], Alexander W. Ensminger[*†], Michael R. Tackett[*†], Peimin Qi[*], Joost Gribnau[*] and Andrew Chess[*†]

*Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142
[†]Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

[‡]These authors contributed equally to this work.

This chapter was published in Nature Genetics 33(3): 339-341 (2003).

Alexander W. Ensminger performed the cell fractionation studies used to confirm the asynchronous replication of a number of loci. This work was done in collaboration with Michael Tackett and Alexander Gimelbrant.

Random monoallelic expression and asynchronous replication define an unusual class of autosomal mammalian genes. We show that every cell has randomly chosen either the maternal or paternal copy of each given autosome pair, such that alleles of these genes scattered across the chosen chromosome replicate earlier than the alleles on the homologous chromosome. Thus, chromosome pair non-equivalence, rather than being limited to X-inactivation, is a fundamental property of mouse chromosomes.

Monoallelically expressed genes fall into three distinct classes. X-inactivation in female cells is a random process resulting in half of the cells choosing the maternal X chromosome and half choosing the paternal X chromosome (Lyon, 1986). By contrast, autosomal imprinted genes such as *Igf2* and *H19* are monoallelically expressed in a parent of origin-specific manner (Efstratiadis, 1994). The third class, randomly monoallelically transcribed autosomal genes, includes the large family of odorant receptor genes (Chess et al., 1994) as well as genes encoding the immunoglobulins (Pernis et al., 1965), T cell receptors (TCRs) (Rajewsky, 1996), interleukins (Hollander et al., 1998; Rhoades et al., 2000), natural killer cell receptors (Held et al., 1995) and pheromone receptors (Rodriguez et al., 1999).

All monoallelically expressed genes examined also share the property of asynchronous replication (Chess et al., 1994; Hollander et al., 1998; Rodriguez et al., 1999) defined as one allele replicating earlier in S phase than the other allele. For most other genes, both alleles replicate synchronously at a defined portion of S phase. Asynchronous replication represents an epigenetic mark that appears prior to transcription and may underlie the differential behavior of two alleles of identical

sequence (Mostoslavsky et al., 2001). For those genes whose transcription is randomly monoallelic, the asynchronous replication is also random. The asynchronous replication appears to be set up early in development prior to the time when tissue-specific transcription is established (Mostoslavsky et al., 2001; Simon et al., 2001) and is therefore found even in tissues in which the genes are not expressed (Chess et al., 1994). For example, the asynchronous replication of odorant receptor genes has been observed in all cell types analyzed including fibroblasts and lymphocytes. The presence of asynchronous replication in a variety of cell types allowed us to compare the replication timing of diverse monoallelically expressed genes that are expressed in different cells. Given that these genes are widely dispersed across autosomes (Young et al., 2002; Zhang and Firestein, 2002), we sought to establish the extent to which their replication asynchrony is coordinated. We focused on four autosome pairs, each containing distinct loci of randomly monoallelically expressed genes (Fig. 1*a*).

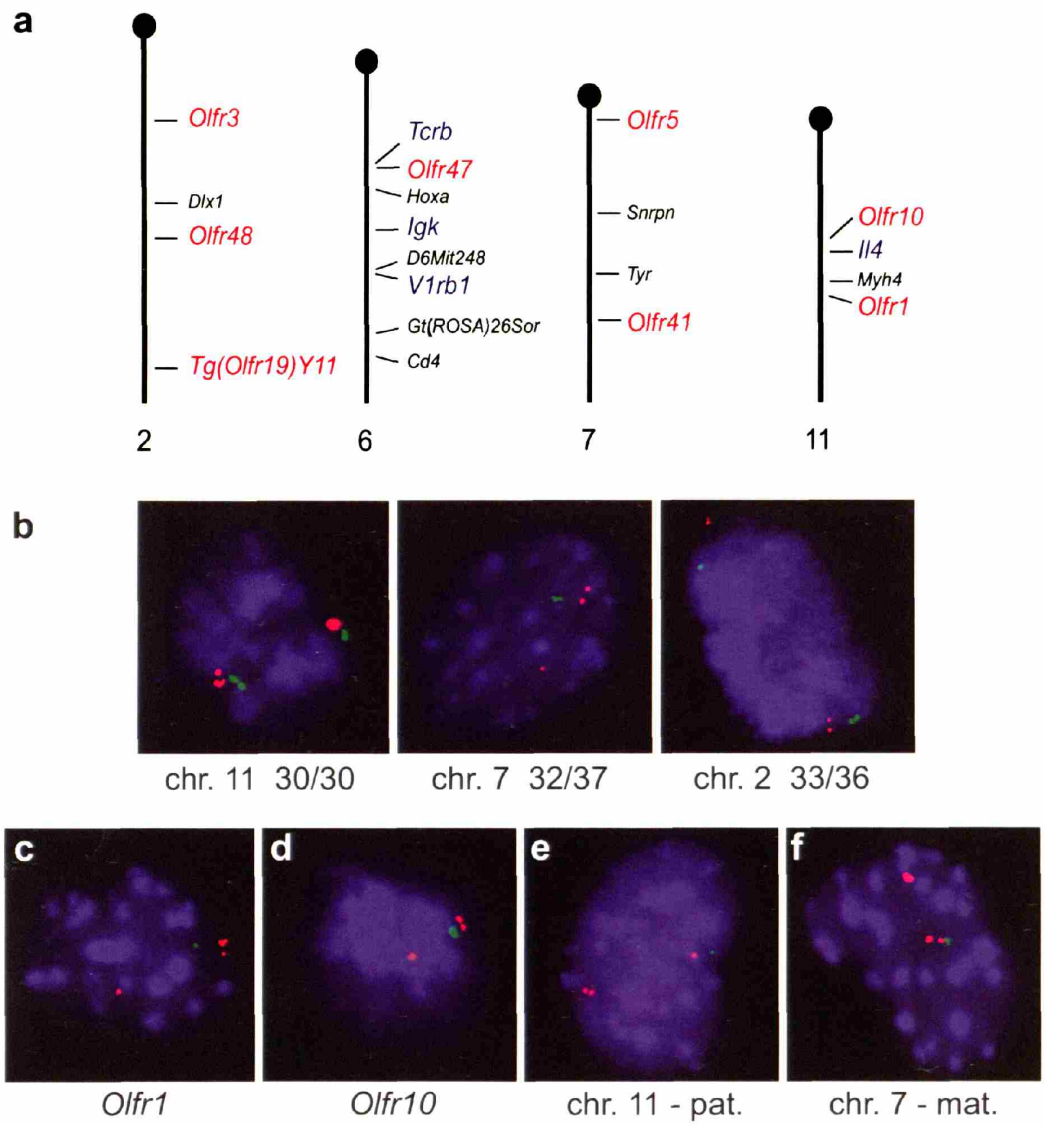
Asynchronous replication can be assayed by fluorescence *in situ* hybridization (FISH) analysis of interphase nuclei (Selig et al., 1992). Replicated loci are visualized as a double-dot hybridization signal, while unreplicated loci reveal a single dot. Asynchronously replicating genes reveal a single dot-double dot (SD) pattern in 30-40% of S phase cells as opposed to around 10-15% SD for synchronously replicating genes (Selig et al., 1992). While the FISH assay is only an indirect assessor of replication timing, asynchronous replication observed with this assay has been corroborated by direct measurements of replication timing (Mostoslavsky et al., 2001; Simon et al., 2001) (supplementary Fig. A). To assess coordination of distant loci on a given chromosome, we performed two-color FISH examining two genes at once and scored cells that reveal a

Figure 1

Coordination of odorant receptor asynchronous replication for individual mouse chromosome pairs. **(a)** Diagram indicating the relative positions of odorant receptor genes (red) and other monoallelically expressed genes analyzed in this study (blue) along with the location of control genes (black). Centromeric ends are at the top. **(b)** Two-color FISH was performed on a population of mouse embryonic fibroblasts. For all the images, blue is from DAPI staining of chromatin. Left panel: Analysis of chromosome 11. The Cy3 labeled (red) probe identifies the *Olfr1* odorant receptor gene and the FITC-labeled (green) probe identifies the *Olfr10* gene. The double-dot signals for the two probes in these images are on the same chromosome indicating coordination of these two distant loci (30/30 cells counted). A control probe *Myh4* located between the *Olfr1*and *Olfr10* loci, is synchronously replicating (9% SD). Center panel: Similar analysis of chromosome 7 for two odorant receptor genes from distinct clusters: *Olfr5* gene (red) and the *Olfr41* gene (green). Coordination is observed (32/37 cells counted). Control probes between the *Olfr5* and *Olfr41* loci included the tyrosinase gene, which in wild-type cells is synchronously replicated (13% SD), and the asynchronous, but imprinted *Snrpn* gene. As expected, the imprinted *Snrpn* gene, when compared to an odorant receptor reveals a lack of coordination (not shown). Right panel: Similar analysis demonstrating that two odorant receptor genes from distinct clusters of chromosome 2 are coordinated (33/36): *Olfr48* (red), *Olfr3* (green). A control probe *Dlx1* is synchronously replicating (10% SD). **(c and d)** FISH analyses of line A.5 detected with a β -geo probe (green) that identifies the maternal chromosome 11. Examples of cells probed with the *Olfr1* odorant receptor gene (red, C) and the *Olfr10* odorant receptor gene (red, D) are shown. Both of these

odorant receptor genes are maternal early replicating in line A.5. **(e and f)** Lack of coordination between chromosomes 7 and 11 is demonstrated. In each case, the maternal chromosome is marked by the green probe. The maternal chromosome 11 has a β-geo insertion and the paternal chromosome 7 has a deletion at the tyrosinase gene site. Line A.1 reveals lack of coordination between the paternally early replicating *Olfr1* odorant receptor gene on chromosome 11 (panel E) and the maternally early replicating *Olfr41* odorant receptor gene on chromosome 7 (panel F).

FIGURE 1


a
Olfr3
Dlx1
Olfr48
Tg(Olfr19)Y11
2
Tcrb
Olfr47
Hoxa
Igk
D6Mit248
V1rb1
Gt(ROSA)26Sor
Cd4
6
Olfr5
Snrpn
Tyr
Olfr41
7
Olfr10
Il4
Myh4
Olfr1
11
b
chr. 11 30/30
chr. 7 32/37
chr. 2 33/36
c
d
e
f
Olfr1
Olfr10
chr. 11 - pat.
chr. 7 - mat.

SD signal for both genes. If the two genes are coordinated, and are replicated during an overlapping portion of S phase, the expectation would be that in cells SD for both genes the double dots for both genes will be on the same chromosome (maternal or paternal) and therefore close to each other in the nucleus. If the two genes are not coordinated, one expects only 50% of the cells with SD for both probes to appear coordinated.

Using this approach we assessed the potential for coordination of asynchronous replication in wild-type primary mouse embryonic fibroblasts (PMEFs), analyzing two distinct odorant receptor loci on chromosome 11 that are 14 centimorgans (cM) apart. Strikingly, we observe coordination in 30/30 cells in which both probes gave the SD pattern (Fig. 1*b*). Similarly we observe coordination for two distant loci on chromosome 7 (32/37 cells) and for two distant loci on chromosome 2 (33/36 cells) (Fig. 1*b*). As expected, in each case genes between the distinct odorant receptor loci reveal synchronous replication. These data indicate that odorant receptor genes have long-range coordination of their replication asynchrony for the three autosomes examined.

We next examined whether coordinated asynchronous replication of odorant receptor genes, once established, is heritable in the progeny of a given cell. Clonal cell lines were derived from embryonic and adult animals with distinguishable maternal and paternal chromosomes for chromosomes 7 and 11. For chromosome 7, we analyzed 7 cell lines and for chromosome 11, we analyzed 8 cell lines. In some cell lines we consistently observe early replication of the maternal allele (Fig. 1*c*) and in the other cell lines we consistently observe early replication of the paternal allele (supplementary Tables A, B). These analyses indicate that for each odorant receptor gene, the random choice of one of the two alleles to replicate early, once established, is heritable. Analyses

of these clonal cell lines also demonstrate coordination of asynchronous replication along a given chromosome. For all the cell lines analyzed both loci replicate the same parental allele early (Fig. 1*c*,*d* and supplementary Tables A, B). These analyses of clonal cell lines thus confirm coordination along given chromosomes.

