The Wistar Institute of Anatomy and Biology

Annual Progress Report: 2009 Formula Grant

Reporting Period

January 1, 2010 – June 30, 2010

Formula Grant Overview

The Wistar Institute of Anatomy and Biology received $1,778,081 in formula funds for the grant award period January 1, 2010 through June 30, 2011. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

Modeling the Epigenetic Changes in Alternative Promoters of Cancer Genes - We propose a combination of computational, statistical and high-throughput experimental approaches to better characterize the use of alternative promoters of human genes in cancer cells, which will lead to improved biomarkers and drug discovery, one of the major challenges of cancer research. We will (1) develop an information resource of alternative promoters of cancer genes; (2) experimentally assess the active or repressed state of alternative promoters in different cancer cells relative to normal cells, using massive parallel sequencing approaches; and (3) develop statistical models to define unique and common epigenetic signatures across the promoters in different cancer cells.

Duration of Project

1/1/2010 - 6/30/2011

Project Overview

Cancer results from the accumulation of genetic and epigenetic events arising over a long period of time. Despite the fact that many important advances have been made in understanding the differences between normal and cancer cells, achieving a high level of specificity and efficacy in cancer diagnosis and treatment still remains a great challenge. A key insight provided by the use of high-throughput molecular technologies, such as microarrays, is the complexity and heterogeneity of cancer. A phenomenon that adds an important dimension to this complexity is the presence of alternative first exons and promoters within a gene. Sparing neighboring healthy tissues is an important consideration when designing therapies that suppress cancer cells. In a specific cell-type or tissue-type, only a subset of genes (or gene isoforms) is active, generally determined by a piece of DNA, known as the promoter, which lies upstream of the corresponding gene. Recent annotations of the human genome suggest that up to 70% of protein-coding genes make one or more alternative forms. Notably, the incidence of multiple-promoters among cancer genes is much higher than the proportion among the remainder of genes, and there

is growing evidence linking aberrant use of multiple promoters and cancer formation ─ several oncogenes and tumor-suppressor genes are already known to have multiple promoters, and moreover, it is known that the aberrant use of one promoter over another in some of these genes is directly linked to cancerous cell growth.

We hypothesize that alternative promoters in a gene locus are aberrantly used (either activated or silenced) in cancer genomes. Our rationale is that regional epigenetic modifications in alternative promoters are tissue-specific and that disruptions to these processes occur during cancer initiation and progression. These aberrant events may lead to the eventual silencing or activation of critical promoters of cancer genes. We propose a combination of computational, statistical and high-throughput experimental approaches to better characterize the use of alternative promoters of human genes in cancer cells, which will lead to improved biomarkers and drug discovery, one of the major challenges of cancer research. Specifically we will (1) develop an information resource of alternative promoters of cancer genes, (2) experimentally assess the active or repressed state of alternative promoters in different cancer cells relative to normal cells using massive parallel sequencing approaches, and (3) develop statistical models to define unique and common epigenetic signatures across the promoters in different cancer cells.

Principal Investigator

Ramana V Davuluri, PhD
Associate Professor
The Wistar Institute
3601 Spruce Street
Philadelphia, PA 19104

Other Participating Researchers

Louise Showe, PhD, Hyunsoo Kim, PhD, Yingtao Bi, PhD, Sharmistha Pal, PhD, Ravi Gupta, PhD - employed by The Wistar Institute

Expected Research Outcomes and Benefits

Two important questions of immediate attention that will be addressed by this research proposal are (i) which alternative promoters of human genes are preferentially silenced or activated in cancer cells, and (ii) what are the differences in the epigenetic modifications across the alternative promoters that cause differential use of the corresponding promoters in different tissues? The successful completion of this research will provide (i) an information resource of the alternative promoters of cancer genes, and (ii) a comprehensive survey of tissue-specific and cancer-specific alternative promoters. The combined use of bioinformatics and massive parallel sequencing technologies will help better understand the epigenetic mechanisms that regulate alternative promoters in a tissue-specific manner and provide clues to their aberrant regulation in cancer cells. We believe this research will identify a large number of novel targets for the diagnosis and treatment of different cancers. For example, imagine a drug that suppresses expression of a specific form of a human gene called CYP19 promoter II, which is aberrantly activated in breast cancer, while sparing the other five promoters that control expression of this

gene in healthy tissues. By targeting the specific problematic promoter, a more effective treatment with fewer side effects can be administered. It is not hard to imagine how specific targeting of individual promoters could lead to more accurate and sensitive diagnoses as well.

