The Wistar Institute of Anatomy and Biology

Annual Progress Report: 2008 Formula Grant

Reporting Period

July 1, 2009 – June 30, 2010

Formula Grant Overview

The Wistar Institute of Anatomy and Biology received $1,461,800 in formula funds for the grant award period January 1, 2009 through June 30, 2010. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

Critical Roles of Transcription Factor Foxp1 in Thymocyte Development and Peripheral T Cell Function - T lymphocytes are one of the two main components of adaptive immunity. Effective immune responses against various infectious diseases require the successful generation of competent T cells, whereas inappropriate differentiation (unsuccessful generation) of T cells can lead to diseases such as those of the autoimmune system. Transcription factors are proteins involved in the regulation of gene expression. We recently demonstrated that transcription factor Foxp1 is essential for T cell development. In the absence of Foxp1, even without an overt antigen challenge, all of the generated T cells have the activated phenotype and have acquired effector functions. Our study of Foxp1 provides a new mouse model to investigate the molecular mechanism(s) underlying T cell development, activation and autoimmunity.

Duration of Project

1/1/2009 - 6/30/2010

Project Overview

The objective of this project is to gain an understanding of the essential roles of transcription factor Foxp1 in T cell development and function.

The specific aims of this project are (1) to elucidate the molecular mechanism(s) by which Foxp1 engages in the thymocyte differentiation, and (2) to further characterize and determine the roles of Foxp1 in peripheral T cell function.

We will establish an *in vitro* thymocyte differentiation model, which will allow us to dissect in detail the molecular mechanism(s) underlying Foxp1 function in thymocyte development. Combining conventional molecular and cellular methods, we will also use a mixed bone marrow chimera approach and generate new mouse lines to further investigate the essential roles of Foxp1 in T cell function.

Principal Investigator

Hui Hu, PhD
Assistant Professor
The Wistar Institute
3601 Spruce Street
Philadelphia, PA 19104

Other Participating Researchers

Xiaoming Feng, PhD, Jessica Willen, Lifeng Tian, PhD - employed by The Wistar Institute

Expected Research Outcomes and Benefits

Our initial study has identified transcription factor Foxp1 as a critical participant in the differentiation of T lineage cells. The detailed analysis of the molecular mechanism(s) underlying Foxp1 functions in thymocyte development will enable us to understand how normal naïve T cells are generated without effector functions. Understanding the principles in generating competent naïve T cells (without abnormal activation/effector function) will help us to understand not only the development and activation of T cells, but also some of the fundamental causes of autoimmune diseases. The further characterization of *Foxp1*$^{-/-}$ T cells in the periphery will help understand T cell homeostasis and survival. Key transcription factors are regulated by external cell signals via specific signaling pathways. Our research on Foxp1 may potentially provide important information for the design of novel and targeted therapies to harness effective immune responses against infectious diseases or treatment for autoimmune diseases.

Summary of Research Completed

1. Studies have shown that activated T cells are more responsive than naïve T cells upon TCR stimulation. We found that higher percentages of Foxp1-deficient T cells up-regulated CD69 and CD25 expression and proliferated than did the control cells to suboptimal stimulation by anti-CD3 antibodies without co-stimulation (**Fig. 1a** and **b**).

Without TCR stimulation, neither Foxp1-deficient nor the control T cells produced any cytokines (data not shown). However, upon transient *ex vivo* stimulation by anti-CD3 plus anti-CD28 antibodies, or by phorbol 12-myristate 13-acetate (PMA) plus ionomycin, higher percentages of Foxp1-deficient CD4^{+} T cells produced IL-2 (**Fig. 1c**), but not IFN-γ or IL-4 (data not shown). Higher percentages of Foxp1-deficient CD8^{+} T cells were found to produce IL-2 and IFN-γ under both stimulation conditions (**Fig. 1c**). These results suggest that Foxp1-deficient T cells in the periphery are not naïve cells masked with an activated phenotype. They respond like activated T cells.

2. To determine whether the abnormal development of Foxp1-deficient T cells is by a cell intrinsic or extrinsic mechanism, we reconstituted lethally irradiated wild-type congenic mice with mixed bone marrow cells from Foxp1$^{f/f}$*Cd4*^{Cre} *and control Foxp1*$^{+/+}$*Cd4*^{Cre} *mice.* Six to ten

weeks after reconstitution, we analyzed the generation and phenotype of donor T cells in the mixed chimeras. In the thymus, the generation and CD4/CD8 profile of Foxp1-deficient thymocytes were normal (**Fig. 2a**). Expression of CD44 and CD62L in both Foxp1-deficient and control DP thymocytes was low/negative (data not shown). However, both Foxp1-deficient CD4 SP and CD8 SP thymocytes had increased expression of CD44 and decreased expression of CD62L (**Fig. 2b**), suggesting that Foxp1-deficient thymocytes acquire an activated phenotype by a cell intrinsic mechanism independent of their environment.

In the periphery of mixed chimeras, Foxp1-deficient T cells exhibited an activated phenotype (**Fig. 2b**). In agreement with the results from *Foxp1*$^{f/f}$*Cd4*^{Cre} mice (data not shown), the percentages and cell numbers of peripheral Foxp1-deficient CD4^{+} T and CD8^{+} T cells in mixed chimeras were reduced (**Fig. 2a** and data not shown), and they produced cytokines rapidly (**Fig. 2c**). In both the thymus and the periphery, Foxp1-deficient T cells (except DP thymocytes) were larger than the control cells (**Fig. 2d**).

3. In an attempt to rescue the cell death of Foxp1-deficient T cells in the periphery, we generated Bcl2^{Tg}*Foxp1*$^{f/f}$*Cd4*^{Cre} *mice.* We found that, although Bcl-2 transgenic Foxp1-deficient CD4^{+} and CD8^{+} T cells also had an activated phenotype, the total numbers of the peripheral T cells in *Bcl2*^{Tg}*Foxp1*$^{f/f}$*Cd4*^{Cre} were still lower than those in the control mice (data not shown), indicating that Bcl-2 transgene may not be able to rescue the cell death of Foxp1-deficient T cells in the periphery.

4. Generation of conditional Foxp1A-transgenic and Foxp1D-transgenic mice. Key features of the conditional Rosa26-locus knock-in construct include:

- A cassette with a stop codon flanked by two *lox*P sites is set in front of the inserted transgene so that the transgene will only be expressed when the stop cassette is deleted by Cre recombinase. Therefore, by using different Cre-deleter mouse strains, the transgene will be expressed in a lineage- and developmental-stage dependent manner
- The inserted transgene is followed by an internal ribosome entry site (IRES) and the sequence encoding the enhanced green fluorescent protein (EGFP); therefore, the cells that actively express the transgene will also express the EGFP as an indicator protein.
- The IRES-EGFP cassette is flanked by frt sites; thus, if expression of EGFP is suspected to affect cell behavior, the EGFP transgene can be deleted with Flp recombinase by crossing the mice with Flp-deleter mice.