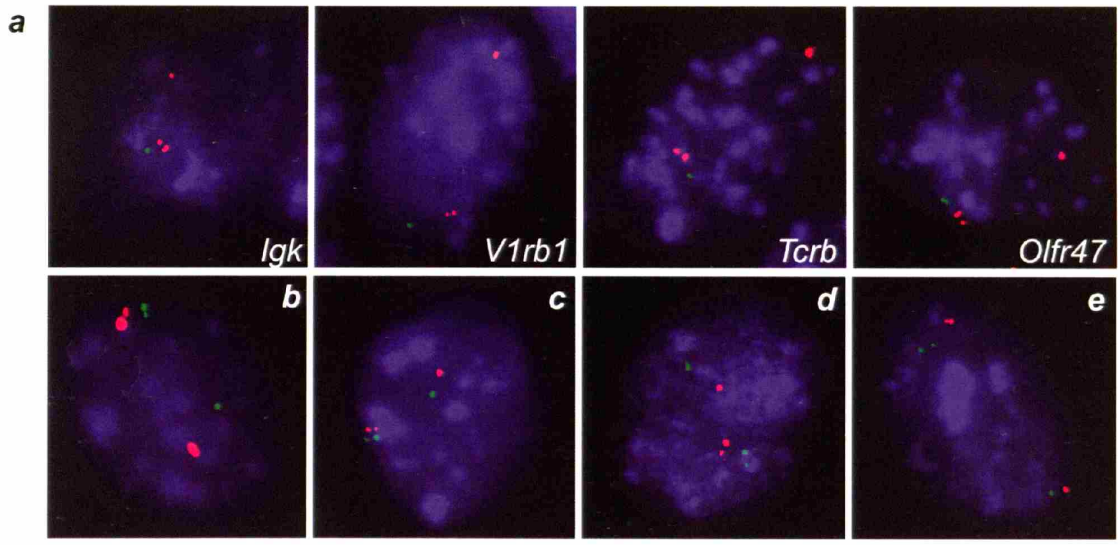
To test for genome-wide coordination, we analyzed clonal cell lines derived from animals carrying marks allowing us to distinguish the parental origins of two chromosome pairs at a time. We used FISH to compare the replication timing of odorant receptor genes on chromosomes 6 and 11 with odorant receptor genes on chromosome 7 in pair wise comparisons. The data show that odorant receptor genes present on different chromosomes are not coordinated in their replicative asynchrony; all possible outcomes were observed (Fig. 1*e*,*f* and supplementary Table C). Thus, rather than genome wide coordination, there appears to be only a chromosome pair by chromosome pair coordination of odorant receptor gene asynchronous replication.

To explore whether other randomly monoallelically transcribed genes are also coordinated in their replicative asynchrony, we examined clonal cell lines in which we can distinguish the maternal and paternal copies of chromosome 6. Figure 2*a* shows asynchronous replication of the *Igk* constant region, a V1R pheromone receptor gene and *Tcrb*. Strikingly, all three of these loci are coordinated with each other as well as with an odorant receptor gene cluster on chromosome 6 (Fig. 2*a*). Some clonal cell lines have the maternal alleles of all four genes replicating early (Fig. 2*a*). In the other clonal cell lines, the paternal alleles of all four genes replicate early (see supplementary Fig. B). Analyses of uncloned populations of cells (similar to the analyses presented in Fig. 1*b*) also demonstrate coordination of genes on chromosome 6 (Fig. 2*b*,*c*). Similar population

Figure 2

A variety of monoallelically expressed genes reveal coordination: analyses of chromosomes 6 and 11. **(a)** Analyses of a clonal cell line that has a marked maternal copy of chromosome 6 (β-geo transgene, green). Probes for the *Igk* cluster, a large *V1R* pheromone receptor cluster (VNO-61), the *Tcrb* and an odorant receptor cluster containing *Olfr47* are shown (red). Line F.1 has all four genes maternally early replicating. Line C.1 has all four paternally early replicating (see supplementary information). Two control probes, *Hoxa* and *Cd4* each revealed synchronous replication (12%SD and 18%SD respectively). **(b)** Analyses of uncloned fibroblasts (similar to the analyses shown in Fig. 1B) also show coordination of a chromosome 6 odorant receptor gene cluster containing *Olfr47* (red) and the VNO *V1R* cluster (green) (30/32 cells). **(c)** Same analysis as in (B), but comparing to *Tcrb* (33/34 cells). **(d)** A similar population analysis reveals coordination of the *Il4* gene (red) and the *Olfr10* odorant receptor gene (green) on chromosome 11 (25/26 cells). **(e)** Example of coordination of the *Olfr48* endogenous odorant receptor gene (green) and a transgenic *Olfr19* odorant receptor gene (*Tg(Olfr19)Y11*, red) on chromosome 2 (30/31 cells, $P<0.0000001$).

FIGURE 2



analyses demonstrate that on chromosome 11, the interleukin-4 (*Il4*) gene is coordinated with the odorant receptor genes (Fig. 2*d*). These data, taken together, allow the remarkable conclusion that all randomly asynchronously replicated genes examined are coordinated along each given chromosome.

An interesting question arising from these observations is whether the mechanisms employed by different chromosomes are capable of communicating with each other if sequences from different chromosomes are artificially placed in *cis*. We have analyzed a small odorant receptor translocation that we created artificially: a 300kb odorant receptor-containing YAC transgene derived from chromosome 16 that is integrated on chromosome 2 (*Tg(Olfr19)Y11*). Previously we demonstrated that this transgenic odorant receptor locus undergoes asynchronous replication (Ebrahimi et al., 2000). Two-color FISH demonstrates that the transgene is coordinated in its asynchronous replication with the endogenous odorant receptor gene loci on chromosome 2 (Fig. 2*e*) suggesting similarities in the mechanisms governing allele-specific replication timing on different chromosomes.

Here, we present data indicating that randomly monoallelically expressed genes have a chromosome pair by chromosome pair coordination of their asynchronous replication. Scattered genes along a given chromosome are coordinated in their asynchronous replication timing in *cis*, leaving unaffected the bulk of the genes (which are synchronously replicated or in rare instances asynchronous, but imprinted). Asynchronous replication is established early in development (Mostoslavsky et al., 2001; Simon et al., 2001) and maintained in the progeny of individual cells in a clonal manner (Fig. 1*c-f* and supplementary Tables A, B) (Mostoslavsky et al., 2001). Randomly

monoallelically expressed genes are expressed in different cells of a given cell type or in different cell types entirely. Therefore, the coordination of replication timing we present does not imply coordination of transcription of these distinct gene families and is instead a consequence of the early developmental mechanisms establishing the asynchronous replication. Each gene family likely makes use of the asynchronous replication, and the chromatin structure differences it reflects, in the complex gene regulation that characterizes these gene families. In the case of the immunoglobulin genes, we have recently shown that early replication correlates with the allele that will first undergo rearrangement and therefore provides a basis for the establishment of allelic exclusion (Mostoslavsky et al., 2001).

X-inactivation has been known for decades. Remarkably, our data indicate that chromosome pair non-equivalence is also found on autosomes and thus is a general, fundamental property of chromosomes affecting a large number of loci dispersed throughout the genome. The autosomal non-equivalence we observe is similar to what is observed with X-inactivation except that a larger fraction of the genes on the X chromosome are affected. It is possible that other similarities in the respective underlying mechanisms of X-inactivation and autosomal non-equivalence will emerge.

ACKNOWLEDGEMENTS

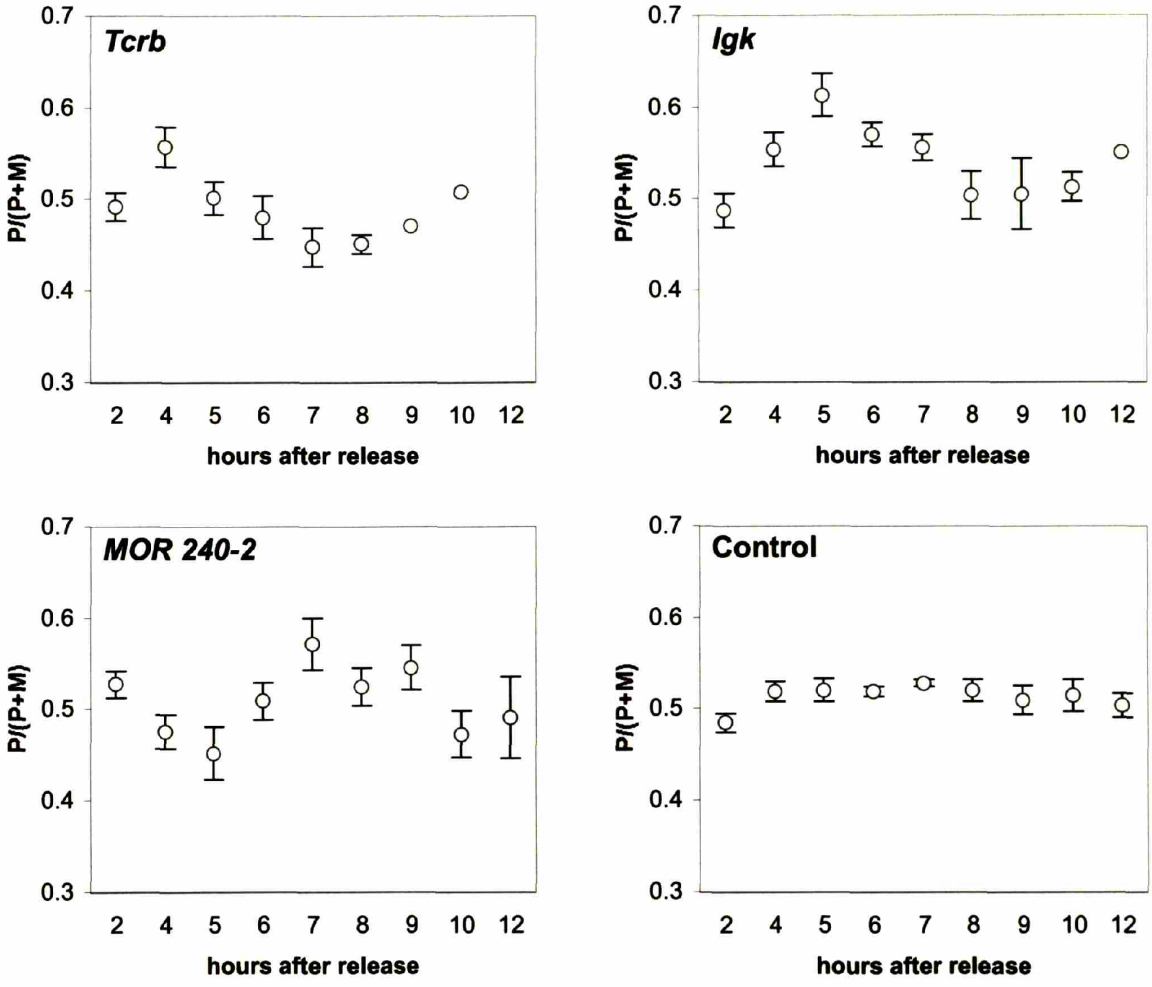
Supported by grants from the NIH (NIDCD) to A.C., F.A.W.E., and M.T.. A.W. Ensminger is a Howard Hughes Medical Institute Predoctoral Fellow. We thank H. Cedar, G. Fink, D. Housman, R. Jaenisch, D. Page, C. Cowles, and members of the Chess laboratory for discussions and comments on the manuscript; H. Higgins for manuscript preparation; G. Paradis for help with FACS analyses; J. Young for chromosome 6 BACs, D. Littman for the Cd4 cosmid, R. Lane for the Olfr41 BAC, B. Holdener for the tyrosinase deletion mice, A. Dunn for the *CSF2* knock-in mice, P. Soriano for the ß-geo transgenic mice, N. Rosenberg for the Abelson leukemia virus producing lines, D. MacAlpine for advice on immunoprecipitation, S. Gabriel, B. Blumenstiel, M. DeFelice and E. Winchester for help with MALDI-TOF genotyping, and S. Rozen for help with sequencing.

