Optimization of Chromatin Immunoprecipitation

Ronak Shah

School of Biological Sciences
Functional Genomics Research Stream, PAI 2.46
Section of Molecular Genetics and Microbiology

The University of Texas at Austin

Spring 2012 Graduate

An Honors Thesis for the
Dean's Scholars Honors Program

Supervising Professor
Dr. Arturo de Lozanne, PhD

Research Mentor
Dr. Patrick Killion, PhD

Abstract

Transcription factors play a vital role in controlling cell growth, and locations of their binding sites at various points of the cell cycle can provide clues about malfunctions in eukaryotic growth, such as cancer. Our research focuses on transcriptional regulation of the eukaryotic cell cycle, using *Saccharomyces cerevisiae* as the model organism. We are observing binding patterns of affinity-tagged MCM1, SWI4, SWI5, FKH1, FKH2, and ACE2 transcription factors. These binding sites are discovered and isolated *in vivo* using chromatin immunoprecipitation (ChIP) followed by high-throughput, next-generation sequencing to map them to the genome for further analysis. The procedure locks transcription factors to their binding sites on DNA, and then eliminates the extraneous DNA to isolate the genes of binding site alone. However, ChIP generates a relatively low yield of DNA, often contaminated, and our research focuses on optimizing elements of the protocol to produce a higher, purer, output. Optimized lysis methods have reduced time and increased output of DNA, and sonication cycles have been adjusted to yield a more uniformly sheared mixture of DNA. The efficacy of sonication is evaluated through diagnostic gel electrophoresis and interpretation of visual results. In addition, the impact of pre-clearing on clarity of final yields was studied. DNA output and purity is repeatedly tested at various breakpoints of the ChIP procedure to ensure that the optimization modifications are delivering higher yields. DNA purity is tested using polymerase chain reaction (PCR) to find previously characterized target regions in the genome along with positive and negative controls to ensure homogeneity of the sample. Thus, by optimizing ChIP, we can obtain a highly accurate DNA sample more suitable for next-generation sequencing.

Background and Introduction

Purpose

The eukaryotic cell cycle consists of four main phases: mitosis, G1, synthesis, and G2. Transcription factors, or proteins that promote or repress the expression levels of various regulatory genes (such as cyclins and cyclin-dependent kinases), mediate the cell's progress through these phases. Each transcription factor has one or more specific binding sites to the cell's genome and can thus exert its effects on multiple genes. When a cell is affected by cancer, it experiences rapid, unregulated growth, likely due to mutations in oncogenes (responsible for reproduction and growth) or tumor suppressor genes (responsible for slowing or preventing growth). As a result, the affected cell no longer smoothly transitions from phase to phase in the cell cycle—the mutation prevents the cell from entering its natural checkpoints, leading to rapid and uncontrolled growth.

The long-term goal of the Functional Genomics research stream at the University of Texas at Austin is to eventually determine which of these genomic sites are involved in triggering cancerous behavior. This is done via chromatin immunoprecipitation (ChIP), a procedure where DNA of select regions can be purified and solely extracted without other contaminants. Previous research has determined the genes responsible for coding the transcription factors involved with regulating the cell cycle, but their binding sites have not been entirely discovered. Discovering the sequences and locations of these binding sites is ultimately the goal of the Functional Genomics research stream. It is hypothesized that a mutation in the sequences of these binding sites prevents proper transcription factor interactions, and this leads to the malfunction of the cell cycle. After the desired genomic DNA has been extracted via ChIP, it is then sent for sequencing.

First Steps

The first step was to choose a model organism, and the yeast strain *Saccharomyces cerevisiae* (which is eukaryotic) was used in this research lab for many reasons. Its ability to replicate quickly, produce easily extractable DNA, and arrest in various phases of the cell cycle were the essential attributes required of the lab's model organism. Most importantly, the entire genome of the yeast strain *Saccharomyces cerevisiae* has been sequenced, thus allowing the lab to determine the exact locations of the binding sites of the desired transcription factors. The YeastGenome database was essential in selecting the primary targets, and previous literature has determined which transcription factors are essential to mediating the cell cycle. Eight factors were chosen: ACE2, FKH1, FKH2, MBP1, MCM1, SWI4, SWI5, and SWI6. For chromatin immunoprecipitation to work, the chosen transcription factors must be tagged with TAP protein. Since the gene sequences of the aforementioned transcription factors have been discovered, it was possible to order specific strains of yeast with specific TAP-tagged transcription factors. These yeast cells have modified genomes, where the sequence for TAP protein is inserted after the sequence specific to the transcription factor, and as a result, the TAP protein is also translated along with the transcription factor to become a part of the overall protein structure.