Experiments needed to generate the conditional Foxp1A and Foxp1D transgenic mice have been completed. The targeted ES clones for both the Foxp1A and Foxp1D transgenes were screened by southern blot analysis using both 5' and 3' probes (**Fig. 3a**, **b** and data not shown). The positive clones for both transgenic mouse lines were selected and the chimeras were obtained. The founder lines of both Foxp1A transgenic (*Foxp1A*^{Tg}) and Foxp1D transgenic (*Foxp1D*^{Tg}) mice have already achieved germline transmission, and *Foxp1D*^{Tg}*Cd4*^{Cre} mice have been generated (**Fig. 3c**). Thymocytes or peripheral T cells from *Foxp1D*^{Tg} mice do not express any EGFP (data not shown). In the few *Foxp1D*^{Tg}*Cd4*^{Cre} mice we obtained, DP thymocytes started to express EGFP, and almost all SP thymocytes were EGFP positive (**Fig. 3c**), indicating that the stop codon in front of the Foxp1D transgene has been deleted by Cre recombinase under CD4

promoter at DP stage and the conditional knock-in construct is working well. We have started to backcross the Foxp1D transgenic mice to C57BL/6 background and are crossing Foxp1A transgenic mice with *Cd4*^{Cre} mice to generate *Foxp1A*^{Tg}*Cd4*^{Cre} mice.

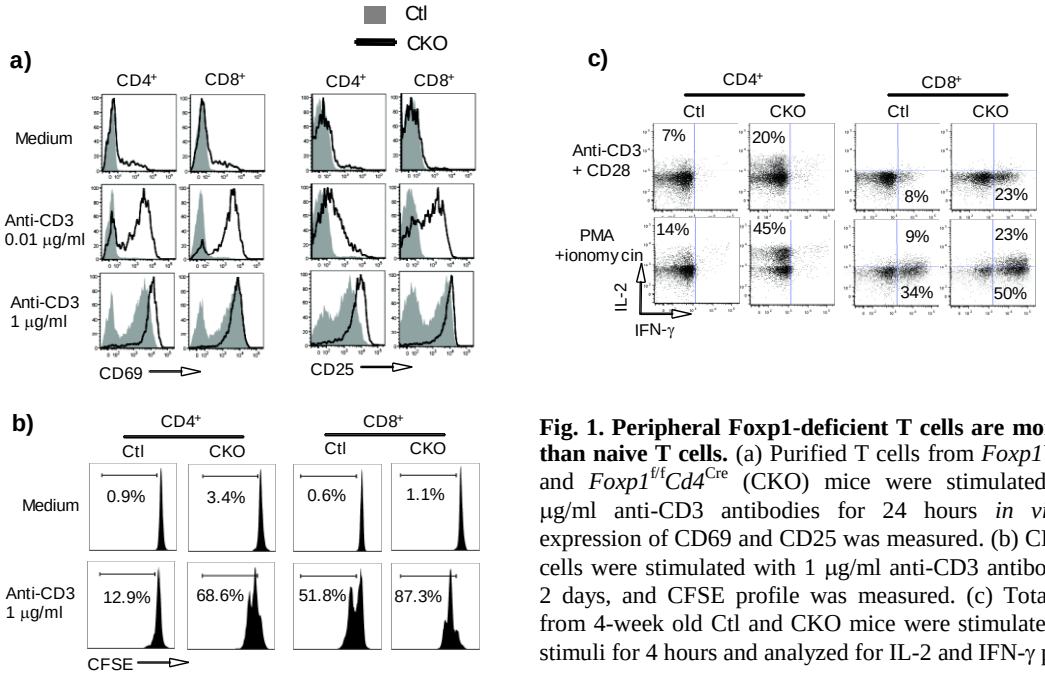

Fig. 1. Peripheral Foxp1-deficient T cells are more responsive than naive T cells. (a) Purified T cells from *Foxp1*$^{+/+}$*Cd4*^{Cre} (Ctl) and *Foxp1*$^{f/f}$*Cd4*^{Cre} (CKO) mice were stimulated with 0.01-1 μg/ml anti-CD3 antibodies for 24 hours *in vitro*, and the expression of CD69 and CD25 was measured. (b) CFSE-labeled T cells were stimulated with 1 μg/ml anti-CD3 antibodies alone for 2 days, and CFSE profile was measured. (c) Total spleen cells from 4-week old Ctl and CKO mice were stimulated by different stimuli for 4 hours and analyzed for IL-2 and IFN-γ production.