SUPPLEMENTARY INFORMATION

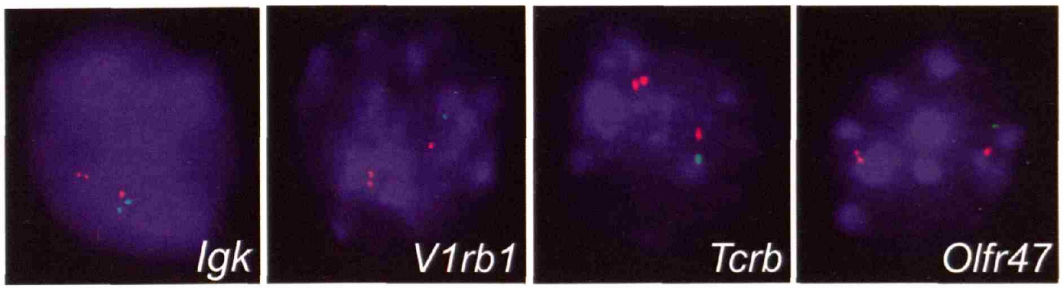
S-phase fractionation analyses of asynchronous replication:

The FISH approach has been previously correlated with direct measures of replication timing for *Igk* (Mostoslavsky et al., 2001) and for imprinted genes (Simon et al., 1999). To directly assess allele-specific replication timing for an odorant receptor gene and *Tcrb*, we have employed an S-phase fractionation approach. Clonal ES cell lines from *M.m domesticus* X *M. m. castaneous* F1 mice were separated into different portions of S phase following release from G1 arrest (Krude, 1999). Cells at each timepoint were harvested after a pulse of bromodeoxyuridine (BrdU) allowing isolation of replicating DNA. We then used PCR followed by single nucleotide primer extension (Cowles et al., 2002) to examine the relative amounts of DNA from the two alleles. Asynchronous replication of the *Tcrb* locus and the kappa immunoglobulin (Igk) locus in line ES51 is shown in Supplemental Figure A. Both genes are paternally early replicating in the ES51 line. As with the FISH experiments, we consistently observe coordination of the replication asynchrony direction of these two genes in clonal cell lines. The *MOR240-2* odorant receptor gene on chromosome 11 is maternally early replicating in ES51 (Supplemental Fig. A).

SUPPLEMENTAL FIGURE A:



SUPPLEMENTAL FIGURE B:



METHODS

Mice and cells

Coordination analyses were performed on PMEFs derived from wild type *Mus musculus* embryos (E12.5) and embryos homozygous for a 300 kb odorant receptor-containing yeast artificial chromosome transgene (Ebrahimi et al., 2000). Clonal cell lines were generated from the following mice: mice with a maternally inherited LacZ-containing transgene on chromosome 6 (Mostoslavsky et al., 2001); mice with a maternally inherited LacZ-containing transgene on chromosome 11 (Stanley et al., 1994); and mice with a paternally inherited deletion on chromosome 7 (Chess et al., 1994). Clonal fibroblast cell lines were also generated from embryos inheriting the paternally inherited deletion on chromosome 7 and either the maternally inherited transgene on chromosome 11 or the maternally inherited transgene on chromosome 6 mentioned above. PMEFs at E12.5 (Csankovszki et al., 2001) or adult primary ear fibroblasts (Mostoslavsky et al., 2001) were immortalized via infection with a retroviral vector encoding the SV40 large-T antigen (Csankovszki et al., 2001). The cells were subcloned by limiting dilution and to check for aneuploidy, the subclones were compared to the untransformed cells in a FACS analysis with propidium iodide staining (Hansen et al., 1993). Clonal cell lines that were diploid by FACS analysis were further analyzed by FISH for a variety of chromosomes. Clonal Abelson cell lines were generated by transforming fetal liver (E12.5) cells or adult bone marrow cells from mice with a paternally inherited deletion on chromosome 7 with Abelson leukemia virus (Rosenberg and Baltimore, 1978) and subsequent limiting dilution. As aneuploidy in Abelson

transformed clones is rarer, these cells were checked only by FISH with probes to a number of chromosomes.

FISH

BACs representing *Il4*, *Hoxa*, *Dlx1* and olfactory receptors *Olfr48*, *Olfr3*, *Olfr5*, *Olfr1* and *Olfr10* (ranging in size from 80-200 kb) were identified in a 129/SVJ mouse BAC library purchased from Incyte Genomics (154F05, 4C06, 199N02, 193J05, 58N20, 23C08, 83E20, 50J07 respectively). The BACs were identified using PCR primers whose sequences were obtained from GENBANK. Identities of these BACs were confirmed by PCR and sequence analyses. The following mouse probes were a gift or purchased: pSPIgB, plasmid carrying the 12kb BamH1 fragment encompassing the J-Ck region (Y. Bergman); pSAbgeofrtpA, plasmid carrying a 5.6kb Xho-1 fragment carrying the bgalneo genes (L. Jackson-Grusby); Cos1, 37.8kb cosmid carrying the *Cd4* gene (D. Littman); 71-1b, cosmid carrying Cb1 region in the *Tcrb* locus (J. Chen); 22B05, BAC carrying the *Olfr41* OR gene (R. Lane); 85M02, BAC with the deleted region on chromosome 7 (ResGen); 168014, BAC from the OR locus on chromosome 6 (J. Young) and 380N16, PAC carrying the *Myh4* gene (Roswell Park Cancer Institute). These probes were all tested to ensure that they yielded a single discrete hybridization signal for each unreplicated allele. For probe preparation, double stranded DNA was labeled by Cy3, FITC or digoxigenin nucleotides by nick translation as described previously (Panning and Jaenisch, 1996). Labeled probes were precipitated with salmon sperm DNA, tRNA and mouse COT-1 DNA. Probe was washed extensively with 100% ethanol and resuspended in hybrisol VII (Ventana). Probe mixture was heated to 65°C for 10

min. and then incubated at 37°C for 30 min. Cells were pulse labeled with BrdU/FdU (AP Biotech) for 30 min. and fixed in 3:1 methanol : acetic acid, dropped on poly-L-lysine coated slides (Sigma) in a humid chamber and denatured for 3 min. at 70°C in 70% formamide, 2X SSC. BrdU labeling was used to identify cells in S phase in experiments measuring the percentage of SD cells for a given probe and was visualized using a mouse antibody against BrdU (Anti-BrdU Pure, Becton Dickinson) in concert with a fluoroscein-conjugated anti-mouse IgG (Vector Laboratories). Hybridizations were carried out in a humid chamber at 37°C overnight. The slides with direct label probes were washed 3 times with 50% formamide, 2 X SSC at 42°C followed by 3 washes with 2X SSC also at 42°C. The following washes were done at room temperature for 10 min. with 1 X SSC, 5 min. with 4X SSC + 0.1% Tween 20 and 5 min. with 4X SSC. The slides were dehydrated in cold 80% ethanol, 95% ethanol and 100% ethanol. The indirect label (digoxigenin) was detected as described earlier (Gribnau et al., 1998). It was detected with the sheep anti-dig antibody (Roche Diagnostics) and the signal was amplified using a FITC conjugated rabbit anti-sheep antibody (Calbiochem) and a FITC conjugated goat anti-rabbit antibody (Roche Diagnostics). For the coordination analysis, more often than not, when we observed the SD pattern for one probe, we also were able to observe the SD pattern for the other probe. We counted only the cells where the S and D signals for the first probe were clearly separated. Also the signals for the second probe were similarly separated and each had a clear association with either the S or the D signal of the first probe. The cells that show non-coordination are likely to be cells where the hybridization signals that are near each other and appear to be linked are actually unlinked. The extent of non-coordination correlates with the distance between the

probes. Also, even for very tightly linked genes, one would expect some cells that seem to be SD are really DD, and one of the replicated alleles is not discernable as a double-dot. Also note that as one would expect, the rare SDs from a synchronously replicating gene are half maternal and half paternal early. For the analysis with the clonal cell lines, we counted only the cells where the S and D signals were clearly separated and there was a close association of either the S or the D signal and the probe used to mark the maternal chromosome.

Cell synchronization and DNA immunoprecipitation

Clonal mouse ES cell lines (129 x CAST) were generated by selection of neomycin resistant subclones after transfection with neo containing vector. Each clone was synchronized by first arresting the cells in late G1 using a 14 hour exposure to 0.5 mM mimosine (Krude, 1999). Subsequently, the mimosine was removed to release the cells from arrest. Cells were labeled with BrdU one hour prior to harvesting by trypsinization (at 2, 4, 5, 6, 7, 8, 9, 10, and 12 hours). DNA was extracted from Proteinase K-treated cells using standard phenol:chloroform extraction methods and immunoprecipitated as previously described (Ren et al., 2000) using a monoclonal mouse anti-BrdU antibody (PharMingen) together with Dynabeads M-280 Sheep anti-Mouse IgG (Dynal).

SNP-based determination of allelic content

In order to distinguish the parental alleles, single nucleotide polymorphisms (SNPs) were identified by sequencing loci of interest using genomic DNA from CAST/Ei

and 129/SvJ (Jackson Labs) as template. In the DNA prepared from each S-phase fraction, regions containing SNPs were PCR amplified. The relative amounts of the two parental alleles were determined by primer extension of PCR products. Detection was by matrix-assisted laser desorption/ionisation - time-of-flight (MALDI-TOF) mass spectrometry (Cowles et al., 2002; Tang et al., 1999). In each individual reaction, the ratio of peak heights corresponding to the two alleles was calibrated to a series of mixes with known composition of genomic DNA from parental strains.

Supplementary Table A. Coordination of two odorant receptor loci on chromosome 11.

Cell Line	*Olfr1*	*Olfr10*
A.1	P (90%, n=40)	P (86%, n=36)
A.2	M (94%, n=32)	M (95%, n=21)
A.3	P (94%, n=32)	P (88%, n=17)
A.4	P (90%, n=40)	P (91%, n=32)
A.5	M (91%, n=33)	M (88%, n=34)
A.6	M (94%, n=32)	M (88%, n=33)
A.7	M (90%, n=39)	M (91%, n=33)
A.8	P (93%, n=46)	P (90%, n=40)

Eight clonal mouse embryonic fibroblast lines were derived from a single mouse embryo (embryo A) heterozygous for an insertion of the *lacZ* gene in the *CSF2* locus[18] (and with distinguishable copies of chromosome 7 due to a paternally inherited tyrosinase gene deletion). FISH was performed using two probes, one to the *lacZ* gene (marks the maternal chromosome 11) and another probe to either the *Olfr1* or *Olfr10* odorant receptor clusters. The letter P or M denotes whether the paternal or maternal allele was early replicating. The number of cells assayed and the percentage of cells in the indicated orientation are also noted. These analyses reveal that the replication of two distinct odorant receptor loci on chromosome 11 are coordinated. Note that, as one would expect, the rare SDs from a synchronously replicating gene are approximately 50% maternal and 50% paternal.

Supplementary Table B. Coordination of two odorant receptor loci on chromosome 7.

Cell Line	*Olfr41*	*Olfr5*
B.1	M (94%, n=16)	M (81%, n=16)
B.2	P (95%, n=20)	P (78%, n=40)
B.3	M (100%, n=13)	M (80%, n=35)
C.1	P (79%, n=28)	P (83%, n=35)
D.1	M (92%, n=38)	M (81%, n=32)
E.1	M (90%, n=31)	M (88%, n=41)
F.1	P (89%, n=45)	P (90%, n=41)

Clonal cell lines were derived from five mice (B, C, D, G and F) each of which is heterozygous for a paternal deletion of the tyrosinase gene on chromosome 7. Thus a probe to the tyrosinase gene identifies the maternal allele. Lines B.1, B.2, B.3, E.1 and F.1 are fibroblasts derived from adult ear tissue. Lines C.1 and D.1 are Abelson transformed B lymphocytes. As in Supplementary Table A, the letter P or M denotes whether the paternal or maternal allele was early replicating and the number of cells assayed and the percentage of cells in the indicated orientation is also noted. (Animals E and F are also heterozygous for an insertion of the ß-geo gene into the maternal *Gt(ROSA)26Sor* locus on chromosome 6.) Analyses of these cell lines reveal coordination on chromosome 7.

Supplementary Table C. Lack of coordination between chromosomes: comparing chromosome 7 with chromosomes 11 and 6.

Cell Line	Ch 7	Other (6 or 11)
A.1	M (90%, n=40)	P (90%, n=40)
A.2	M (91%, n=33)	M (94%, n=32)
A.3	M (91%, n=33)	P (94%, n=32)
A.4	P (92%, n=37)	P (90%, n=40)
A.5	P (87%, n=39)	M (91%, n=33)
A.6	P (92%, n=39)	M (94%, n=32)
A.7	M (88%, n=34)	M (90%, n=39)
E.1	M (90%, n=31)	M (91%, n=35)
F.1	P (89%, n=45)	M (98%, n=41)

Lines A.1-A.7 allow comparisons of *Olfr41* on chromosome 7 and *Olfr1* on chromosome 11. The maternal copy of chromosome 7 is identifiable because of the paternal deletion of the tyrosinase gene and the maternal copy of chromosome 11 is identifiable because of the maternal insertion of the *lacZ* gene in the *CSF2* locus (Stanley et al., 1994). As in Supplementary Tables A and B, the letter P or M denotes whether the paternal or maternal allele was early replicating and the number of cells assayed and the percentage of cells in the indicated orientation is also noted. Lines E.1 and F.1 allow a comparison of *Olfr41* on chromosome 7 and a gene on chromosome 6; chromosome 6 is identifiable in these clonal cell lines because animals E and F are heterozygous for an insertion of the ß-geo gene into the maternal *Gt(ROSA)26Sor* locus (Soriano, 1999) and as mentioned before, the maternal copy of chromosome 7 is identifiable because of the paternal deletion of the tyrosinase gene. A lack of coordination between different chromosomes is apparent.