The aforementioned transcription factors are active at different genomic locations throughout the cell cycle. Thus, the DNA must be extracted when all of the yeast cells are in the exact same phase of the cell cycle for uniformity. The drugs nocodazole and hydroxyurea allow for this, and microscopy is done to confirm that the cells have arrested, all in the same phase. Once the cells are all in synchronized growth, ChIP can begin.

Chromatin Immunoprecipitation

The first step of chromatin immunoprecipitation involves the cross-linking of proteins to DNA. This is done with formaldehyde, which binds all proteins to the nucleic acid. This allows for an in vivo capture of the transcription factors' binding sites.

After the cross-linking, the yeast cells must be broken down to expose the contents of the nucleic acid inside. This is done through a process called lysis, where glass beads are used to puncture the cell walls. Two lysis methods were studied to determine the most efficient—traditional, and via mini-bead beater. In the traditional method, the microcentrifuge tube containing the yeast cells in solution was fastened to a pulsing vortexor and placed in a 4°C refrigerator for varying times, to determine the optimal duration necessary to achieve a thorough lysis. Additionally, a mini-bead beater, or device that vigorously shakes the microcentrifuge tube at speeds of over a thousand pulses per minute, was introduced and tested. The rapid pulsing, coupled with glass beads, creates high friction, heating up the microcentrifuge tubes. Samples were not pulsated for over 90 seconds to prevent cellular degradation, and were placed on ice between trials of varying durations. This also justifies use of the refrigerator in the traditional lysis method.

Next, the genomic DNA is sheared in a process called sonication, where a needle-tipped rod is immersed in the suspension of cross-linked yeast cells, and resonates at a high frequency to break the phosphodiester bonds of the DNA backbone. The proteins are still able to remain bound to the DNA, and fragments are roughly 300 to 1000 bases in size. The vibration of the sonicator produces heat, and thus the samples needed to be cooled on ice in between cycles to prevent nucleotide or protein denaturation. One of the purposes of

this project was to improve the efficacy of sonication. Time length of sonication, as well as number of cycles, were manipulated to determine the optimal yield in the shortest time.

The next step is the most critical—immunoprecipitation. Here, the proteins of interest are isolated from everything else (cellular debris, organelles, and other proteins/nucleic acids) to yield a suspension only containing the TAP-tagged transcription factors along with their cross-linked DNA. Prior to true immunoprecipitation, a pre-clearing step may or may not be necessary. This is where pure sepharose beads are added to the cell lysate to remove the excess organic compounds that could potentially interfere with immunoprecipitation. The beads are left with the solution overnight on a gentle, refrigerated tube rocker. The next day, these sepharose beads are removed, taking with them excess proteins not involved with the immunoprecipitation. The "precipitation" step is then made possible by immunoglobulin G (IgG), an antibody. The TAP tag contains Protein A, a compound that has a high affinity for IgG. Protein A is not naturally produced by yeast cells, and is therefore able to accurately serve as a distinct marker for the selected transcription factors. The antibodies are coated onto the sepharose beads, which are added to the yeast cell suspension and left to mix on the same rocker overnight. This way, each of the binding sites on the IgG-laced beads can get exposed to the TAP-tagged elements in the mixture.

Another goal of the project was to optimize this step, where time left on the rocker could be minimized and more TAP elements could be obtained. Preclearing the lysate prior to immunoprecipitation seems to be an appropriate step, in theory. Removing the extraneous fragments of the cell that may interfere with the efficacy of IgG, prior to immunoprecipitation, would allow for a more thorough extraction of TAP-tagged

transcription factors. However, the IgG beads may nonetheless have their strongest affinity to Protein A, rendering the preclearing step unnecessary or trivial. Overnight rocking, even, may not be necessary to maximize the binding of transcription factors to the IgG. The experiments were conducted with and without preclearing steps (in varying lengths of time) to determine if it improved or even tarnished the overall ChIP yield.

Final Steps

The final steps after immunoprecipitation are to isolate the DNA and transcription factors bound to the antibody-coated beads. Three wash steps are completed, each of increasing stringency, to purify the mixture and remove all components except for the immunoglobulin beads with the TAP protein and its bound nucleic acids. This step is essentially a "post-clearing", though much faster than the preclearing step. After this mixture is as pure as possible, the components are eluted. This is where the antibodies and beads are separated from the transcription factors and DNA. The resulting solution ideally will consist of solely the TAP-tagged transcription factors bound to their genomic loci.