Summary of Research Completed

Development of CancerPromDb database: Integrated resources for annotation and visualization of mammalian cancer gene promoters from ChIP-seq and mRNA-seq experimental data.

The dissection of promoters and first-exons, including tandem alternative first exons of a gene responsible for the temporal and spatial expression in different tissues and cell-types, is a central focus of many research programs today. Although extensive molecular research has provided important details about several gene promoters -- their epigenetic changes (e.g., DNA methylation, histone methylation and acetylation patterns) and the expression pattern of corresponding genes -- the information generated over the years is highly fragmented. In order to better integrate this vast amount of information with the genome sequences, we are developing CancerPromDb (Promoter Database of Cancer Genes), an information resource of mammalian gene promoters involved in cancer.

CancerPromDb (Cancer gene Promoter Database) is a curated database that strives to integrate gene promoters with experimentally supported annotation of transcription start sites, cis-regulatory elements, CpG islands, and ChIP-Seq experimental results, focused on genes involved in cancer. The long-term goal of this database is to contribute to the understanding of disrupted mammalian genetic and epigenetic regulation, eventually leading to cancer initiation and progression. The initial version of CancerPromDb consists of computationally annotated and known active promoters from various published and unpublished ChIP-seq datasets generated for different tissues and cell lines publicly available at the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository. The workflow of the CancerPromDb 1.0 database is shown in Figure 1. Computational prediction is performed by integrating DNA sequence composition, DNA structural information, H3K4me3 ChIP-Seq experiments and CAGE tags. The predicted promoters are then annotated to an integrated list of known gene and transcripts models generated by combining RefSeq, Ensembl, Vega, UCSC gene and MGI representative transcripts. Based on analyzed ChIP-Seq datasets, we found 50% of genes in human and 39% of genes in mouse have alternative promoters. For the unannotated promoters, the present version of the database provides supporting evidence collected from spliced ESTs and CAGE promoters (FANTOM4 project). CancerPromDb is a web-based application with many layers: the core application (designed in Django), a backend database (MySQL), a visualization component (Gbrowse), and a web server (Apache). Users can search the database based on Entrez gene id, gene symbol, or by tissue/cell specific activity, and results can then be filtered based on any combination of tissue/cell specificity, Known/Novel, CpG/NonCpG, Coding/NonCoding promoters, etc. It also provides multiple visualization and downloading options. At present, the Gbrowse show RNA Pol-II and H3K4me3 ChIP-Seq wiggle profiles for embryonic stem cells and various cancer cell lines (Table 1). The currently released version of CancerPromDb 1.0 can be accessed at http://bioinformatics.wistar.upenn.edu/CancerPromDb/.