REFERENCES

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834.

Cowles, C. R., Hirschhorn, J. N., Altshuler, D., and Lander, E. S. (2002). Detection of regulatory variation in mouse genes. Nat Genet *32*, 432-437.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J Cell Biol *153*, 773-784.

Ebrahimi, F. A., Edmondson, J., Rothstein, R., and Chess, A. (2000). YAC transgene-mediated olfactory receptor gene choice. Dev Dyn *217*, 225-231.

Efstratiadis, A. (1994). Parental imprinting of autosomal mammalian genes. Curr Opin Genet Dev *4*, 265-280.

Gribnau, J., de Boer, E., Trimborn, T., Wijgerde, M., Milot, E., Grosveld, F., and Fraser, P. (1998). Chromatin interaction mechanism of transcriptional control in vivo. Embo J *17*, 6020-6027.

Hansen, R. S., Canfield, T. K., Lamb, M. M., Gartler, S. M., and Laird, C. D. (1993). Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell *73*, 1403-1409.

Held, W., Roland, J., and Raulet, D. H. (1995). Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. Nature *376*, 355-358.

Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998). Monoallelic expression of the interleukin-2 locus. Science *279*, 2118-2121.

Krude, T. (1999). Mimosine arrests proliferating human cells before onset of DNA replication in a dose-dependent manner. Exp Cell Res *247*, 148-159.

Lyon, M. F. (1986). X chromosomes and dosage compensation. Nature *320*, 313.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221-225.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. J Exp Med *122*, 853-876.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. Nature *381*, 751-758.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306-2309.

Rhoades, K. L., Singh, N., Simon, I., Glidden, B., Cedar, H., and Chess, A. (2000). Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. Curr Biol *10*, 789-792.

Rodriguez, I., Feinstein, P., and Mombaerts, P. (1999). Variable patterns of axonal projections of sensory neurons in the mouse vomeronasal system. Cell *97*, 199-208.

Rosenberg, N., and Baltimore, D. (1978). The effect of helper virus on Abelson virus-induced transformation of lymphoid cells. J Exp Med *147*, 1126-1141.

Selig, S., Okumura, K., Ward, D. C., and Cedar, H. (1992). Delineation of DNA replication time zones by fluorescence in situ hybridization. Embo J *11*, 1217-1225.

Simon, I., Tenzen, T., Mostoslavsky, R., Fibach, E., Lande, L., Milot, E., Gribnau, J., Grosveld, F., Fraser, P., and Cedar, H. (2001). Developmental regulation of DNA replication timing at the human beta globin locus. Embo J *20*, 6150-6157.

Simon, I., Tenzen, T., Reubinoff, B. E., Hillman, D., McCarrey, J. R., and Cedar, H. (1999). Asynchronous replication of imprinted genes is established in the gametes and maintained during development. Nature *401*, 929-932.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70-71.

Stanley, E., Lieschke, G. J., Grail, D., Metcalf, D., Hodgson, G., Gall, J. A., Maher, D. W., Cebon, J., Sinickas, V., and Dunn, A. R. (1994). Granulocyte/macrophage colony-

stimulating factor-deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology. Proc Natl Acad Sci U S A *91*, 5592-5596.

Tang, K., Fu, D. J., Julien, D., Braun, A., Cantor, C. R., and Koster, H. (1999). Chip-based genotyping by mass spectrometry. Proc Natl Acad Sci U S A *96*, 10016-10020.

Young, J. M., Friedman, C., Williams, E. M., Ross, J. A., Tonnes-Priddy, L., and Trask, B. J. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. Hum Mol Genet *11*, 535-546.

Zhang, X., and Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. Nat Neurosci *5*, 124-133.

Appendix Two

Using RNase-treatment to examine the structural details of RNA transcripts within chromatin

In order to aid in our attempts to determine whether autosomal analogs of *Xist* are involved in autosomal coordination, I performed a set of pilot experiments designed to further characterize the *Xist* transcript. One historically under explored characteristic of *Xist* is that it is highly resistant to RNase digestion (J. Gribnau, personal communication), presumably due to either its association with inactive chromatin and/or protein complexes. This has been noted in the process of performing RNA FISH to *Xist*. One common negative control for background in RNA FISH is to treat slides with RNase A prior to hybridization in order to disrupt signal. However, RNA FISH controls for *Xist* signal involved the exposure of slides to RNase H post-hybridization, presumable to degrade RNA-DNA hybrids (Clemson et al., 1996).

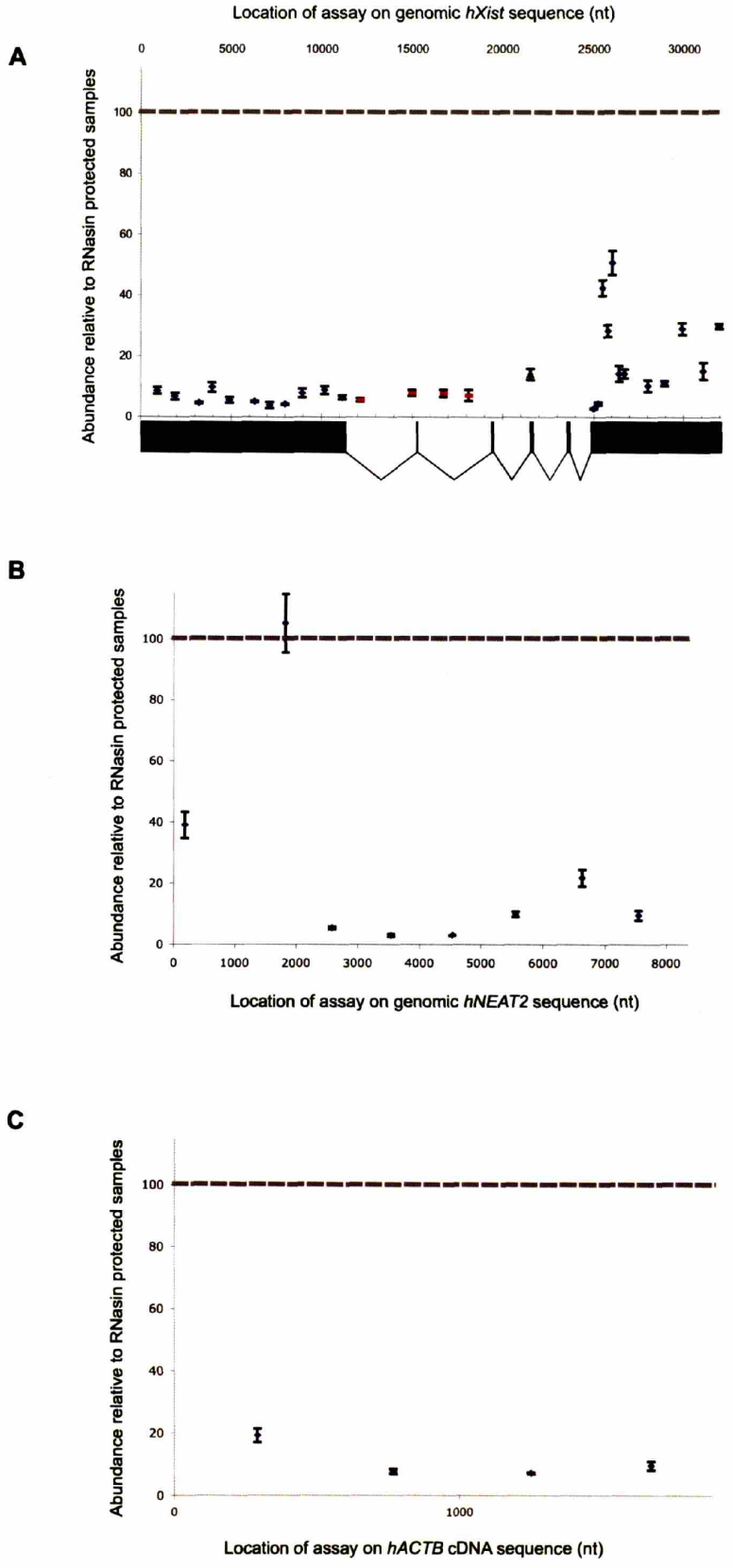
In control experiments I developed to further enrich for *Xist* in a population of nuclear RNAs, we began to explore the possibility of exposing chromatin to RNase in an attempt to enrich for *Xist* relative to introns and other superfluous nuclear RNA species. While such approaches appeared to enrich for *Xist* relative to other RNAs, they also led me to ask whether particular segments of the *Xist* RNA are more or less sensitive to such a treatment.

Through the use of real-time PCR analysis, we compared relative levels of specific segments of *Xist* in random-primed cDNA derived from both RNase-treated and non-treated human chromatin preparations. Primers have been designed across the entire length of the human *Xist* spliced transcript, with each assay within a 1 kb interval along the cDNA of other primer pairs. Strikingly, specific regions of the *Xist* transcript appear to be quite protected from RNase-digestion, whereas other regions appear indiscernible from intronic sequences (Figure 1a). Such protected sequences do not appear to simply

Figure 1.

Relative levels of specific regions of RNA transcripts after treatment with RNase A and T1. **(A)** Chromatin prepped from human female lymphoblasts (GM00131) was exposed to RNase A and T1 cocktail (see materials and methods). After reverse transcription using random decamer primers, the population of cDNAs from treated and untreated samples were compared. Real-time PCR primers designed every kilobase along the *hXist* transcript identifies a region of relative insensitivity to RNase A and T1 cleavage near the beginning of exon 6. The same region has previously been described to selectively immunoprecipitated with anti-BRCA1 antibodies (Ganesan et al., 2002) when compared to other regions of the *hXist transcript.* This may reflect a functional role for this region of the *hXist* transcript, however our results also raise the possibility that the selective enrichment of this region in immunoprecipitation may result from degradation of the transcript during the course of the experiment. **(B)** The conserved noncoding RNA *hNEAT2* is also protected from RNase degradation in chromatin preps. A 5' region shows very little, if any degradation relative to RNasin-protected controls. **(C)** The human beta actin (*hACTB*) transcript shows very little protection from RNase digestion. Other open reading frame transcripts should be examined to determine whether RNase protection is a property unique to noncoding transcripts.

FIGURE 1:


A
Location of assay on genomic *hXist* sequence (nt)
0
5000
10000
15000
20000
25000
30000
Abundance relative to RNasin protected samples
100
80
60
40
20
0
B
Abundance relative to RNasin protected samples
100
80
60
40
20
0
0
1000
2000
3000
4000
5000
6000
7000
8000
Location of assay on genomic *hNEAT2* sequence (nt)
C
Abundance relative to RNasin protected samples
100
80
60
40
20
0
0
1000
Location of assay on *hACTB* cDNA sequence (nt)

correlate with some of the well-described highly conserved secondary structures within the *Xist* transcript. Directing such an approach to one of the nuclear noncoding RNAs identified in our earlier screen, *hNEAT2*, we identified regions that appear to be even more protected from RNase (Figure 1b). Extension of these analyses to *hNEAT1* were complicated by earlier studies which indicated the presence of antisense transcription surrounding *hNEAT1* in somatic human cells. However, careful real-time PCR analysis (see Chapter 3), has indicated that these early results were likely erroneous, opening the way for the analsis of both *hNEAT* transcripts. The localized protection from RNase A and T1 observed in *Xist* and *hNEAT2* is unique from that observed within the human beta actin (hACTB) locus (Figure 1c). In addition to extending our analyses to *hNEAT1*, we will examine other protein coding transcripts to determine the uniqueness of these results. Similar studies performed with either alkaline hydrolysis or additional RNase enzymes, such as RNase V1 (which specifically degrades only double-stranded RNA) and RNase H (which degrades DNA-RNA hybrids) should also place these results in a richer context.