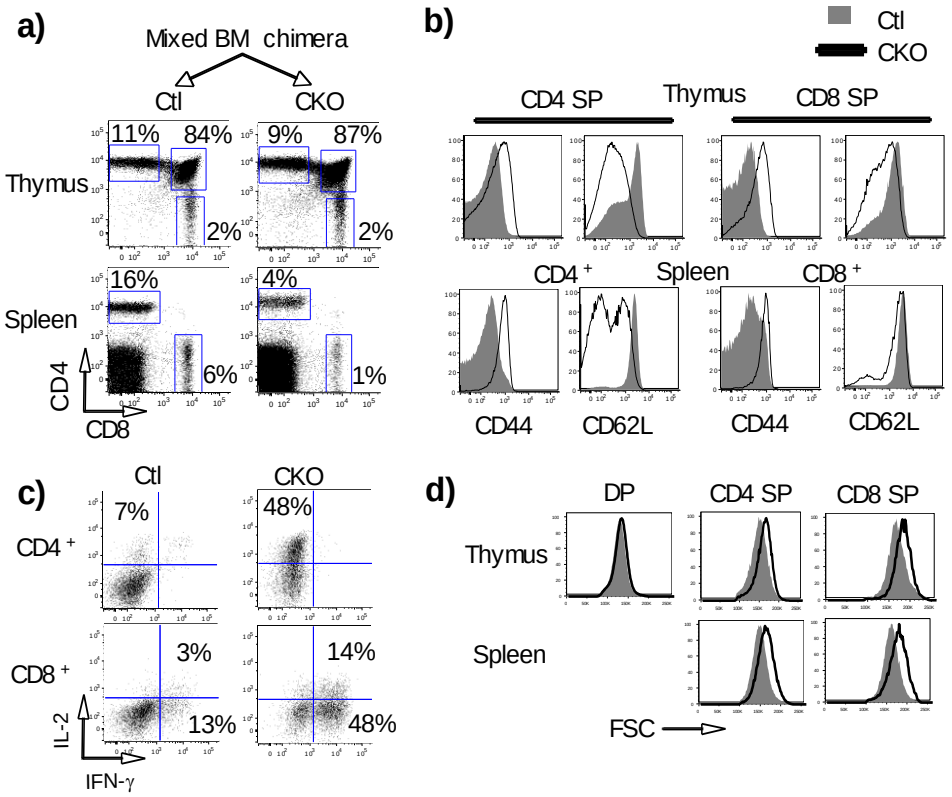

Fig. 2. Foxp1-deficient T cells acquire an activated phenotype by a cell autonomous mechanism. Bone marrow cells from control *Foxp1*$^{+/+}$*Cd4*^{Cre} (Ctl) and CKO mice are mixed to reconstitute the irradiated wild-type recipient mice. Six weeks after reconstitution, gated on donor congenic cell surface markers, the thymocytes and splenic T cells of the recipient mice were analyzed for (a) CD4 and CD8 profiles; (b) expression of CD44 and CD62L; (c) IL-2 and IFN-γ production after PMA plus ionomycin stimulation for 4 h in splenic T cells by intra-cellular staining; (d) cell size.

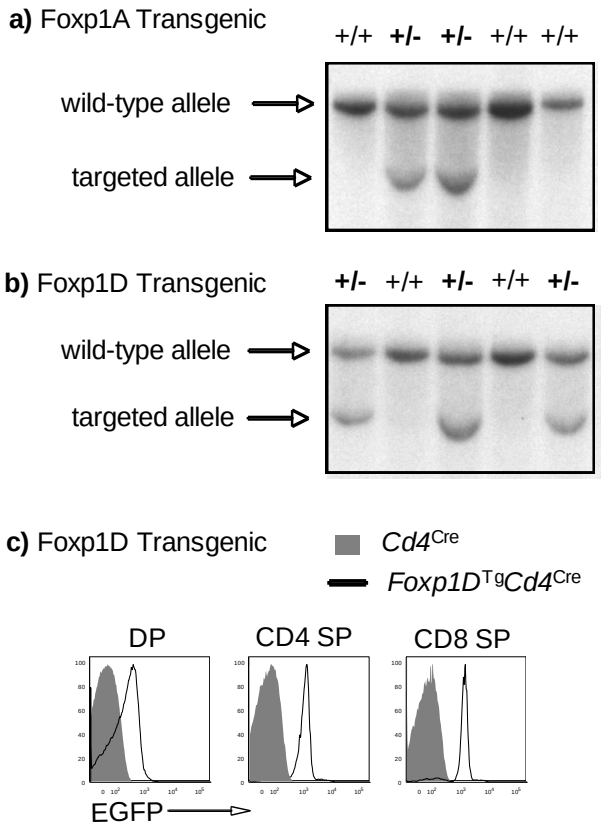

Fig. 3. Successful ES cell targeting of conditional Foxp1A (a) and Foxp1D (b) transgenes and the generation of *Foxp1D*^{Tg}*Cd4*^{Cre} mice (c). (a) Southern blots of Foxp1A transgenic ES cell screening (5' probe) including positive clones were shown. (b) Southern blots of Foxp1D transgenic ES cell screening (5' probe) including positive clones were shown. (c) EGFP expression in DP and SP thymocytes from *Foxp1D*^{Tg}*Cd4*^{Cre} and control *Cd4*^{Cre} mice.

Research Project 2: Project Title and Purpose

Impact of Latent Cytomegalovirus Immediate Early Protein Expression on Embryonic Development - Development of the human fetus can be severely affected by infection with human cytomegalovirus during pregnancy, resulting in mental retardation, vision loss and deafness. To test the hypothesis that proteins produced by latent virus are responsible for the abnormal embryonic development, we will establish and investigate a mouse model system. The work will eventually allow screening in a mouse model for small molecules that can prevent the functions of major immediate early proteins and testing of interference strategies during development.

Duration of Project

1/1/2009 - 6/30/2010

Project Overview

Human cytomegalovirus (HCMV) infection causes life-threatening disease in immunocompromised hosts and has serious neurological consequences for fetal development in women who suffer primary infection during pregnancy. Previous work has shown that HCMV immediate-early protein 1 (IE1-72) is essential for the virus' replication. Thus, blocking the

function of IE1 is expected to inhibit progression of the viral replication program by allowing silencing of the viral genome and appropriate apoptotic activity in cells that are latently infected.

The objective of this project is to determine whether inactivation of the host's repressors and gene-silencing proteins by the IE1 gene products affect differentiating neuronal cells of the developing embryo.

To test our hypotheses, we will *determine whether inactivation of HDAC and Daxx-mediated activity by IE1 alters the differentiation program to induce abnormal neuronal development during embryogenesis.*

We will evaluate the effect of IE1 and IE3 on differentiating brain cells, using a system (Cre/lox recombinase) that enables conditional expression of the major immediate early proteins during different stages of mouse development and in large parts of the brain. These studies will provide further information on the molecular mechanism of CMV-induced abnormalities and on the relative importance of blocking the major immediate early protein functions in congenital infection and also in reactivation circumstances.

This project is designed to test the novel concept that the viral gene product IE1 is essential in circumventing host defenses and is detrimental during fetal development. The expected outcome of our experiments will identify therapeutically useful mechanisms. Our studies will shift current emphasis on antiviral interference strategies that focus on viral replication to IE-related events preceding replication. HCMV-infected embryos and immunocompromised individuals are especially likely to benefit from redefined therapeutic strategies.

Principal Investigator

Gerd G. Maul, PhD
Professor
The Wistar Institute
3601 Spruce Street
Philadelphia, PA 19104

Other Participating Researchers

Dmitri Negorev, PhD - employed by The Wistar Institute

Expected Research Outcomes and Benefits

Our research will establish a mouse model that can be used to test the efficacy of blocking compounds. *Such effective blocks will prevent progression to the viral replicative phase and prevent interference with the differentiation program of latently infected cells. Importantly, appropriate strategies for interference during congenital infection can then be promulgated to prevent the detrimental effects on neuronal cells.* The expected findings will imply a paradigm shift in the search for appropriate drugs away from the focus on viral replication and toward IE-related preceding events. Future application of such drugs promises to dramatically improve the

chances that healthy babies will be born to cytomegalovirus-infected mothers, with all the attendant cost savings of postnatal care and increases in quality of life. As a test system, such engineered mice will also be of enormous value for testing the efficacy of pharmacologically intervening in the reactivation program during tissue transplantation.