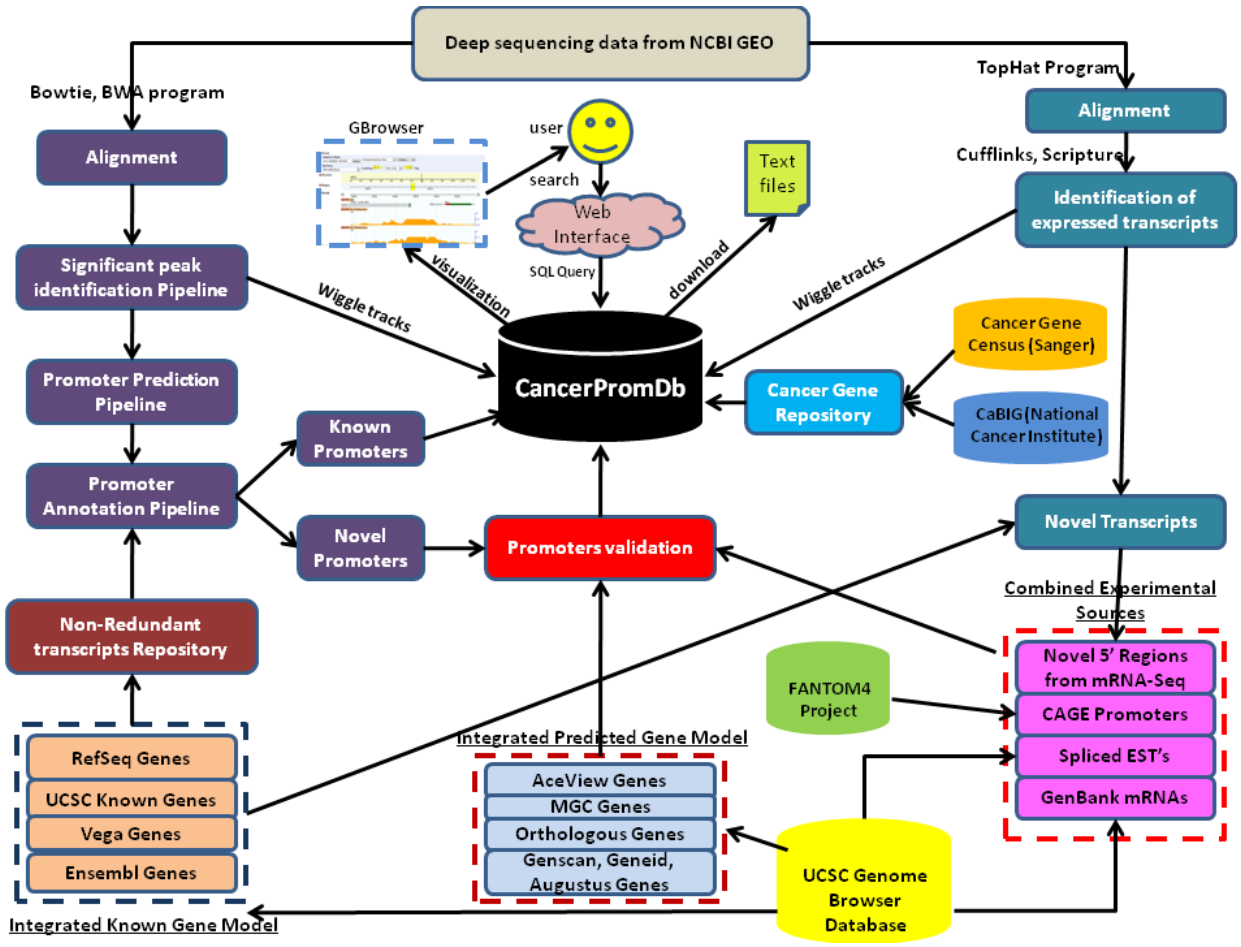

Figure 1: The basic block diagram and workflow of CancerPromDb 1.0 database. Deep sequencing datasets were downloaded from the NCBI GEO server and processed by our analysis and annotation pipeline. The identified promoters are deposited in CancerPromDb relational tables. Novel promoters are compared to various existing experimental and predicted gene promoter regions, and the status of novel promoters is deposited in the relational tables. The database is accessed through a user friendly webpage. The database is integrated with open source genome browser (Gbrowse) to visualize the promoter and various ChIP-Seq enrichment profiles.

Table 1: ChIP-Seq and mRNA-seq datasets used to generate promoter annotation information in CancerPromDb

RNAP-II Chip-Seq Datasets:

Tissues/Cell Lines	GEO Source IDs
Hela	SRR014999 - SRR015009
K562	SRR006807
NB4	SRR016107-SRR016113
CD4+T	SRR029419,SRR029420, SRR037859-SRR037861
LnCaP	GSM353617
VCaP	GSM353607

H3K4me3 ChIP-Seq Datasets:

Tissues/Cell Lines	GEO Source IDs
Hela	SRR039066
LnCaP	GSM353613
hES: H1	SRR018232, SRR018455, SRR020515, SRR029609
hES: H9	SRR015123, SRR015124, SRR015125
CD4+T	SRR029409
Hs68-FBS	SRR020057
IMR90	SRR029610, SRR029618
Sperm	SRR019953, SRR019954, SRR019955
VCaP	GSM353603

mRNA-Seq Datasets:

Tissues/Cell Lines	GEO Source IDs
Brain	SRR015256-SRR015262, SRR036966
Liver	SRR015293-SRR015299, SRR014264
Heart	SRR015300-SRR015306
skeletal muscle	SRR015307-SRR015313
Colon	SRR015314-SRR015320
Adipose	SRR015321-SRR015327
Breast	SRR015270-SRR015273
CD4+T	SRR017233-SRR017240
B Cell	SRR023855, SRR023856
HEK Cell	SRR023853, SRR023854
Jurkat	SRR017241-SRR017247
Lymphnode	SRR015335-SRR015341
HME	SRR015263-SRR015269
MCF7	SRR015274-SRR015277
MB435	SRR015278-SRR015281
BT474	SRR015342-SRR015347

Research Project 2: Project Title and Purpose

Characterizing Mechanisms of Transcriptional Activation Using Live Cell Imaging - Numerous diseases, including cancer, are caused by aberrations in transcriptional regulation. Gene activation requires the tight coordination of regulatory factors at the site of transcription. We propose to study the recruitment dynamics of these factors at a transcription site in single living cells in order to produce a high-resolution map of their recruitment order. A comprehensive understanding of transcriptional regulation is important for the development of new strategies to cure human diseases.

Duration of Project

1/1/2010 - 6/30/2011

Project Overview

We will investigate the recruitment dynamics of transcriptional regulatory factors to a transcription site in single living cells during activation because the exact timing and order of recruitment is not known and cannot be fully determined using biochemical techniques alone. We have developed a method to visualize the activation of a chromatinized transcription site in single living cells with high temporal and spatial resolution. We plan to use it to better understand the temporal mechanisms regulating transcription initiation. It is our goal to use this methodology to develop a high-resolution regulatory pathway map of transcriptional activation, much like what has been done in the signal transduction field. These studies will also provide new insight into the kinetic control of human transcription activation pathways.

In order to achieve these goals, we propose the following specific aims:

Aim 1. Interrogate by shRNA depletion the requirement of known regulatory factors on (i) chromatin decondensation, (ii) pre-initiation complex assembly and (iii) RNA synthesis, using kinetic, live cell imaging of single cells.

Aim 2. Use p53 as a model transcriptional activator to interrogate the timing of its activation kinetics using live cell imaging analysis.

Principal Investigator

Susan M. Janicki, PhD
Assistant Professor
The Wistar Institute
3601 Spruce Street
Philadelphia, PA 19104

Other Participating Researchers

Ilona Rafalska-Metcalf, PhD, Alyshia Newhart, BS - employed by The Wistar Institute

Expected Research Outcomes and Benefits

We will use shRNA depletion to evaluate the requirement of known regulatory factors on chromatin decondensation, pre-initiation complex assembly and RNA synthesis. We will also determine the kinetic and spatial dynamics of post-translational modifications (PTM) and PTM binding proteins during transcription activation and mitosis. In order to expand the use of the live cell imaging methodology, we also propose to use p53 as a model transcriptional activator and to interrogate the timing of its activation kinetics. These studies are significant because they cannot be done using other existing techniques that average effects in cell populations. Additionally, they will provide a portrait of unperturbed gene regulation in single cells. A comprehensive understanding of transcriptional regulation is important for the development of new strategies to cure human diseases.

Summary of Research Completed

Results

Activator Strength Regulates the Degree of Activation

Ribonucleic Acid (RNA) levels increase simultaneously with the activator, Cherry-tTA-ER, at the transcription site following transcriptional activation. This suggests that the VP16 activator can rapidly overcome the heterochromatic repression at the site and initiate transcription. However, as there are 96 activator-binding sites upstream of the Cytomegalovirus (CMV) promoter, it is also possible that the rapid and strong activation is due to the number of VP16 molecules recruited.

To distinguish between these two possibilities, we compared cells expressing Cherry-tTA-ER (VP16 activator) to cells expressing Cherry-TetR-p53-ER, a construct in which the VP16 acidic activation domain was replaced by the first 70 amino acids of p53, the p53 transactivation domain (p53TAD) (Fig. 1A). Western blotting of flag-tagged versions of these constructs showed that they are comparably expressed (Fig. 1B). Similar to the VP16 activator, the p53 activator induced the accumulation of YFP-MS2 at the transcription site (Fig. 1C, panels a-c), as well as the recruitment of GCN5 and Brd2 (Fig. 1C, panels d-i).