In addition to better understanding the structure and function of these noncoding RNAs, such an approach may identify novel features related to RNA-mediated epigenetic silencing that may provide insights into whether other *Xist*-like molecules are present within the nucleus. In order to examine a large number of noncoding RNAs at a high resolution, we are in the process of developing tiling arrays for application to these studies.

MATERIALS AND METHODS

Nuclei were prepped from apparently normal human female lymphoblasts, GM00131 (Coriell Cell Repositories, NIGMS), using a kit based on the sucrose cushion based method of isolation (Nuclei PURE, Sigma). Chromatin was further purified by modifications to a Triton X-100 based method chromatin isolation from whole cells (Fujita et al., 1997), which, when applied to previously purified nuclei likely permeabilizes nuclear membranes as DAPI staining and fluorescence microscopy of the resultant preparations shows gross morphological conservation of the shape of round nuclei. Specifically, 1 x 10^7 nuclei were transferred to prechilled 1.7 ml eppendorf tubes and centrifuged for 5 minutes at 500 x g (4°C). Nuclear pellets were resuspended by pipette in 1 ml ice-cold, fresh cytoskeletal (CSK) buffer (10 mM Pipes, pH 6.8, 100 mM NaCl, 300 mM sucrose, 1 mM $MgCl_2$, 1 mM EGTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 μg/ml aprotinin) supplemented with 0.5% Triton X-100 detergent and incubated on ice for 10 minutes. This incubation was followed by a 3 minute centrifugation at 1300 x g at 4°C. Chromatin pellets were washed twice with 1 ml ice-cold CSK buffer containing no Triton X-100 and once with ice-cold 1 ml SSPB buffer (10 mM Tris-HCl pH 7.0, 10 mM $MgCl_2$, 100 mM KCl). Centrifugation steps in between washes were all the same (1300 x g, 4°C). After washing, chromatin pellets were resuspended in 350 μl SSPB buffer containing either 700 U of SUPERasin RNase inhibitor (Ambion) or 0.01 U of RNase A and 0.4 U of RNase T1 (RNase cocktail, Ambion). Samples were incubated at 30°C for 15 minutes, then placed on ice. Chromatin was pelleted as before (1300 x g, 4°C) and washed 2 timed with SSPB buffer. After removal of the last wash, RNA was purified from the pellet using 1 ml TRIzol

reagent (Invitrogen) according to the manufacturer's directions. RNA was DNase treated to reduce genomic DNA contamination using Ambion's DNAfree kit. Real-time PCR analysis was performed using a Bio-Rad iCycler and Applied Biosystems 2X SYBR green master mix supplemented with a small amount of fluorescein dye according to Bio-Rad's instructions. Standard curves of each primer pair were created using a 10-fold dilution series of SUPERasin protected cDNA. cDNA from chromatin exposed to RNase was quantified relative to this protected sample using small real-time PCR primers designed to amplify 60-80 nt products in a two-step 95°C for 15 seconds, 60°C for 1 minute PCR program.

REFERENCES

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol *132*, 259-275.

Fujita, M., Kiyono, T., Hayashi, Y., and Ishibashi, M. (1997). In vivo interaction of human MCM heterohexameric complexes with chromatin. Possible involvement of ATP. J Biol Chem *272*, 10928-10935.

Ganesan, S., Silver, D. P., Greenberg, R. A., Avni, D., Drapkin, R., Miron, A., Mok, S. C., Randrianarison, V., Brodie, S., Salstrom, J., *et al.* (2002). BRCA1 supports XIST RNA concentration on the inactive X chromosome. Cell *111*, 393-405.

Appendix Three

Monoallelic Expression and Asynchronous Replication of p120 catenin in Mouse and Human Cells

Alexander A. Gimelbrant[1], Alexander W. Ensminger[1,2], Peimin Qi[1], Jacob Zucker[1] and Andrew Chess*[1,2]

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142
[2]Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

This chapter was published in the Journal of Biological Chemistry 280(2): 1354-1359 (2005).

Alexander W. Ensminger performed the DNA FISH assay of replication timing for the human loci.

ABSTRACT

The number of autosomal mammalian genes subject to random monoallelic expression has been limited to genes highly specific to the function of chemosensory neurons or lymphocytes, making this phenomenon difficult to address systematically. Here we demonstrate that asynchronous DNA replication can be used as a marker for the identification of novel genes with monoallelic expression, and identify p120 catenin, a gene involved in cell adhesion, as belonging to this class. p120 is widely expressed; its presence in available cell lines allowed us to address quantitative aspects of monoallelic expression. We show that the epigenetic choice of active allele is clonally stable, and that biallelic clones express p120 at twice the level of monoallelic clones. Unlike previous reports about genes of this type, we find that expression of p120 can be monoallelic in one cell type and strictly biallelic in another. We show that in human lymphoblasts, the silencing of one allele is incomplete. These unexpected properties are likely to be widespread, as we show that the Tlr4 gene shares them. Identification of monoallelic expression of a nearly ubiquitous gene indicates that this type of gene regulation is more common than previously thought. This has important implications for carcinogenesis and definition of cell identity.

INTRODUCTION

While the majority of mammalian genes are expressed from both parental alleles, there are some notable exceptions. In genomic imprinting, one allele is transcriptionally silenced, dependent on its parental origin (Reik and Walter, 2001). By contrast, X-inactivation in females is a random process that leads to the silencing of most of the genes along one of the two copies of the X chromosome (Lyon, 1986). This random choice is made independently by multiple cells around the time of implantation and descendant cells maintain the choice, leading to mosaicism.

In the last decade a class of genes has emerged with properties similar to the genes subject to random X-inactivation. Initially, the only known randomly monoallelically expressed genes on autosomes were the antigen receptors on lymphocytes (Pernis et al., 1965) which were considered a special case because they undergo DNA rearrangement. Interest in this type of regulation was stimulated by the discovery that the members of 1,000-gene family of olfactory receptor genes are also expressed in a random monoallelic fashion (Chess et al., 1994). Pheromone receptors, and a number of immune system molecules, including natural killer cell receptors and interleukins, have now been shown to be monoallelically expressed (Bix and Locksley, 1998; Held et al., 1995; Hollander et al., 1998; Rhoades et al., 2000). Interleukins differ from other genes in this class in that their transcription can be monoallelic in some cells and biallelic in other cells (Bix and Locksley, 1998; Rhoades et al., 2000)

How one could identify novel genes that are subject to random monoallelic expression? Technical challenges of single-cell analysis of allelic choice in a mosaic tissue make identification of more genes with this type of regulation difficult. However,

all classes of monoallelically expressed genes share the property of asynchronous DNA replication: during S-phase, one allele of such genes is replicated before the other. This DNA replication asynchrony has been assayed by fluorescence in situ hybridization (FISH) and S-phase fractionation methods (Kitsberg et al., 1993; Mostoslavsky et al., 2001; Selig et al., 1992; Singh et al., 2003; Taylor, 1960). For genes subject to random monoallelic expression, asynchronous DNA replication is coordinated in a chromosome-wide fashion (Ensminger and Chess, 2004; Singh et al., 2003). Importantly, the asynchrony is independent of expression of the gene in the assayed tissue, e.g. olfactory receptor gene replication is asynchronous in fibroblasts or ES cells, where these genes are not expressed. Thus, an attractive approach to identification of novel loci is to detect asynchronous replication in a clonal cell line, and then characterize expression of the candidate genes in the appropriate cell types.

Using this strategy, we find that the p120 catenin gene, which encodes a component of the cell adhesion machinery, is an asynchronously replicated gene. We then show that it is monoallelically expressed in a subset of mouse and human clonal cell lines, in a manner similar to interleukins. The p120 catenin gene (gene symbol: catns in mouse, ctnnd1 in human; for uniformity, we refer to it as p120 in this paper) is widely expressed (Montonen et al., 2001), in contrast to the previously characterized monoallelically expressed genes. Expression of this gene in cell types that are easy to isolate and subclone allows for characterization of its transcriptional regulation.

MATERIALS AND METHODS

Animals and cells. To obtain 129xCsF1 animals, 129/SvJ females were crossed with CAST/Ei males (Jackson Labs). Abelson cells were prepared from bone marrow of 6-week old F1 animals, or from embryonic day 14 liver, by infecting primary culture with Abelson virus (Rosenberg et al., 1975). Primary fibroblasts were generated simultaneously from ears of the same adult animals, or from the embryonic tissue of the same embryos, respectively; after primary culture establishment, they were transformed with SV-40 virus. Cells were further cloned by limiting dilution (to the average of 0.3 cells per well) or by FACS. Normal human lymphoblastoid lines were obtained from Coriell Cell Repository, grown according to instructions, and cloned by FACS. Clonal ES cell lines used for primer extension replication timing analysis were provided by Dr. A.Wutz. E1A-transformed fibroblasts were provided by Dr. Elsa Flores.

Replication timing – FISH was performed essentially as described (Selig et al., 1992). Briefly, cells were given a 45 min pulse of BrdU before fixation. The DNA probe (BAC, cosmid, or PCR product as noted) was nick-translated (Amersham kit) with Cy3-dCTP (Molecular Probes) or FluorX-dCTP (Amersham). After overnight hybridization and stringent washes, the nuclei were stained with anti-BrdU antibodies (BD), with secondary antibody (Covance or JIRL) coupled to fluorophore complementary to that of the probe, and counterstained with DAPI. For replication timing analysis, only BrdU-positive cells were counted. The complete list of probes is in the Supplement.

Replication timing – primer extension was performed as described in (Xiong et al., 1998), with the following modifications. Unsynchronized clonal populations of ES cells from 129XCsF1 mouse were grown in the presence of LIF. 10^8 cells were fixed in

75% EtOH in a standard procedure for cell cycle analysis (Latt, 1973), stained with DAPI in the presence of RNAse A and FACS sorted into eight fractions: G_0/G_1, six equally wide S phase fractions, and G_2, each sorted fraction containing at least $5x10^5$ cells. Genomic DNA was prepared from these fractions in a standard procedure (Sambrook et al., 1989). Informative SNP in mouse: CTTGACCTGGC(T/C)GTTTTGCAAG (Lindblad-Toh et al., 2000). The list of all used polymorphisms and primers (for PCR and primer extension) is in the Supplement. In order to distinguish the parental alleles, single nucleotide polymorphism (SNP) was used in a manufacturer-recommended genotyping procedure. In the DNA prepared from each cell cycle fraction, regions containing the SNP were PCR amplified. The relative amount of the two parental alleles was determined by primer extension of PCR products. Detection was by matrix-assisted laser desorption/ionisation - time-of-flight (MALDI-TOF) mass spectrometry on Sequenom (San Diego, CA) MassArray platform (Cowles et al., 2002; Tang et al., 1999). In each individual reaction, the ratio of peak heights corresponding to the two alleles was calibrated to a series of mixes with known composition of purified genomic DNA from parental strains (Jackson Labs): 60:40, 50:50, and 40:60 mix. DNA from each point in cell cycle was amplified and measured in quadruplicate.

RT-PCR and RFLP analysis. Total RNA from 2-5 million cells was prepared with standard Trizol (Invitrogen) protocol, with additional genomic DNA digestion (DNAfree kit, Ambion). 0.5-1 μg of RNA was reverse transcribed using MMLV polymerase (Invitrogen) and oligo-dT primers (Ambion); 0.5-1 μl of RT reaction was used as PCR template, with 5-10 μl added directly into endonuclease digestion reaction, results of which were resolved on 2% agarose gel, and stained with EtBr. The complete list of all

used polymorphisms and primers is in the Supplement. Gel quantitation was performed using the public domain NIH Image program (developed at the U.S. National Institutes of Health and available on the Internet at http://rsb.info.nih.gov/nih-image) with the included set of gel analysis macros.

Mass spectrometric quantitation of RT-PCR products was performed as described for genomic DNA, except that 0.1μl of RT reaction was used as template for PCR, and that mixes of 1:3; 1:1; and 3:1 of paternal:maternal genomic DNA were used for calibration.

RESULTS

Asynchronous replication of p120 Catenin

We identified the p120 catenin gene as an asynchronously replicating gene (Fig.1a, b). The first approach involved S-phase fractionation, PCR and primer extension to analyze allele specific DNA replication in cells derived from an F1 mouse as described previously (Singh et al., 2003; Xiong et al., 1998). We confirmed the asynchronous replication of p120 catenin using a standard fluorescence in situ hybridization (FISH) assay (Chess et al., 1994; Selig et al., 1992; Singh et al., 2003) in primary mouse fibroblasts (Fig.1c,d) as well as SV40 and E1A transformed fibroblasts, ES cells, and Abelson cells (not shown). In this FISH analysis, replicated loci are visualized as a double-dot hybridization signal, while unreplicated loci reveal a single dot. Asynchronously replicating genes reveal a single dot-double dot (SD) pattern in 30-40% of S phase cells as opposed to around 10-20% SD for synchronously replicating genes.