Summary of Research Completed

Analysis of the effect of IE1 expression on mouse brain development.

In order to analyze effects of mouse cytomegalovirus (MCMV) IE1/3 on the mouse brain, we first needed to breed to homozygosity the heterozygotic transgenic mice developed during the first six months of this project (first year's milestone). Since the viral gene had been placed into the Hprt locus of the X-chromosome, genotyping for IE1 and Hprt was used to identify homozygotic animals from which a colony was established.

For initial characterization, we crossed these mice with nestin-CreER recombinase mice. Nestin is activated in the early mammalian neural tube, and we expected damage to the embryos' development. Except for a slightly skewed sex ratio, we evaluated 10 litters and found no obvious defects.

Suspecting low penetration of tamoxifen-based activation of the nestin promoter, we crossed the IE1/ mice with the pax3 promoter activated Cre transgenics. Pax3 is activated very early in development of the neural tube. Six litters of these double transgenics were evaluated. The cross as set up should have resulted in half control pups with IE1/3 untranslated and half Pax3-IE1/3 pups with IE1/3 activated. We found a lower than expected litter size and slightly increased embryonic death of the experimental genotype. Neither observation was significant, i.e., no obvious effect was recognized from this cross.

Not knowing which cell type allows expression of the viral promoter, we evaluated the ubiquitin (universal expression of Cre) activated IE1/3 expression of the major immediate early MCMV gene. IE1/3 mice were crossed with ubiquitin activated Cre-ER mice. Timed pregnancy injection of tamoxifen had unexpected results. We found no birth when the pregnant female was injected during the first 8 days of pregnancy, strongly variable number of pups between injections on day 9-12, and early or strongly delayed birth on injections after day 12. The result indicated a potential effect of the viral genes on development. However, pregnant mice had a higher death rate (4 out of 20). Injected non-pregnant mice showed no effect. This result indicated that we could not control for the effect of tamoxifen on the pregnant mouse.

To eliminate the potential tamoxifen effect on pregnant mice, we eliminated the IE1/3 stop codons. We injected males with tamoxifen and waited 38 days until sperm from the time of injection had matured and supposedly had gone through the tamoxifen induced Cre recombination. These males transmitted the open normal IE1/3 gene, i.e., either IE1 or IE3 protein could be expressed as long as the cellular context allowed the IE1/3 promoter to be active. Healthy-appearing pups were born. This indicates that as yet unknown expression patterns during embryonic development in mice show no substantial effect.

The question if and when IE1/3 is expressed was tested in adult mice. Images are presented at the end of the text. Frozen sections of various organs were evaluated for the presence of the GFP labeled viral proteins.

In the brain we find green marked Purkinje cells outlining the dense granular areas imaged by the DNA staining DAPI (see Fig. 1). Only in the cerebellum could such clear and consistent expression be observed. In the cortex, a few (< 1/10.000) cells glowed green. TNFα – used to imitate a cytokine shock during transplantation – resulted in additional cells of the meninges to actively produce the viral protein (Fig.2). In addition, another at present unknown cell type is actively expressing GFP-IE1/3 (Fig. 2).

The other organ with substantial IE1/3 expression was the kidney. There, TNFα produced expression of IE1/3 in the proximal tubules which were clear without the cytokine injection (Fig. 3). No expression was found in the distal tubules. Surprisingly, a TNFa independent high expression was found in an as-yet-unidentified kidney associated tissue type. This clearly shows that only specific cell types can express GFP-IE1/3 with important consequences for virus replication latency and reactivation potential.

In the lungs, viral protein expression was found only in the epithelial cells of the bronchi (Fig. 5). Surprisingly, these cells showed differential expression of IE1 or 3. IE1 has a diffuse distribution in the nucleus, whereas IE3 has an aggregated distribution. Cells in close proximity showed either the IE1 or IE3 phenotype (see higher magnification images in Fig. 6 and 7).

In the liver, no expression was evident through fluorescent microscopy. Expression in other organs and their positive cell types remain to be identified.

A general conclusion is that the viral proteins IE1 and IE3 are expressed in specific cell types and in organ specific regions, some of which can be activated through specific stress. It seems significant that not all cells of the same type express the viral protein within a given organ; rather, different cell types have strongly varying ratios of active expression. This transgenic mouse developed by us is clearly an important resource to define which cell types can silence the viral promoter and which cell types keep the promoter open all the time. In addition, the ability to visually detect the cell types that are activated upon the cytokine storm unleashed during kidney or any other transplantation is significant for transplantation research.

One of the reasons that the expression of IE1/3 has no effect on the mouse during development could be that, during establishment of the mouse or during development of the embryonic stem cells to produce these transgenics, IE1 or IE3 gene copies were selected that expressed non-functional proteins. To test this possibility, we produced embryonic fibroblasts from mice where the stop sequence had been removed through germ cell transmission. Approximately 3% of these fibroblasts express IE1/3. That demonstrates that most of the cells can silence the viral promoter.

This system provides a great opportunity to determine how the promoter is silenced (protein acetylation or methylation or DNA methylation) and whether there is stochastic activation or if certain cells remain permissive over a long time (time lapse studies). This result also indicates

that the intrinsic cellular defenses are successfully silencing most of the viral genes. How a stochastic activation is achieved can be investigated with these cells.

A trivial explanation for the inability to influence the mouse brain development could have been the selective loss of functionality of either IE1 or IE3. The IE1 expressing fibroblasts were, therefore, tested as to the function of IE1, i.e., the dispersion of ND10 or PML bodies using immunostaining for PML. None of the GFP-IE1 expressing cells had the typical ND10/PML body dots. This clearly shows that at least one observable effect of IE1 is that it is functioning as in the wild type situation. The function of IE3 is its activation of the early protein E1. This activation is being investigated.

The produced transgenic mouse failed to provide an easy assay system for problems in brain development due to cytomegalovirus infection. Subtle defects have not been investigated due to fiscal and time constraints; however, these might exist and will be further investigated by those parties who specialize in this. Distribution of this resource animal is in progress.