In cells activated for three hours, Cherry-tTA-ER and YFP-MS2 accumulated at the transcription site in 87.6 ± 3.4% and 84.7 ± 4.2% of cells, respectively (Fig. 1D). In contrast, Cherry-TetR-p53-ER and RNA were detected at the site in only 32.9 ± 3.0 and 36.2 ± 6.3% of cells, respectively (Fig. 1D). Similar results were seen when GCN5 and Brd2 recruitment were evaluated (Fig. 1E). These results, therefore, suggest that activator strength, not binding site number, regulates the percentage of cells in which a gene is activated.

Interestingly, the number of cells activated by VP16 was approximately the same at both the 30-minute and 3-hour time points, suggesting that the decision to activate is made early and that the number does not increase over time (Fig. 1D). The lower rate of p53 activator-induced accumulation is likely due to its lower ability to access the TRE binding sites in the condensed

inactive transcription site. Although the p53 activator cannot be detected at the site 15 minutes after activation, accumulations of YFP-MS2 are detected in 12.7 ± 2.2% of cells, indicating that it is a more sensitive detector than the activator.

Additionally, the VP16 activator induced a significantly higher degree of chromatin decondensation than the p53-activator, as determined by measuring the pixel area of the transcription site (Fig. 1C, panels a, d and g, and 1F). The area occupied by the p53-activator bound transcription site was ~1.4 times larger than the inactive site bound by Cherry-lac repressor, but significantly smaller than the VP16-activated site (Fig. 1F), which is ~6 times the size of the inactive site. CFP-SKL protein levels in p53-activated cells were also lower than in VP16-activated cells, demonstrating that activator strength also affects the quantity of protein synthesized (Fig. 1G). Taken together, these results indicate that the extent of activation and the ability to overcome a condensed chromatin configuration is due largely to activator strength and not to the number of binding sites upstream of the promoter.

Materials and Methods

Plasmid Construction

To make mCherry-TetR-p53-ER, TetR was Polymerase Chain Reactioned (PCRed) from pTet-Off and cloned into pCMV-mCherry-C3 (XhoI/HindIII). The ER hormone domain was PCRed and cloned into pCMV-mCherry-TetR-C3 (SacII/ BamHI) to make pCMV-mCherry-TetR-ER-C3. The p53 TAD (aa 1-70) was PCRed and cloned into pCMV-mCherry-TetR-ER-C3 (EcoRI/KpnI). Flag-mCherry-tTA-ER and Flag-mCherry-TetR-p53-ER were made by PCRing mCherry-tTA-ER and mCherry-TetR-p53-ER and cloning them into p3xFLAG-CMV-10 (NotI /BamHI).

Cell Culture

The U2OS derived cell line, 2-6-3, was cultured in high glucose Glutamax media (Invitrogen) with 10% tetracycline-free Fetal Bovine Serum (FBS) (BD Biosciences), 1% pen/strep and 100 μg/ml hygromycin B. Transfections were done using FuGENE 6 (Roche) according to the manufacturer's protocol or electroporation as previously described. Cells transfected with Cherry-tTA-ER were activated with 1 μM tamoxifen (Sigma).

Imaging

Images were acquired using a Leica DMI 6000 B inverted automated microscope with HCX PL APO 100x/1.40-0.70 oil objective lens using a 457/488/514nm 30mW Argon-Ion laser for Yellow Fluorescent Protein (YFP) imaging, a 561nm/25mW diode laser for mCherry and a 442nm/70mW diode laser for Cyan Fluorescent Protein (CFP) imaging.

For fixed images, cells were plated on 1.5 coverslips, transfected overnight, activated as indicated, and fixed for 15 minutes in 3% Paraformaldehyde (PFA) in 1X Phosphate Buffered Saline (PBS). Coverslips were mounted in antifade fluorescence mounting medium (1 mg/ml p-phenylenediamine, 90% glycerol in PBS, pH 8.0–9.0 adjusted with Sodium Carbonate /

Bicarbonate Buffer, pH 9.2). Image stacks were taken (0.4 μm increments) using Hamamatsu ORCA-AG camera (1 × 1 binning; 1344 × 1024 pixels). Maximum projections were obtained using COMPIX SimplePCI software. Image contrast adjustments were performed using COMPIX SimplePCI and Adobe Photoshop software.

Image Analysis

To determine the area of the locus, a point inside and a point outside but nearby the transcription site were manually selected. A threshold calculated from the difference in intensity (in arbitrary units) between the two points was used to define the boundary of the transcription site. The "imreconstruct" function in Matlab was used to obtain the region of the transcription site with pixels above the threshold value. The number of pixels identified in the region was then summed for the total locus area of the transcription site.