Monoallelic expression of mouse p120 Catenin

We analyzed p120 catenin expression in Abelson murine leukemia virus (Abl-MLV) transformed pre-B cell clonal cell lines derived from an F1 cross between a female 129 mouse and a male Mus castaneous (Cs) mouse. Clonal cell lines were made using a FACS to place a single cell per well, or by limiting dilution. Both approaches gave similar results. After determining that these cell lines express all four of the previously described splice variants, we proceeded to assess the relative mRNA levels of the two alleles of the p120 gene. We took advantage of a single nucleotide polymorphism (SNP) present in the 3' untranslated region of the message that can be distinguished either using

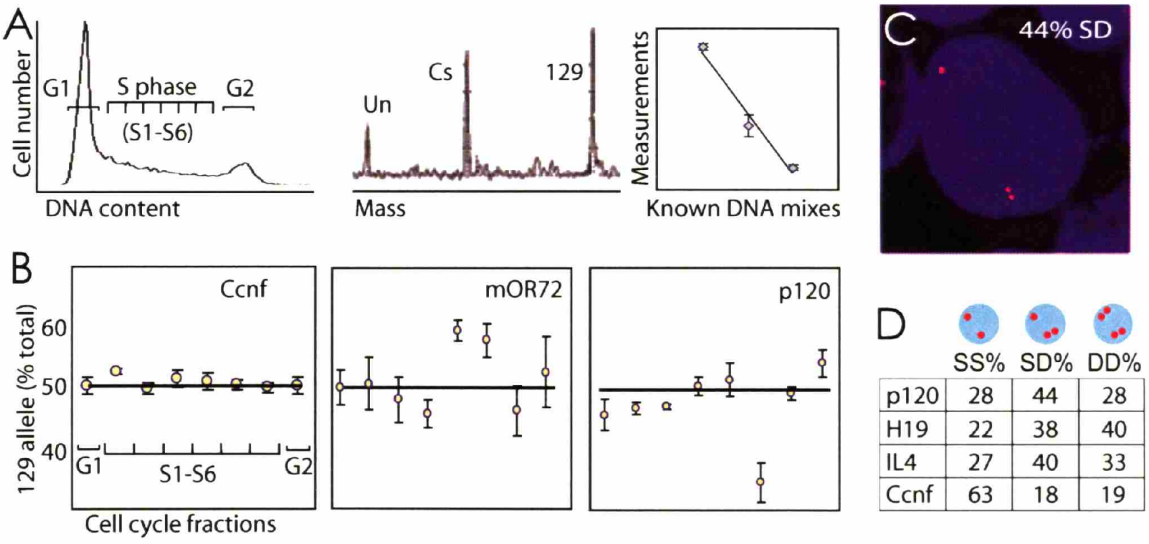
Figure 1

Asynchronous replication of mouse p120 catenin.

A. S-phase fractionation, primer extension assay for replication asynchrony. An unsynchronized clonal population of ES-23 cells from 129xCast F1 animals was FACS sorted according to DNA content into 8 cell cycle fractions: G1, G2, and six S-phase fractions, S1 through S6 (*left panel*, typical FACS profile); DNA from each fraction was PCR amplified with primers flanking a SNP distinguishing the 129 allele of an assayed gene from the Cast (Cs) allele; products of allele-specific primer extension were detected using mass spectrometer (*center*; Un - mass peak of unextended primer); to calibrate for bias and confirm linearity of allele detection, each experiment included known mixes of parental DNA (*right*; mean ± s.e.m., n=4). **B.** Primer extension assay readout. For each of the eight cell cycle fractions, the relative content of 129 (maternal) allele is presented as mean ± s.e.m. (n=4); no significant difference from 1:1 ratio of each allele marks synchronously replicated gene Adam1a (*left panel*), while for asynchronously replicated genes, such as olfactory receptor mOR72 (*center*), overabundance of the early replicated allele causes relative allele content in some S-phase fractions to differ significantly from 50%; p120 (*right panel*) was paternally early. **C.** FISH-based assay for asynchronous replication. Mouse fibroblast nucleus with single dot-double dot (SD) pattern; red – 10 kb PCR probe for p120 labeled with Cy3, blue – DAPI. **D.** Summary of the FISH assay in primary mouse fibroblasts with probes for p120 and for the asynchronously replicated imprinted gene H19, the randomly asynchronously replicated gene IL4, and a synchronously replicated locus Ccnf: SS – single dot-single dot, DD – double dot-double

dot; in each case, n=50-100. Similar results were observed with BAC and PCR probes for p120, as well as in ES cells, E1A-transformed fibroblasts and Abelson cells.

FIGURE 1:



	SS%	SD%	DD%
p120	28	44	28
H19	22	38	40
IL4	27	40	33
Ccnf	63	18	19

primer extension or by using a polymorphic Hae III restriction endonuclease site (RFLP analysis). Examples of restriction endonuclease analyses revealing monoallelic and biallelic cell lines are shown in Figure 2a. In total, we analyzed 78 clonal Abl-MLV transformed pre-B cell lines (derived from 8 mice) for p120 expression using RFLP. Analyses of allele specific expression were also performed using Sequenom mass spectrometric primer extension genotyping of RT-PCR products (see examples in Fig. 2d). While this genotyping platform is usually used to call heterozygotes and homozygotes, it has been demonstrated to detect the presence of a rare allele at 2% in a complex mixture (Buetow et al., 2001; Ross et al., 2000): therefore, for the Abl-MLV transformed lines in which only one allele is detected there is at least fifty-fold skewing. Sixty percent of the lines monoallelically express the p120 catenin gene (32% 129 and 28% Cs) with the remaining 40% of the lines showing biallelic expression. This distribution of monoallelic vs. biallelic expression is similar to what has been observed for the IL-2 and IL-4 genes (Bix and Locksley, 1998; Rhoades et al., 2000).

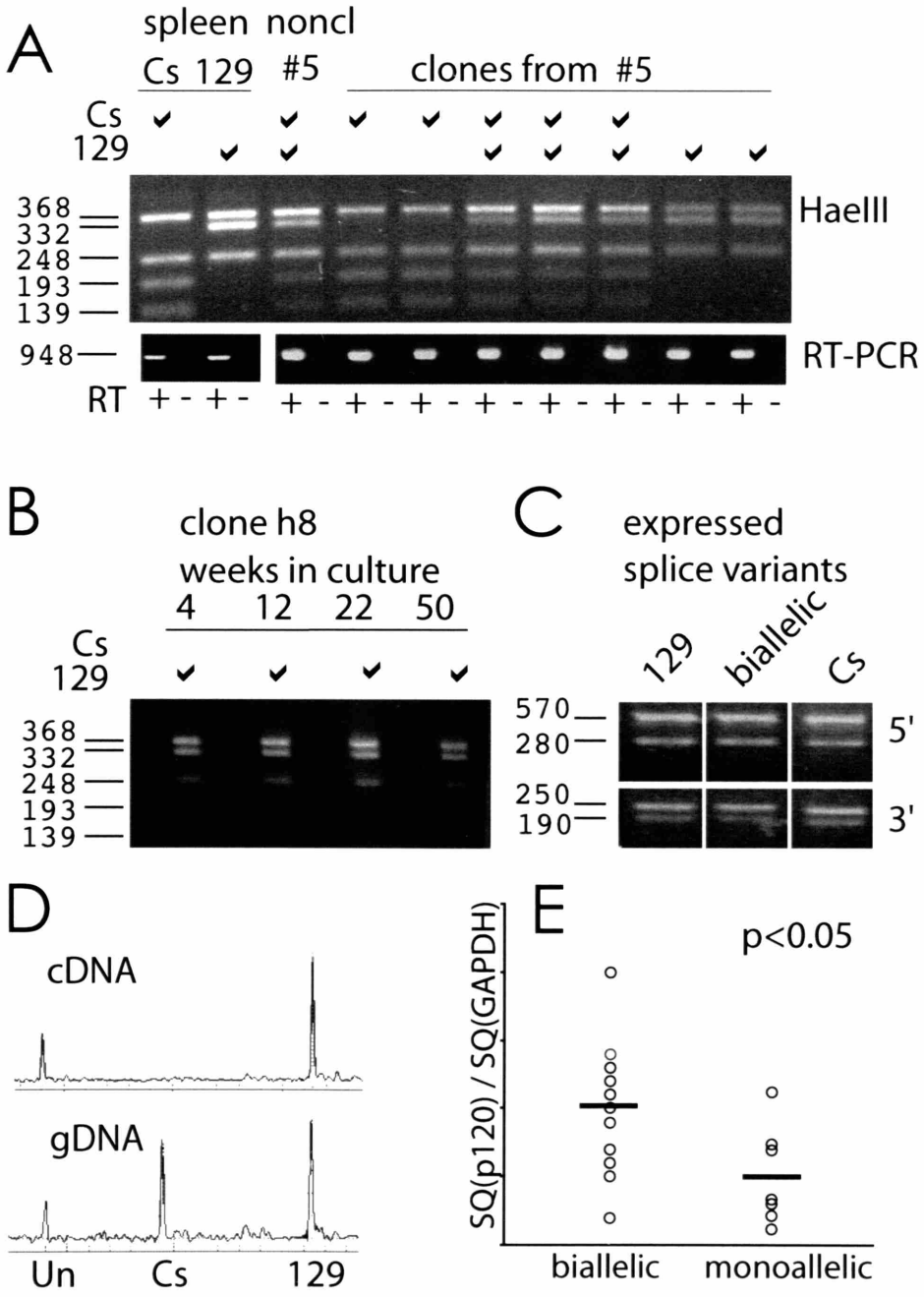
The stability of the allele-specific expression pattern was assessed both for monoallelic and biallelic clonal cell lines. Monoallelic lines were stable for as many as 50 weeks maintained in continuous culture (Fig. 2b). Biallelic clonal cell lines were also stable and in one case, we performed subcloning of a biallelic line and analyzed 15 subclones, all of which maintained biallelic expression qualitatively similar to the parental clonal cell line. Abl-MLV-transformed cell lines express all four reported splice variants of p120 catenin (Montonen et al., 2001), regardless of whether the paternal, maternal, or both alleles are expressed (Fig. 2c).

Figure 2

Expression of p120 catenin in Abelson cells.

A. Restriction digestion (Hae III) analysis of p120 RT-PCR products from spleen of Cast (Cs) or 129 mouse, from a polyclonal cell line derived from bone marrow of 129xCast F1 mouse (#5), and from individual clones from the same line. The calls (129, Cs, or biallelic) are indicated above the gel (3.5% agarose with ethidium bromide). Lower gel: under each digest lane, corresponding undigested PCR product and no-RT control are shown (1.5% agarose, ethidium bromide). Band size is indicated in base pairs. **B**. Restriction digestion (Hae III) of p120 RT-PCR products from the same clone (h8) expressing only the maternal allele of p120 catenin at different time points (as noted, in weeks) in continuous culture. 3.5% agarose gel with ethidium bromide. **C**. Expression of p120 catenin splice variants in Abl-MLV clonal lines expressing either allele, or biallelic. **D**. Both alleles are present in the genomic DNA indicating that monoallelic expression we observe is not due to the loss of the other allele. Primer extension/mass spectrometry traces from p120 RT-PCR from line h8 (*upper trace*) or genomic DNA from the same line (*lower trace*). Designations of the mass peaks are the same as in Fig.1. **E**. Real-time quantitative RT-PCR measurements of starting quantity (SQ) of p120 normalized to SQ of GAPDH in biallelic and monoallelic Abelson clonal lines. Horizontal bar represents the mean value.

FIGURE 2:


A
spleen noncl
Cs 129 #5
clones from #5
Cs
129
368
332
248
193
139
HaeIII
948
RT-PCR
RT + - + - + - + - + - + - + - + - + -
B
clone h8
weeks in culture
4 12 22 50
Cs
129
368
332
248
193
139
C
expressed
splice variants
129
biallelic
Cs
570
280
250
190
5'
3'
D
cDNA
gDNA
Un
Cs
129
E
p<0.05
SQ(p120) / SQ(GAPDH)
biallelic
monoallelic

To assess the relative levels of transcription in monoallelic vs. biallelic lines, we employed real-time quantitative PCR analysis of RT products, in each case normalizing the p120 level to measured level of the GAPDH transcript. On average, the biallelic clones express p120 catenin at twice the level that the monoallelic clones express (Fig. 2e). While there is a relatively high variation in the level of expression among both monoallelic and biallelic cell lines, a t-test analysis (two tailed, with unequal variance) indicates that the difference between the two distributions is significant at the $p<.05$ level. The distributions are consistent with a two-fold difference in expression between monoallelic and biallelic expression, suggesting that if a given allele is transcribed, its level of expression is independent of whether or not the other allele is active.