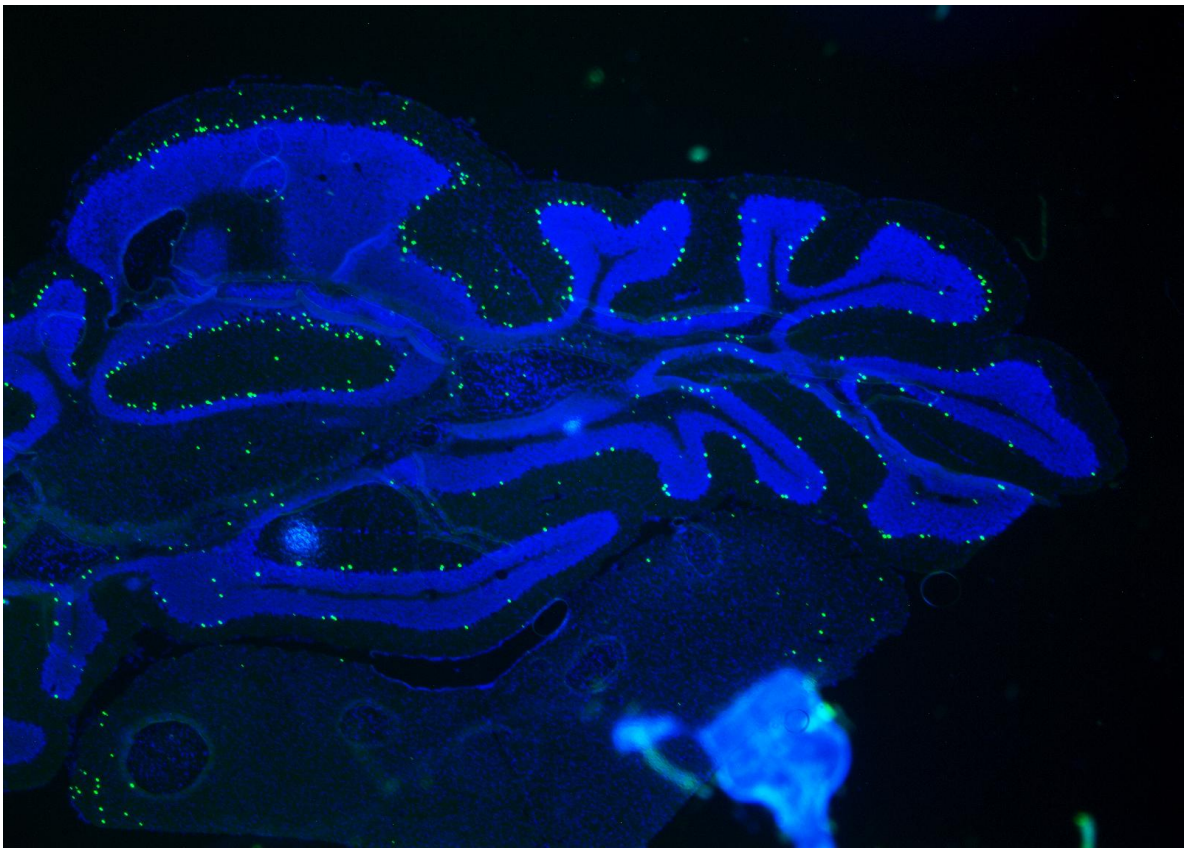

Figure 1. Adult mouse cerebellum frozen sectioned and stained to show nuclei with DAPI.
Green dots surrounding the granular layer are IE1/3 expressing Purkinje cells.

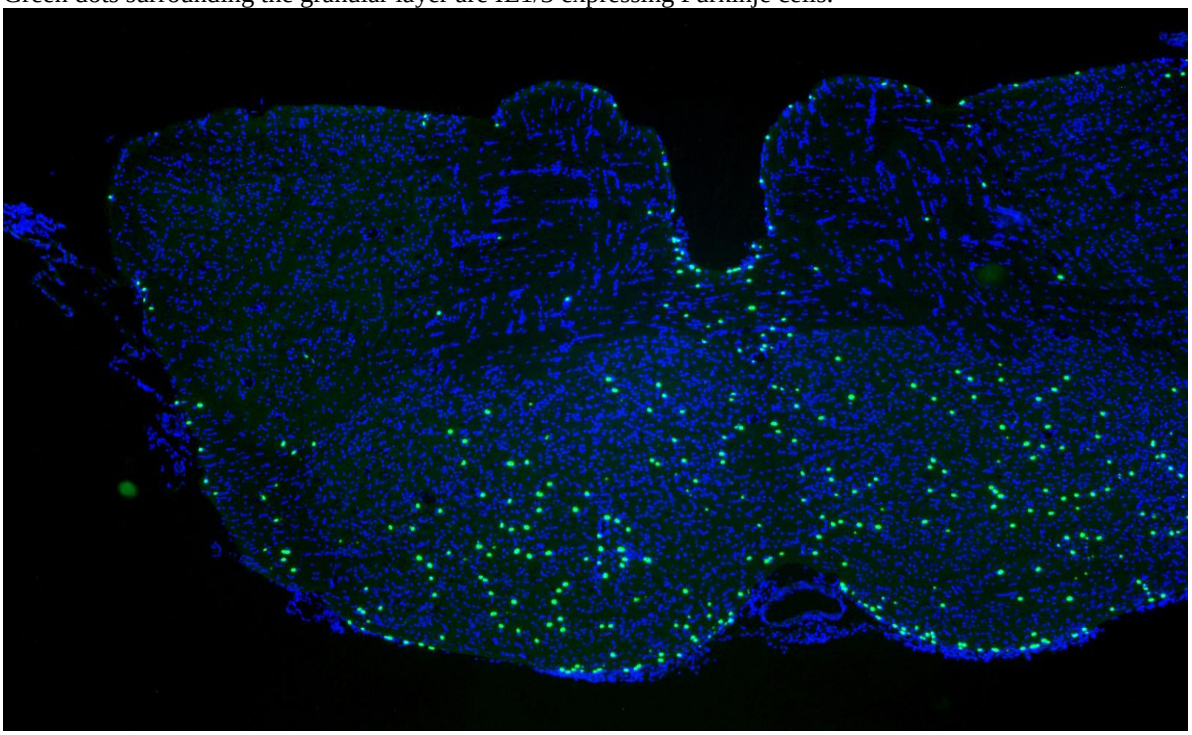

Figure 2. Cross section of mouse brain stem. Cells of the meninges express GFP-IE1/3 and an as-yet cluster of unidentified cells.