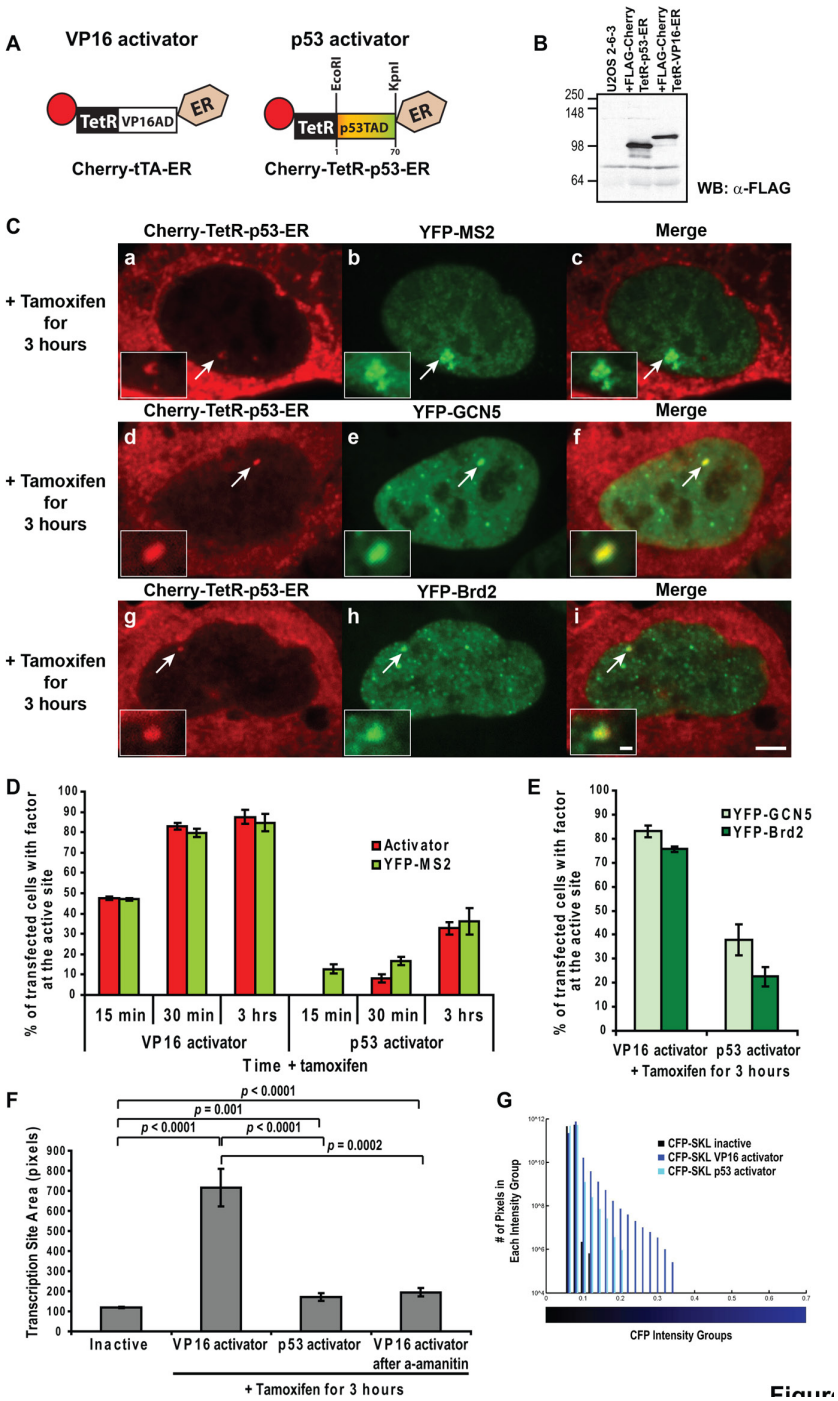

Figure 1. ***Single cell analysis of transcriptional activation induced by VP16 and p53 activators.*** (**A**) Schematic representation of the VP16 and p53 activators. (**B**) Western blot analysis of Flag-tagged versions of the p53 and VP16 activators showing that they are expressed at comparable levels in 2-6-3 cells. (**C**) YFP-MS2 (panels **a-c**), YFP-GCN5 (panels **d-f**), and YFP-Brd2 (panels **g-i**) are enriched at the transcription site, marked by Cherry-tTetR-p53-ER. Scale bar represents 5 µm. Scale bar in enlarged inset represents 1 µm. (**D**) Analysis of the percentage of cells with the activators and YFP-MS2 enriched at the transcription site 15 min, 30 min, and 3 hrs after activation. 100 cells were analyzed from 3 independent experiments; SEMs (in the form of error bars) are presented in the graphs. (**E**) Analysis of the percentage of cells with accumulations of YFP-GCN5 and YFP-Brd2 at the transcription site 3 hrs after activation. 100 cells were analyzed from 3 independent experiments; SEMs (in the form of error bars) are presented in the graphs. (**F**) Measurement of the pixel area of the transcription site in the inactive state (marked by Cherry-lac repressor) and after 3 hrs activation by the VP16 activator (Cherry-tTA-ER with and without α-amanitin pre-treatment) and the p53 activator (Cherry-TetR-p53-ER). SEM (in the form of error bars) and *p* values are presented in the graph. (**G**) Frequency histogram showing the distribution of the blue pixel intensity levels (blue bars) as a measure of the CFP-SKL protein in cells activated for 3 hrs by the VP16 and p53 activators. Black bars represent the background signal. The x-axis is the average fluorescence pixel intensity in each bin on a scale from 0 to 1 and divided into bin sizes of 0.02; the y-axis is the number of pixels in each bin, on a logarithmic scale. The bar beneath the histogram shows the intensity range. Measurements are from 5 independent experiments.

Research Project 3: Project Title and Purpose