Asynchronous replication and monoallelic expression of the human p120 catenin gene

Replication timing of the human p120 catenin gene was examined using the FISH based replication timing assay on primary human fibroblasts. The human p120 gene reveals 41% of S-phase cells with a SD pattern, consistent with other asynchronously replicating genes (Fig. 3a,b). Thus, as was observed for the mouse gene, the human p120 catenin gene is asynchronously replicating.

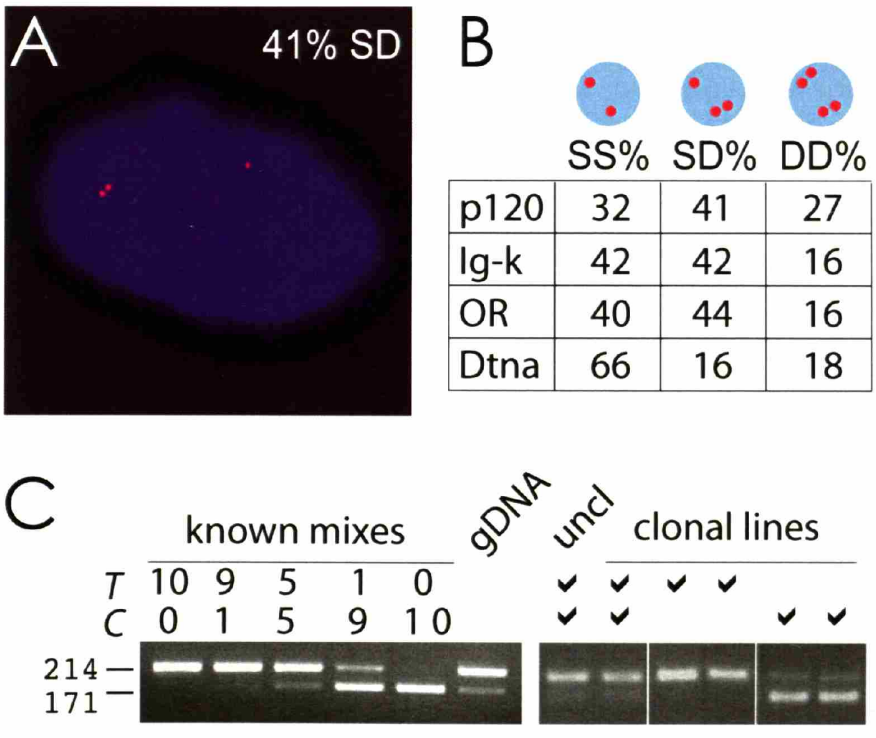
The relative expression levels of the two alleles of the human p120 gene were determined by analyzing clonal EBV transformed B cell lines derived from an individual heterozygous for a *Bst*NI RFLP in the p120 catenin cDNA. Subclones were generated using single cell FACS sorting and subjected to RT-PCR after sufficient cell expansion. Digestion of the RT-PCR product with the *Bst*NI endonuclease reveals that one allele is

Figure 3

Asynchronous replication and monoallelic expression of p120 catenin in human lymphoblasts.

A. Asynchronous replication of human p120. FISH was performed on an apparently normal human 46, XX primary fibroblast population (WI38). Red – p120 probe, blue – DAPI. A nucleus with single-double (SD) pattern is shown. **B**. Summary of results of FISH assay with probes for p120, as well as for asynchronously replicated genes Ig-kappa and OR6B3, and synchronously replicated gene Dtna; in each case, n=100 nuclei were counted; designations same as in Fig.1. Note that the counts for control genes Ig-kappa and OR6B3, which were performed simultaneously with the others, were reported elsewhere (Ensminger and Chess, 2004). **C**. Restriction digestion analysis (BstNI) of p120 PCR products. Note that PCR is biased towards the undigested (T allele) product, seen in a series of known mixes of cloned PCR products, as well as from genomic DNA from polyclonal lymphoblastoid line from apparently normal human (Coriell GM10849). Restriction digestion analysis of RT-PCR products from an uncloned line (uncl) and individual FACS-generated clones from it shows that different clones reveal patterns similar to the uncloned line and genomic DNA (biallelic), or preferentially express T or C allele. Genomic DNA from all tested cells was uniformly biallelic.

FIGURE 3:


A
41% SD
B
SS%
SD%
DD%
p120 32 41 27
Ig-k 42 42 16
OR 40 44 16
Dtna 66 16 18
C
known mixes
gDNA
uncl
clonal lines
T 10 9 5 1 0
C 0 1 5 9 10
214
171

digested and the other allele is not digested. Analyses of 23 single cell-derived subclones show a variety of patterns (Fig. 3c), with some cell lines preferentially expressing one allele and others preferentially expressing the other allele. Other cell lines express roughly equal levels of both alleles. A control analysis of genomic DNA derived from the parental cell line indicates that the primers amplify both alleles with a slight bias towards the undigested allele. The extent of allele-specific bias of the expression of the human p120 gene is not as great as the extent of skewing observed in the mouse. We estimate the extent of bias of the human gene to be 10-20 fold in some lines (Fig. 3c). By contrast, in the murine Abl-MLV transformed clones monoallelically expressing p120, the skewing was at least 50 fold if not absolute (Fig.2d).

Monoallelic expression of the mouse Tlr4 gene

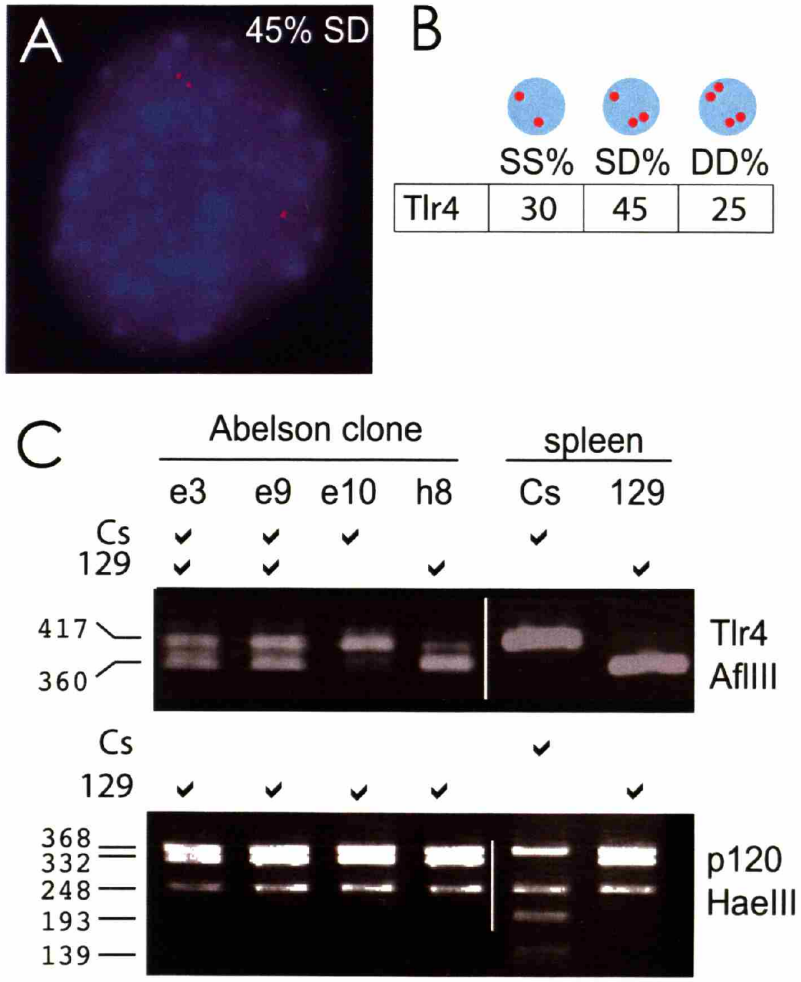
To examine a second monoallelically expressed gene in the same murine Abl-MLV transformed clones analyzed above, we examined the Tlr4 receptor gene which has been reported to be monoallelically expressed in B cells (Pereira et al., 2003). The Tlr4 receptor is involved in innate immunity. We demonstrated that Tlr4 is asynchronously replicating in mouse cells (Fig. 4a,b). Analyses of clonal cell lines indicated that one-fourth of them have monoallelic expression for either the maternal or paternal allele (Fig. 4c). However, these cell lines with monoallelic expression are not absolute in their monoallelic expression. Rather, we observe skewing of expression reminiscent of the skewing observed in analysis of the human p120 catenin gene. Based on control PCR experiments with known mixes of the two alleles as input, we estimate that the skewing observed in some Abl-MLV transformed B cell lines is around twenty fold (data not

Figure 4

Asynchronous replication and monoallelic expression of Tlr4 in murine cells.

A. Asynchronous replication of Tlr4. The FISH assay was performed on mouse fibroblast line, 5.2.2. A nucleus with a SD pattern is shown; red – Tlr4 probe labeled with Cy3, blue – DAPI. **B**. Summary of the FISH assay with probe against Tlr4; n=40 nuclei were counted, the designations are the same as in Fig.1. Similar results were observed in primary fibroblasts. **C**. Example of restriction digestion analysis (Afl III) of RT-PCR products from cloned Abelson lines e3, e9, e10 and h8, and parental controls from spleen of mice of parental strains. Below, RT-PCR and restriction digestion (Hae III) analysis of the same lines for p120 shows their maternal (129) monoallelic expression of p120.

FIGURE 4:


A
45% SD
B
SS%
SD%
DD%
Tlr4
30
45
25
C
Abelson clone
spleen
e3
e9
e10
h8
Cs
129
Cs
129
417
360
Tlr4
AflIII
Cs
129
368
332
248
193
139
p120
HaeIII

shown). Recall that the skewing observed for the human p120 catenin gene is around ten fold. The mouse Tlr4 is therefore a second example of an autosomal randomly monoallelically expressed gene that is not absolute in its allele-specific transcription patterns. We also examined clonal fibroblast cell lines for allele specific transcription of Tlr4 and found only biallelic lines (twelve lines). Note that similar analyses of (fourteen) clonal fibroblast lines revealed only lines biallelic for expression of the p120 catenin gene.

Tlr4 and p120 catenin are on different chromosomes; therefore, based on our prior observation that asynchronous replication is independently regulated for different chromosomes, we asked whether the allele specificity of monoallelic transcription of Tlr4 and p120 catenin are also independently regulated. We addressed this question by analyzing clonal Abl-MLV transformed B cell lines for allele-specific expression of both p120 catenin and Tlr4. In total, 43 lines were examined both for the Tlr4 and p120 catenin genes: 6 were monoallelic for both genes; 9 were biallelic for both genes; and 28 were monoallelic for one gene and biallelic for the other. These analyses reveal that the two genes are independent in their monoallelic expression (Fig. 4c). In the four examples shown, clonal Abl-MLV transformed lines express p120 catenin from the maternal allele. The maternal and paternal alleles of p120 can be distinguished using an RFLP; the maternal (129 allele) is digested by the *Afl* III restriction endonuclease. One of these lines is expressing predominantly the paternal Tlr4 allele, one line expresses predominantly the maternal allele and two lines express equal amounts of the two alleles. We also analyzed clonal cell lines prepared by limiting dilution which also reveal monoallelic expression for both genes (Fig. 4c). Thus, both in terms of whether there is monoallelic expression or

not, and in terms of which parental allele is monoallelically transcribed, these two genes are independently regulated.

DISCUSSION

We have demonstrated a new strategy to identification of a novel gene with random monoallelic expression: we detected asynchronous replication of p120 catenin, confirmed it by an independent method, and then characterized the gene's allele-specific expression in fibroblasts and cells of the B-cell lineage. The advantage of this approach is that it does not depend on whether the candidate gene is expressed in the cells used for initial analysis. This should allow for the scaling up of this approach to identify candidate genes in a systematic fashion, including genes expressed in an extremely restricted manner.