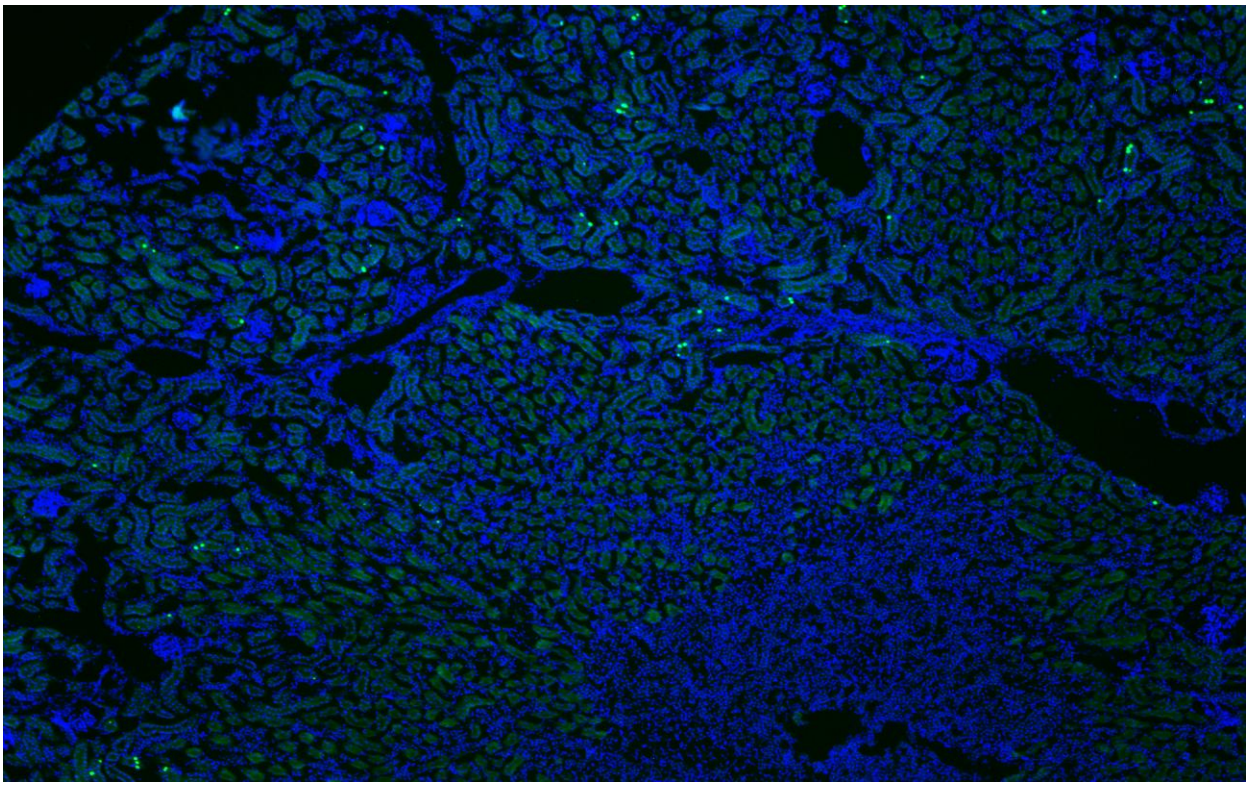

Figure 3. Frozen sectioned mouse kidney stained for DNA by DAPI. Some proximal tubule cell nuclei express GFP-IE1/3 after TNDa exposure (top and left outer layer). Distal tubular nuclei (lower right) do not show any viral protein expression.

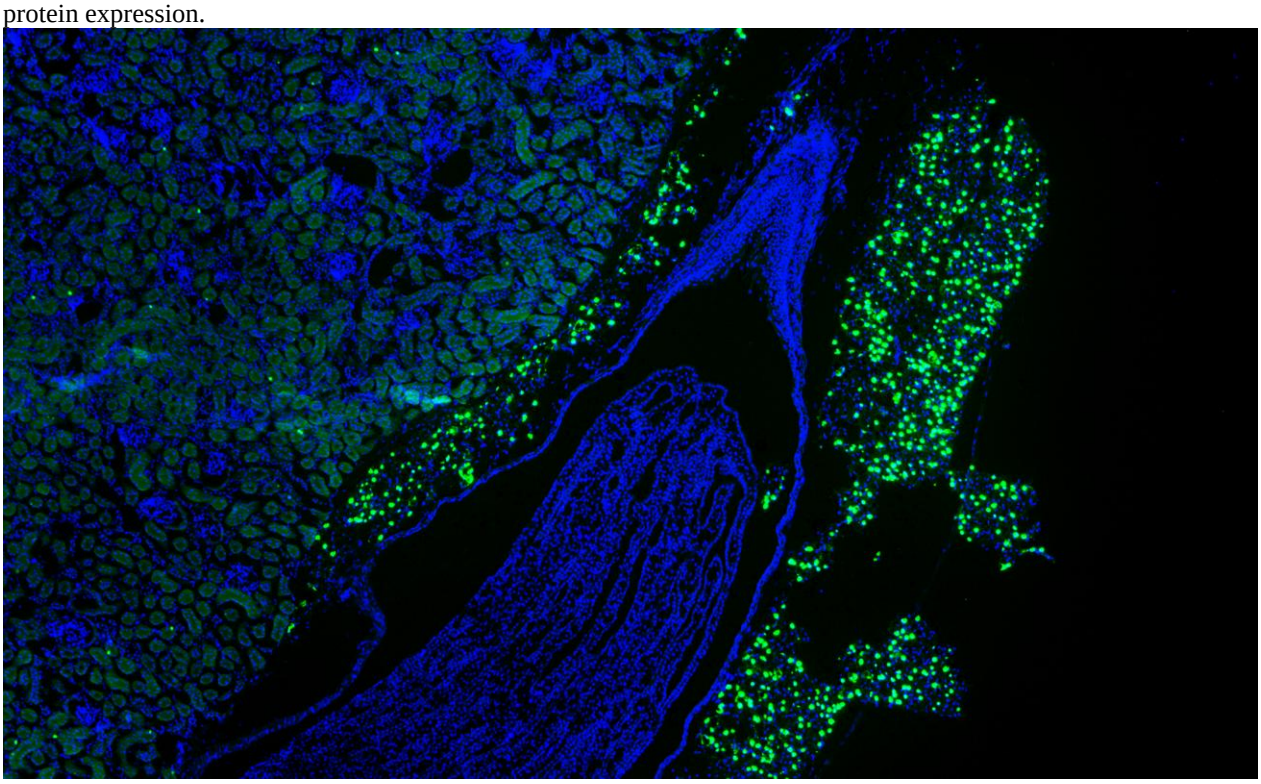

Figure 4. Section as above but showing specific high expressing nuclei at outer as yet unidentified accessory kidney tissue, which is independent of TNFa exposure.