Lastly, in order to remove the transcription factors and just end up with the nucleic acids (which can then be sent to sequencing), the crosslinks are reversed. This breaks the bond created by formaldehyde initially between the transcription factor and DNA. RNase A and Proteinase K are then added to the mixture to remove contaminating traces of RNA or any other proteins. The DNA is then isolated from all other organic components by a phenol-chloroform wash and is then purified via the normal extraction protocol.

Once the DNA has been isolated, the yield will be checked by NanoDrop (in nanograms per microliter). Obtaining a sample density of over 100ng/μL is ideal, for this

quantity is sufficient for NextGen sequencing. After the sequencing step, the stream will use methods of computational biology to map the sequences to the genome and find regions of high overlap. Finally, the lab will elucidate peaks, or regions of high similarity between the genome and the newly sequenced sample. These peaks, finally, are assumed to be the key binding sites of the transcription factor at that specific point in its cell cycle where crosslinking occurred.

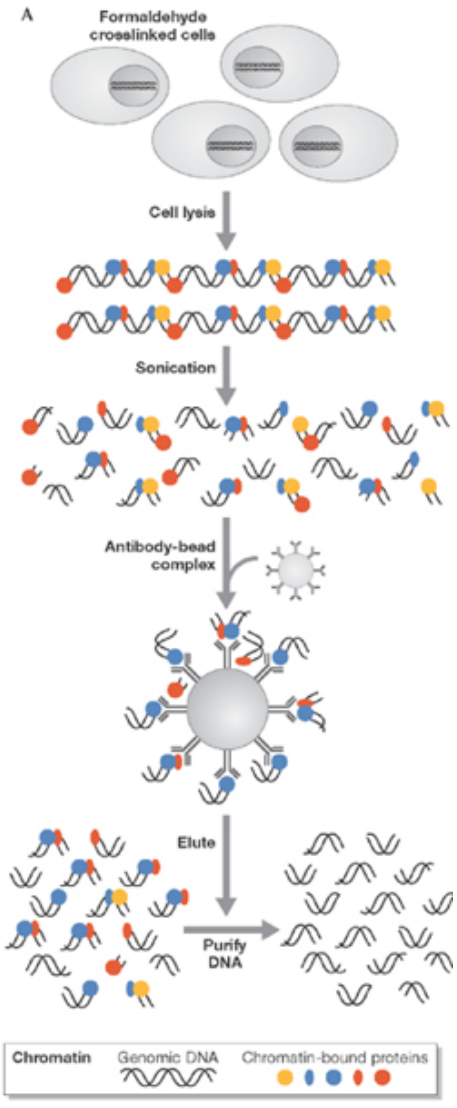

Figure 1: ChIP overview (Massie, 2008).

Results

Traditional, vortex-based lysis times were tested in varying lengths, with the most significant depicted in Table 1. Microcentrifuge tubes with the samples and glass beads were placed on a vortexor in a 4°C refrigerator overnight previously, but this was discovered to be inefficient. Additionally, the sonication protocol, set at 50% duty cycle, half power for 30 seconds, originally required four rounds, but more rounds were necessary.

	Sonication			
Cell Lysis		*4 rounds*	*6 rounds*	*8 rounds*
	2 hours	Non-homogenous shearing	Non-homogenous shearing	Strong results
	1 hour	Non-homogenous shearing	Non-homogenous shearing	Strong results
	30 minutes	Non-homogenous shearing, low yield	Non-homogenous shearing, low yield	Low yield

Table 1: Cycles of lysis time coupled with sonication rounds

Gel electrophoresis was used to evaluate the results, as shown in Figures 2-7 below. Immediately following the lysis and sonication procedures, the samples were loaded into agarose gels. Electrophoresis took 45 minutes to complete. Strong results would consist of a bright band at the bottom of the gel with minimal smear. The narrower the band, the more effective the sonication, indicating that the DNA was all uniformly sheared. A smear in the gel pointed to a less effective sonication, as the DNA fragments were sheared to varying lengths. The brighter the bands, the more thorough the lysis—more DNA was extracted from the yeast cells.