Our observation of allele-specific transcription of the p120 catenin gene has added several unexpected facets to our understanding of the epigenetic phenomenon of random monoallelic expression. Unlike the other genes in this class, such as olfactory receptors and T- and B-cell receptors, p120 is very widely expressed in human and mouse tissues. This expands the class of such genes in a crucial way, suggesting that other widely expressed genes may share this mode of transcriptional regulation. p120 catenin is similar to interleukins in that while a majority of Abelson cells express it from a single allele, a significant fraction of cells are expressing both alleles equally.

We found that p120 catenin is asynchronously replicated and monoallelically expressed in both mouse and human cell lines. This suggests that natural selection favored the features responsible for this unusual type of gene regulation over the about 60 million years separating primates and rodents from their last common ancestor. Monoallelic expression of p120 may be directly advantageous for mammals. It is also possible that it is a consequence of the genome architecture in this region: in both

mouse and human genome, p120 is less than 1 Mb away from clusters of olfactory receptors.

We observed, for the first time, that a gene can be randomly monoallelically expressed in one cell type (p120 in Abelson cells) and completely biallelically expressed in another (fibroblasts; see Fig.2). In practical terms, this shows that observing biallelic or monoallelic expression of a given gene in one cell type is not necessarily predictive for other cell types. More generally, this observation raises an important mechanistic question: what is the difference between the cells that are monoallelic or biallelic with respect to a given gene? Either their DNA carries different epigenetic modifications, or the modifications are identical, but cells of one type are competent for transcriptional readout of such modifications, and cells of another type are not. Note that some epigenetic mark is always present, as asynchronous DNA replication of p120 is ubiquitous. Further study of p120 and similar genes will help to distinguish these possibilities.

The question of causal relationship between asynchronous DNA replication and random monoallelic expression is very intriguing. To our knowledge, all known genes subject to random monoallelic expression are replicated asynchronously, an observation further confirmed by asynchronous replication of Tlr4 locus (Fig.4). Asynchronous replication thus appears to be a necessary condition of random monoallelic expression. It is, however, not a sufficient condition: for example, transcription of p120 is completely biallelic in fibroblasts, even as its replication is asynchronous in these cells. Moreover, further analysis should determine whether early or late replication status of a p120 allele determines the allele's activity. We have shown previously that for Ig-κ,

rearrangement occurs on the early replicating allele, rendering it active (Mostoslavsky et al., 2001); this may hold true for non-rearranged loci, as well.

Determining the relative allele activity is an inherent challenge in the analysis of random monoallelic expression. Expression in tissues is mosaic, and single-cell approaches provide limited ability for quantitative study. The fact that p120 is transcribed from either paternal or maternal allele in stable clonal cell lines allowed us to analyze its expression in more detail. The quantitative analysis of p120 expression in human lymphoblastoid lines revealed an unexpected feature: incomplete silencing, where in some cell lines, one allele is about ten-fold less active then the other one (Fig.3). This resembles the incomplete and variegated silencing of X-linked gene Rep1 in female cells (Carrel and Willard, 1999).

Another intriguing observation is that cell lines expressing p120 from both alleles have a level of p120 transcript about two-fold higher than in the lines expressing p120 from a single allele (Fig. 2e). Thus, similar to X-inactivation, it appears that the transcriptional regulation is centered not on the total level of transcript in the cell but on each given promoter; when the second allele is on, both alleles are equally active, resulting in double dose of the transcript. Therefore, monoallelic expression in a fraction of cells of a given type creates variability in the dose of expression of the affected gene. It is especially noteworthy that even with the two alleles absolutely identical, the difference in the transcript level of the affected gene(s) between cells expressing the gene from one allele and cells expressing it from both alleles, can cause these cells to be functionally different. This is potentially relevant to tumorigenesis, as discussed below.

Both incomplete silencing and cell-type dependence of monoallelic expression are not unique features of p120, as we show that Tlr4 shares these properties in mouse cells (Fig.4), suggesting that these properties may be relatively widespread. Another point arising from our analysis of monoallelic expression of Tlr4 and p120 in the same cells, is that these genes are independently regulated. We have shown previously that the epigenetic mark responsible for asynchronous replication is independent for different autosomes, even as it is coordinated on a given autosomal pair (Singh et al., 2003). Considering that p120 and Tlr4 are on different chromosomes, it is thus not surprising that their transcriptional choice is independent of one another. However, an important point arises with monoallelic genes that could be coexpressed in the same cell (similar to p120 and Tlr4). Even the cells with exactly the same expression profile would differ in their complement of alleles expressed from multiple genes of this type. The combinatorial possibilities are large and add an extra dimension to definition of cell identity.

An especially intriguing hypothesis concerns variability in allele-specific expression in genes involved in carcinogenesis. Increased epigenetic heterogeneity has been suggested as a contributing factor in metastatic potential of tumor cells (Ohlsson et al., 2003). The exclusive expression of a defective allele of a tumor suppressor gene would increase chances of malignant transformation in this cell (Ohlsson et al., 1998). A more subtle possibility is presented by p120 catenin. As part of the E-cadherin complex (Anastasiadis and Reynolds, 2001), p120 catenin interacts with tyrosine kinases (Cozzolino et al., 2003; Piedra et al., 2003), whose activity impacts on the strength of cell-cell adhesion. Overrepresentation of individual splice variants of human

p120 catenin has been shown to dramatically affect cell shape (Aho et al., 2002), and the observation of changes in expression of p120 in tumors has led to the suggestion of a role for p120 in tumorigenesis and/or metastasis (Anastasiadis and Reynolds, 2001; Mayerle et al., 2003; Mo and Reynolds, 1996; Thoreson and Reynolds, 2002). Monoallelic expression could cause even a subtle mutation – e.g., a mutation affecting preferential outcomes of alternative splicing – to change the function of the cell expressing only the mutated allele. Moreover, as noted above, even with two identical alleles, random monoallelic expression can cause variation in levels of expression, including that of p120. These ideas raise an interesting possible connection between monoallelic expression that is present in normal cells and early events in tumorigenesis.

ACKNOWLEDGEMENTS

This work was supported by grants from NIH to AC. AE was supported by HHMI predoctoral fellowship.

We thank S.Gabriel, M.Defelice, B.Blumenstiel for assistance with Sequenom genotyping system; G.Paradis for help with FACS sorting; C.Cowles for advice with primer-extension genotyping; E.Winchester, H.Skaletsky, G.Bell and B.Yuan for help with software; H.Higgins for help in preparing the manuscript; R.Jaenisch, T.Orr-Weaver, A.Bortvin, and members of the Chess lab for useful discussions.

REFERENCES

Aho, S., Levansuo, L., Montonen, O., Kari, C., Rodeck, U., and Uitto, J. (2002). Specific sequences in p120ctn determine subcellular distribution of its multiple isoforms involved in cellular adhesion of normal and malignant epithelial cells. J Cell Sci *115*, 1391-1402.

Anastasiadis, P. Z., and Reynolds, A. B. (2001). Regulation of Rho GTPases by p120-catenin. Curr Opin Cell Biol *13*, 604-610.

Bix, M., and Locksley, R. M. (1998). Independent and epigenetic regulation of the interleukin-4 alleles in CD4+ T cells. Science *281*, 1352-1354.

Buetow, K. H., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., Little, D. P., Strausberg, R., Koester, H., Cantor, C. R., and Braun, A. (2001). High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip- based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Proc Natl Acad Sci U S A *98*, 581-584.

Carrel, L., and Willard, H. F. (1999). Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. Proc Natl Acad Sci U S A *96*, 7364-7369.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834.

Cowles, C., Hirschhorn, J. N., Altshuler, D., and Lander, E. S. (2002). Detection of regulatory variation in mouse genes. Nature Genetics *32*, 432-437.

Cozzolino, M., Stagni, V., Spinardi, L., Campioni, N., Fiorentini, C., Salvati, E., Alema, S., and Salvatore, A. M. (2003). p120 Catenin Is Required for Growth Factor-dependent Cell Motility and Scattering in Epithelial Cells. Mol Biol Cell *14*, 1964-1977.

Ensminger, A. W., and Chess, A. (2004). Coordinated replication timing of monoallelically expressed genes along human autosomes. Hum Mol Genet *13*, 651-658.

Held, W., Roland, J., and Raulet, D. H. (1995). Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. Nature *376*, 355-358.

Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998). Monoallelic expression of the interleukin-2 locus. Science *279*, 2118-2121.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D. J., Nicholls, R. D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. Nature *364*, 459-463.

Latt, S. A. (1973). Microfluorometric detection of deoxyribonucleic acid replication in human metaphase chromosomes. Proc Natl Acad Sci U S A *70*, 3395-3399.

Lindblad-Toh, K., Winchester, E., Daly, M. J., Wang, D. G., Hirschhorn, J. N., Laviolette, J. P., Ardlie, K., Reich, D. E., Robinson, E., Sklar, P., *et al.* (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet *24*, 381-386.

Lyon, M. F. (1986). X chromosomes and dosage compensation. Nature *320*, 313.

Mayerle, J., Friess, H., Buchler, M. W., Schnekenburger, J., Weiss, F. U., Zimmer, K. P., Domschke, W., and Lerch, M. M. (2003). Up-regulation, nuclear import, and tumor growth stimulation of the adhesion protein p120 in pancreatic cancer. Gastroenterology *124*, 949-960.

Mo, Y. Y., and Reynolds, A. B. (1996). Identification of murine p120 isoforms and heterogeneous expression of p120cas isoforms in human tumor cell lines. Cancer Res *56*, 2633-2640.

Montonen, O., Aho, M., Uitto, J., and Aho, S. (2001). Tissue distribution and cell type-specific expression of p120ctn isoforms. J Histochem Cytochem *49*, 1487-1496.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B. E., Chess, A., Cedar, H., and Bergman, Y. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221-225.

Ohlsson, R., Kanduri, C., Whitehead, J., Pfeifer, S., Lobanenkov, V., and Feinberg, A. P. (2003). Epigenetic variability and the evolution of human cancer. Adv Cancer Res *88*, 145-168.

Ohlsson, R., Tycko, B., and Sapienza, C. (1998). Monoallelic expression: 'there can only be one'. Trends Genet *14*, 435-438.

Pereira, J. P., Girard, R., Chaby, R., Cumano, A., and Vieira, P. (2003). Monoallelic expression of the murine gene encoding Toll-like receptor 4. Nat Immunol *4*, 464-470.

Pernis, B., Chiappino, G., Kelus, A. S., and Gell, P. G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. J Exp Med *122*, 853-876.

Piedra, J., Miravet, S., Castano, J., Palmer, H. G., Heisterkamp, N., Garcia de Herreros, A., and Dunach, M. (2003). p120 Catenin-associated Fer and Fyn tyrosine kinases regulate beta-catenin Tyr-142 phosphorylation and beta-catenin-alpha-catenin Interaction. Mol Cell Biol *23*, 2287-2297.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet *2*, 21-32.

Rhoades, K. L., Singh, N., Simon, I., Glidden, B., Cedar, H., and Chess, A. (2000). Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. Curr Biol *10*, 789-792.

Rosenberg, N., Baltimore, D., and Scher, C. D. (1975). In vitro transformation of lymphoid cells by Abelson murine leukemia virus. Proc Natl Acad Sci U S A *72*, 1932-1936.

Ross, P., Hall, L., and Haff, L. A. (2000). Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. Biotechniques *29*, 620-626, 628-629.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

Selig, S., Okumura, K., Ward, D. C., and Cedar, H. (1992). Delineation of DNA replication time zones by fluorescence in situ hybridization. Embo J *11*, 1217-1225.

Singh, N., Ebrahimi, F. A., Gimelbrant, A. A., Ensminger, A. W., Tackett, M. R., Qi, P., Gribnau, J., and Chess, A. (2003). Coordination of the random asynchronous replication of autosomal loci. Nat Genet *33*, 339-341.

Tang, K., Fu, D. J., Julien, D., Braun, A., Cantor, C. R., and Koster, H. (1999). Chip-based genotyping by mass spectrometry. Proc Natl Acad Sci U S A *96*, 10016-10020.

Taylor, J. H. (1960). Asynchronous duplication of chromosomes in cultured cells of chinese hamster. J Biophys Biochem Cytol *7*, 455-463.

Thoreson, M. A., and Reynolds, A. B. (2002). Altered expression of the catenin p120 in human cancer: implications for tumor progression. Differentiation *70*, 583-589.

Xiong, Z., Tsark, W., Singer-Sam, J., and Riggs, A. D. (1998). Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension. Nucleic Acids Res *26*, 684-686.

~The End~