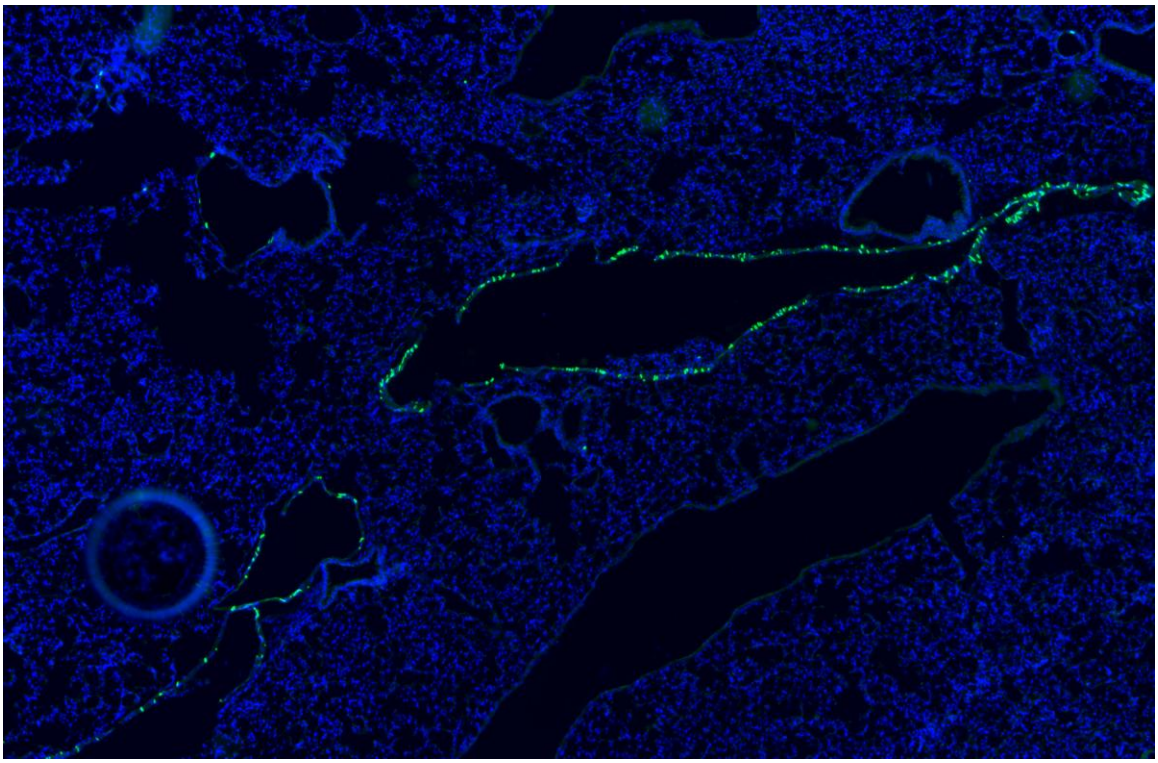

Figure 5. Frozen sectioned mouse lung stained for DNA by DAPI. Only bronchi have an epithelial expression of GFP-IE1/3.

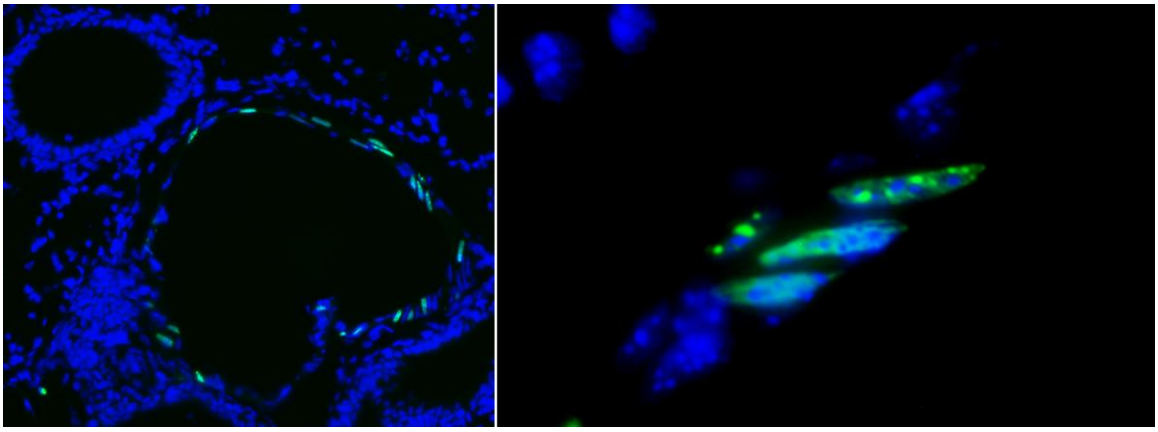

Figure 6 and 7. Higher magnification shows that not all the epithelial cells are expressing IE1/3 and closely situated cells express IE1 (diffuse) or IE3 (granular).

Research Project 3: Project Title and Purpose

Frequency, Fate and Significance of MicroRNA Editing - Within the last decade a new type of ribonucleic acid (RNA) was discovered, microRNA (miRNA), which plays a major role in the regulation of gene expression. miRNAs are produced as larger gene transcripts (pri-miRNA) which are subsequently processed to mature miRNA. We found that pri-miRNAs are subject to editing by conversion of adenosine residues to inosine residues (A→I editing). Editing of pri-miRNAs results in inhibition of their processing and alters the genes targeted by the unedited

miRNAs, resulting in naturally occurring "off-targeting effects". The overall goal of this project is to gain a better understanding of the biological significance of miRNA editing.

Duration of Project

1/1/2009 - 6/30/2010

Project Overview

The broad, long-term goal of this project is to increase our understanding of ribonucleic acid interference (RNAi) biology and improve applications of the microRNA (miRNA)-based RNAi technology. Our recent studies have revealed that primary transcripts of certain miRNA genes (pri-miRNAs) are subject to one type of RNA editing that converts adenosine residues into inosine in double-stranded RNA (dsRNA). Editing of pri-miRNAs results in inhibition of their processing to mature miRNAs or expression of edited mature miRNAs and silencing of a set of genes different from those targeted by the unedited miRNAs, resulting in naturally occurring "off-targeting effects" of miRNA-based RNAi gene silencing. It is currently not known how widely pri-miRNA editing takes place. In this project, therefore, we will first estimate the A→I editing frequency of mammalian and viral pri-miRNA, determine its biological significance, and assess its "off-targeting effects". A search through an miRNA database identified a large number of mammalian and viral candidate pri-miRNAs containing a "preferred" A→I editing site within a consensus sequence predicted from a set of pri-miRNAs proven to have editing sites. These candidate pri-miRNAs will be tested experimentally in order to determine what fraction of mammalian and viral pri-miRNAs are edited. The effects of editing on miRNA processing or expression of the edited miRNAs will be evaluated by subjecting these newly identified pri-miRNAs containing specific A→I editing sites to an *in vitro* miRNA processing assay. Targeted cloning and identification of edited mature miRNAs that would be processed from edited pri-miRNAs will be attempted.