Elucidation of the Integrator Composition and Function - This project is focused at elucidating the role of the Integrator, a complex that was discovered by research investigators at The Wistar Institute, in the control of transcription and cellular proliferation. We showed that Integrator is critical for transcription and processing of small nuclear ribonucleic acid (RNA) genes. Elucidation of the mechanism through which small nuclear RNAs are transcribed and processed to mature RNAs is of seminal importance in understanding regulation of gene expression and growth control. Aberrant regulation of small nuclear RNA processing could result in defective messenger RNA processing through deregulation of gene splicing and may lead to cellular transformation and oncogenesis.

Duration of Project

1/1/2010 - 6/30/2011

Project Overview

Small nuclear ribonucleic acids (snRNAs) are components of the spliceosome, a complex assembly of ribonucleoparticles involved in pre-micro RNA (mRNA) processing. The majority of snRNAs (i.e., U1 to U5) are transcribed by RNA polymerase II (RNAPII) to yield short non-polyadenylated 3'-extended precursors. These precursors are then exported to the cytoplasm for further 3' trimming and incorporation into the small nuclear ribonucleoparticles (snRNPs). The formation of proper snRNA precursors depends on a *cis*-acting sequence, named the 3' box, located 9-19 nucleotides downstream of the 3' end of the mature snRNA and on the presence of an snRNA compatible promoter at the 5' end of the gene. Furthermore, the C-terminal domain (CTD) of the largest subunit of RNAPII has been shown to play a crucial role in snRNA 3'-end processing. The identity of the factor(s) that associate with the CTD and result in the processing of the snRNA genes remained elusive for over two decades. Recently, we isolated a multi-protein complex termed the Integrator that associates with the CTD of RNAPII and mediates the 3'-end processing of snRNAs. This project is focused on characterization of the Integrator polypeptide composition and functional elucidation of its mechanism of action.

Specific Aims:

1. *Analysis of subunits of the Integrator and reconstitution of the complex using recombinant components.* Several of the proteins in the Integrator complex do not correspond to genes whose function or expression patterns have been described. Biochemical and immunological analyses will be performed to determine the role of these proteins in regulation of small nuclear RNA transcription and processing.

2. *Identification of the subunit(s) of the Integrator that mediate association with the CTD.* We will examine the role of the CTD phosphorylation in modulating the association of the Integrator complex and RNAPII.