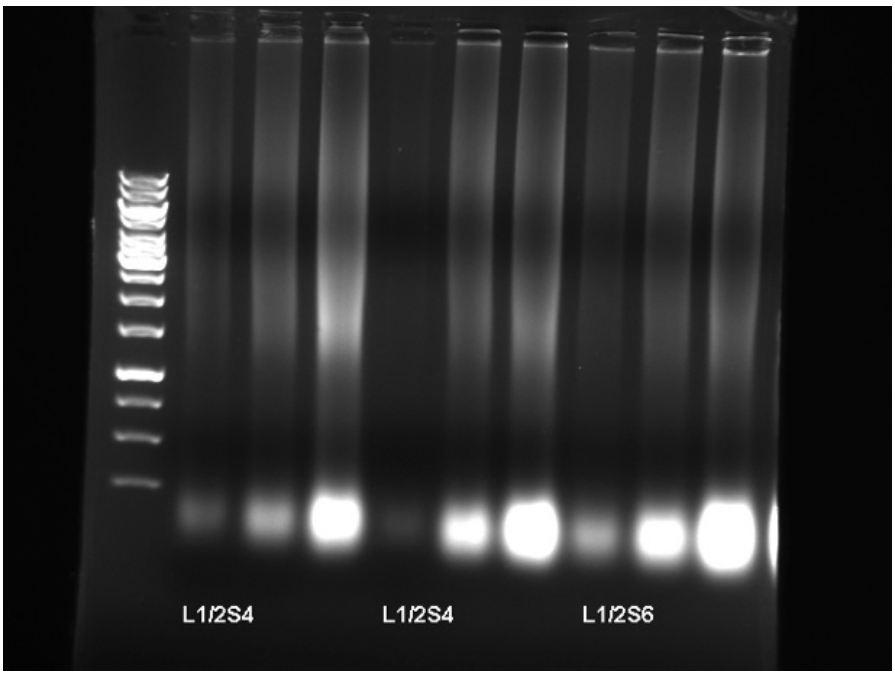

Figure 2

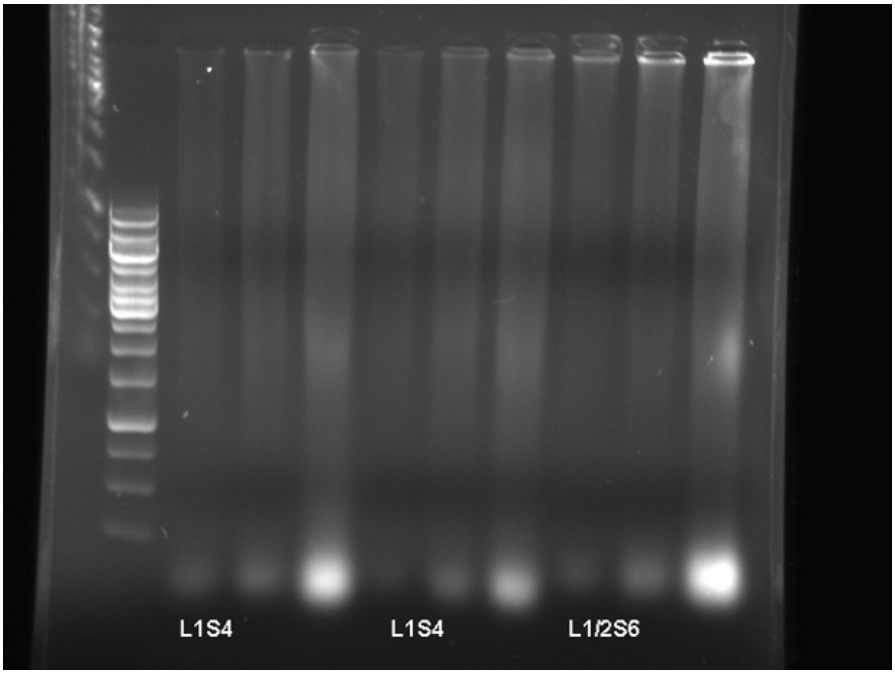

Figure 3

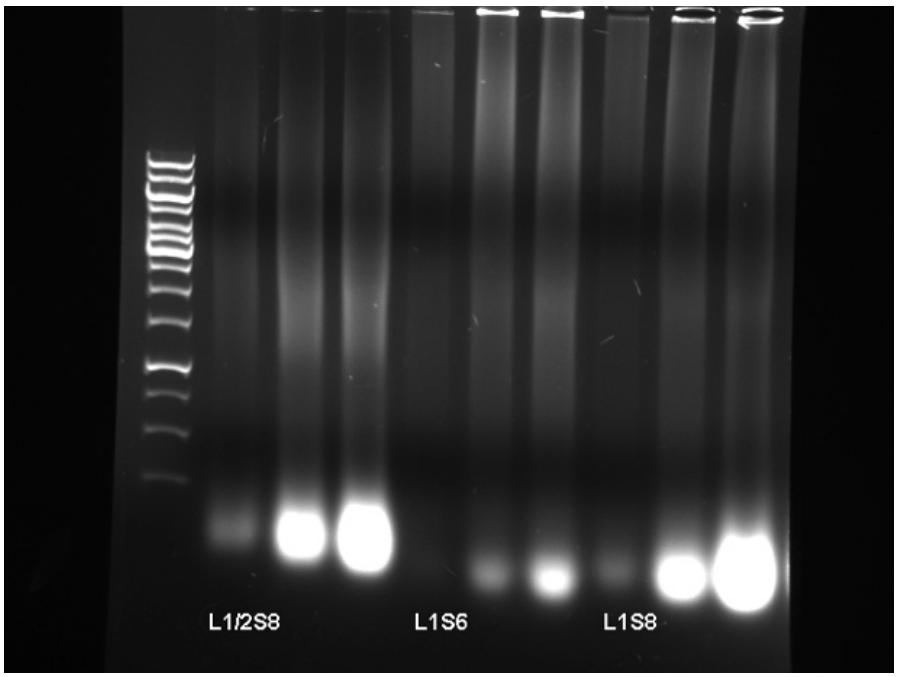

Figure 4

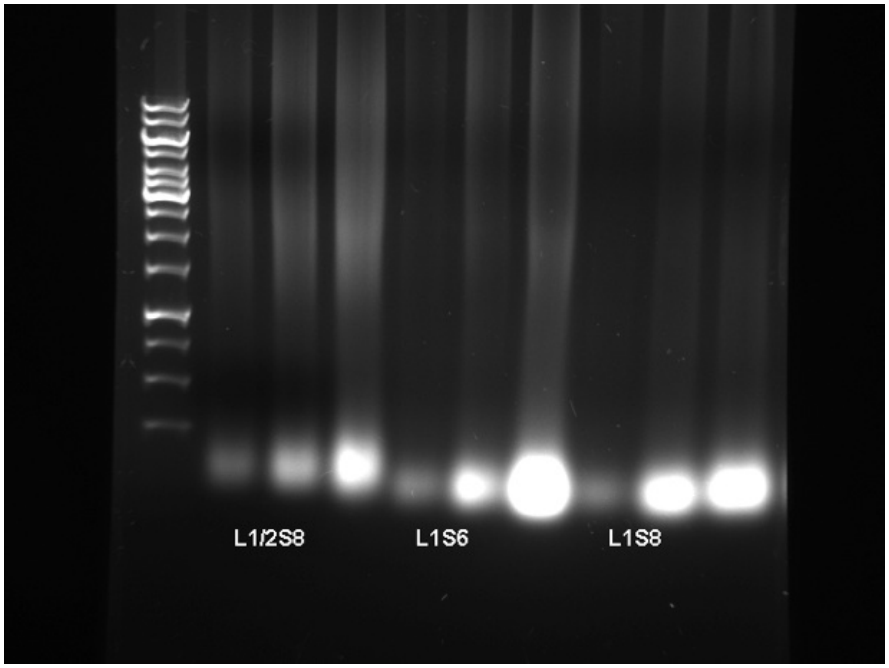

Figure 5

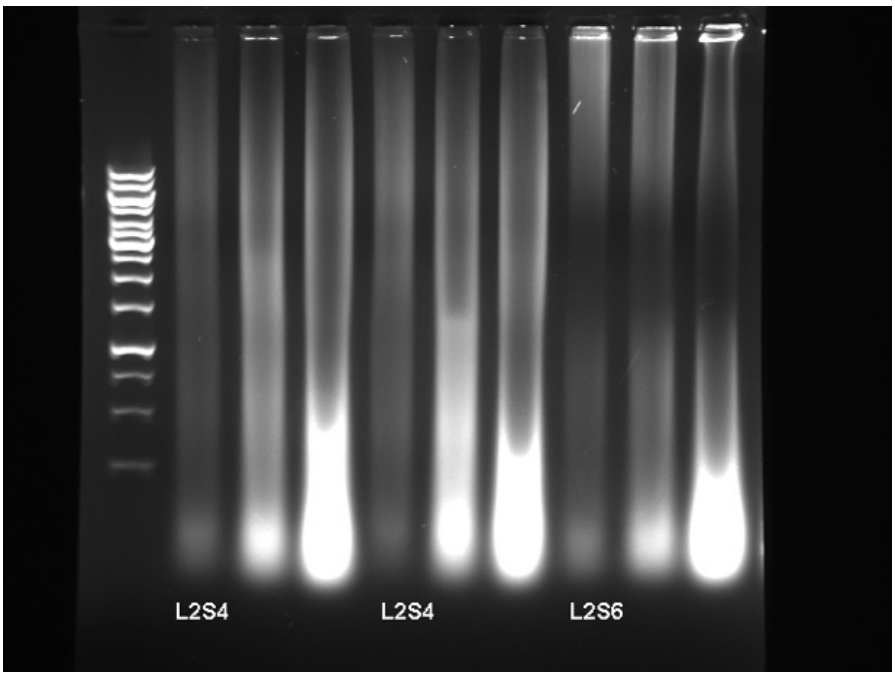

Figure 6

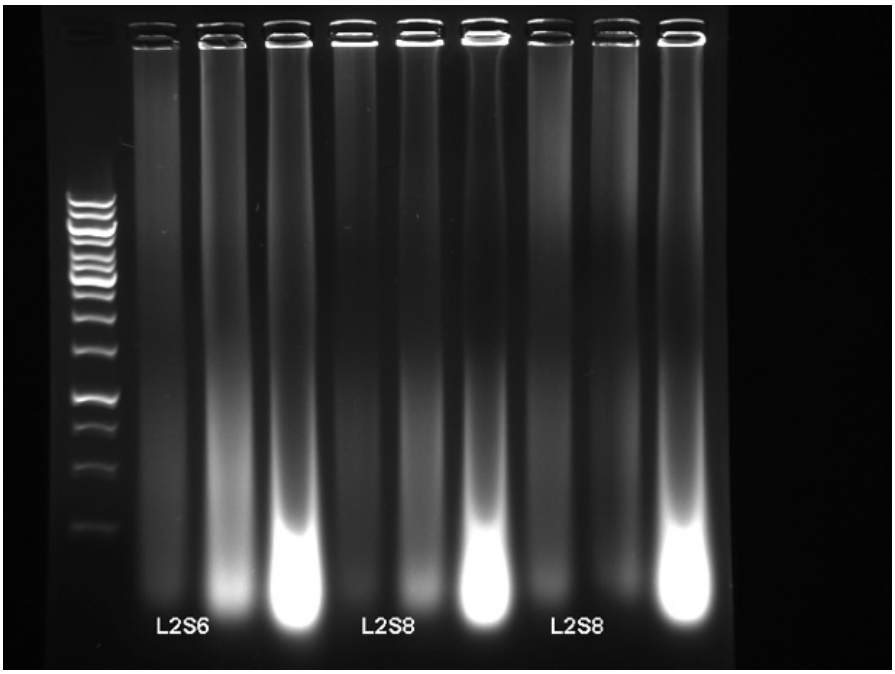

Figure 7

Discussion

The first lane of each gel contained a DNA ladder, serving as a marker for molecular weights. Additionally, the notation "L1S8," for example, indicates the time of lysis, in hours, followed by the number of sonication cycles. Three amounts of cell lysate, in increasing order (5μL, 10μL, and 15μL, respectively) were added per trial directly after sonication.

Figures 4 and 5 indicate that the one-hour lysis, coupled with eight cycles of sonication, was most efficient. As shown in the gel images, the most homogenous mixture of DNA fragments occurs with more sonication cycles. Moreover, it was proven that longer rounds of lysis do not always produce higher yields of DNA. The traditional laboratory protocol for ChIP was shortened by a full day after discovering that one hour of lysis was sufficient to thoroughly break apart the yeast cells and obtain the DNA. After many manipulations of the variables affecting DNA yields in the ChIP procedures, and rounds of trial and error, optimizing the lysis and sonication steps produced the greatest improvement of results at minimal cost.

Bibliography

Iyer et al. *Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF*. Nature (2001).

Massie, Charles E, and Ian G. Mills. "ChIPping away at gene regulation." *EMBO Reports.* European Molecular Biology Organization, 2008. Web. 13 Apr. 2012.

Nagalakshmi et al. *The transcriptional landscape of the yeast genome defined by RNA sequencing*. Science (2008).

Park. *ChIP-seq: advantages and challenges of a maturing technology.* Nature reviews Genetics (2009).

Simon et al. *Serial regulation of transcriptional regulators in the yeast cell cycle*. Cell (2001).

Spellman et al. *Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization*. Mol. Biol. Cell (1998).