Principal Investigator

Ramin Shiekhattar, PhD
Professor
The Wistar Institute
3601 Spruce Street
Philadelphia, PA 19104

Other Participating Researchers

David Baillat, PhD, Clement Carre, PhD, Kaiping Yan, PhD, Matteo Cesaroni, PhD, Alessandro Gardini, PhD, Derly Acevedo, Fan Lai, PhD – employed by The Wistar Institute

Expected Research Outcomes and Benefits

This project will lead to a greater understanding of the mechanism by which eukaryotic gene expression is regulated. A few years ago, we discovered a novel complex termed the Integrator that mediates the transcription and processing of small nuclear ribonucleic acid (RNA) genes. Small nuclear RNA genes are critical components of the machinery that regulates gene splicing of nearly all human messenger RNAs. Therefore, understanding the scope and mode of action of Integrator will provide us with a greater insight into the precise mechanism by which genes are regulated. This knowledge will be important as aberrant regulation of both transcription and splicing could result in development of cancer.

Summary of Research Completed

To address Milestones 1.1 and 1.2, analysis of individual Integrator subunits and analysis of protein-protein interactions, we isolated the Integrator from the chromatin fraction and began our subunit interaction experiments and global analysis of its genome occupancy.

We initially purified the Integrator complex using a nuclear extract of a stably expressing Flag-INTS10 293T-derived cell line. In order to gain further insight into the composition of the Integrator complex that is associated with the chromatin fraction, we purified Integrator using a new protocol that was described for purification of chromatin bound factors. As expected, we found that Integrator was quite enriched in the chromatin fraction from its stable association with RNAPII; interesting, however, we also were able to obtain a cleaner purification of the complex using chromatin enriched fractions (also called nuclear pellet). The purification enabled us to avoid the excess of over-expressed flag-tagged protein. Flag-INTS10 is now eluted at nearly stoichiometric ratio with the other subunits (Fig. 1a). Mass spectrometric sequencing of this chromatin bound Integrator complex confirmed the subunit composition we had previously established for the Integrator (Fig. 1b). The enriched association of the Integrator with the chromatin persuaded us to analyze its global genome occupancy using chromatin immunoprecipitation followed by high throughput Solexa sequencing. We anticipate a greater role for this complex on RNAPII-mediated transcription beyond that of snRNA transcription and RNA processing.
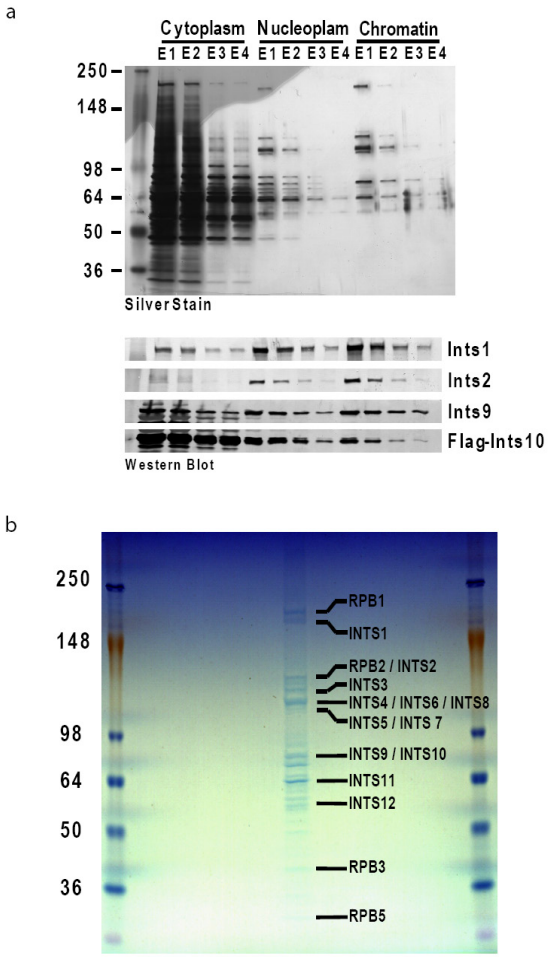
Figure 1. Purification of Integrator from chromatin fraction.
The close association of Integrator and chromatin described above and the recent identification of Integrator in large scale purifications of the spliceosome or the mRNA 3´end processing machinery suggested a broader function for the Integrator complex in RNAPII transcription. To investigate this possibility, we are designing a ChIP-sequencing experiment. A stable 293T cell line stably expressing INTS11 fused to two HA epitopes was established and tested for enrichment of U1 snRNA gene 3´end as a control.