The specific aims of this project are:

Specific Aim 1. Survey miRNA gene transcripts for A→I RNA editing and identify edited miRNAs in order to determine the overall frequency of mammalian and viral pri-miRNA editing.

Specific Aim 2. Determine the effects of the pri-miRNA editing *in vitro* by miRNA processing assay using recombinant Drosha-DGCR8 and Dicer-TRBP complexes.

Principal Investigator

Kazuko Nishikura, PhD
Professor
The Wistar Institute
3601 Spruce Street
Philadelphia, PA 19104

Other Participating Researchers

None

Expected Research Outcomes and Benefits

Since its discovery in 1998 RNA interference (RNAi) has become a major focus of biomedical research and its powerful gene silencing effects have been demonstrated in recent years. More recently, miRNA-mediated gene silencing has become the technique of choice for future gene therapy for viral infection and human diseases, including cancer. Furthermore, recent studies indicate that mutation and mis-expression of miRNAs correlate with various human diseases, including cancer. Outcomes of this project will reveal how frequently mammalian and viral miRNAs are subjected to A→I RNA editing, as well as the biological significance of miRNA editing. Thus, this project has relevance to cancer biology. In addition, information obtained in this project has the potential to reveal the factors determining the efficacy of RNAi technology and its application for future gene therapy.

Summary of Research Completed

Overview

During the current period, we completed our second milestone, *Specific Aim 2*: Determine the effects of the pri-miRNA editing *in vitro* by miRNA processing assay using recombinant Drosha-DGCR8 and Dicer-TRBP complexes. We found that a majority of pri-miRNA editing is likely to interfere with the miRNA processing steps. In addition, four new edited miRNAs with altered seed sequences were identified by targeted cloning and sequencing of the miRNAs that would be processed from edited pri-miRNAs. Thus, miRNA editing could have a large impact on the miRNA-mediated gene silencing.

Results

Expression of edited mature miRNAs

Many pri-miRNA editing sites were located within the mature miRNA sequence or corresponding partner miRNA sequence in the opposite strand, which could be processed to edited mature miRNAs as we previously reported for miR-376 cluster RNAs. We applied the targeted cloning strategy as described previously, attempting to isolate cDNA clones corresponding to all of these potential edited mature miRNAs. Only four mature miRNAs containing U<u>A</u>G and/or non-U<u>A</u>G editing sites not previously reported were identified – miR-379-5p, miR-411-5p, miR-607-3p, and miR-99b-3p (Figure 1). Interestingly, both miR-379 and miR-411 are members of the miR-379 cluster. Simultaneous expression of several edited mature miRNAs of the miR-376 cluster has been reported, indicating the possibility of simultaneous editing of clustered miRNAs. The A→I editing sites identified in these mature miRNAs were located in the seed sequence known to play a major role in the hybridization of a miRNA with the target site, thus affecting selection of the target genes. As we have demonstrated for edited miR-376 RNAs, application of target prediction programs is likely to

reveal that these newly identified edited mature miRNAs target genes are different from those targeted by the unedited miRNAs.

A recent global survey for miRNAs from various human and mouse tissues identified a large number of miRNAs that had nucleotide changes within the mature miRNA sequence, including those already reported previously as edited miRNAs such as miR-368, miR-376a1, and miR-376a2. However, four edited miRNAs identified in the present study, as well as those identified previously (hs-miR-376b, mu-miR-376a, mu-miR-376b, and mu-miR-376c) by the targeted cloning strategy, were not included in the list of the miRNAs with A→G changes determined by the unbiased cloning strategy, indicating a limitation of the unbiased cloning strategy for detection of certain miRNA editing sites. Nevertheless, this deep sequencing study for unbiased expression profiling identified a few new miRNAs that had A→G nucleotide changes (potential A→I editing) at a frequency higher than 5%, i.e., hs-19b-2, hs-miR-512-1 and hs-miR-512-2, mu-miR-17, mu-miR-450b, and mu-miR-483. We tested some of these miRNAs containing the A→G sequence changes independently to see whether they are A→I edited miRNAs. Through analysis of the pri-miRNA sequence corresponding to the mature miRNAs with A→G changes, we found that the reported sequence heterogeneity of hs-miR-19b-2 and mu-miR-17 is not due to A→I editing, indicating that they are most likely sequencing errors. Together, expression of edited mature miRNAs appears to be relatively rare, although a large fraction of human pri-miRNAs undergoes editing at the sequence corresponding to mature miRNAs. Thus, processing of a majority of edited pri-miRNAs must be suppressed as we previously reported for miR-142 and miR-151.

Processing of edited pri-miRNAs

We then investigated the effects of miRNA editing on processing of pri- to pre- or pre- to mature miRNAs by subjecting unedited and edited pri-miRNAs to an *in vitro* processing assay using purified Drosha-DGCR8 or Dicer-TRBP complexes as described previously. We randomly selected six pri-miRNAs, i.e. pri-let-7g, pri-miR-33, pri-miR-133a2, pri-miR-197, pri-miR-203, and pri-miR-379, among 43 pri-miRNAs that have at least one U<u>A</u>G triplet editing site (Figure 2). None of the edited mature miRNAs expected to be processed from pri-let-7g and pri-miR-33 were detected, although editing sites were located within the mature miRNA sequences of these pri-miRNAs. Although the edited miR-379 RNAs were detected (Figure 1B), the frequency of detection of edited mature miR-379 (15%) versus unedited miR-379 was much lower than that expected from the editing frequency of pri-miR-379 (60%), indicating that processing of edited pri-miR-379 RNAs may not be as efficient as that of unedited pri-miR-379 RNAs. We also noted this proportionally less frequent detection of edited mature miR-411 RNAs (Figure 1C). Furthermore, neither unedited nor edited miR-203 was detected in human brain, perhaps due to their low expression levels. We tested a set of unedited and "pre-edited" pri-miRNAs for *in vitro* processing assays. In the "pre-edited" pri-miRNA used for the assay, the adenosine residue of the editing site was replaced with guanosine. We previously demonstrated that the A→G alteration of pri-miRNA is treated the same as A→I editing by Drosha-DGCR8 complexes and also by the Dicer-TRBP complexes.

Both +11 and +14 sites of pri-let-7g were edited at 10% and 30% *in vivo* in human brain. The effects of editing at the major +14 site were tested for *in vitro* processing (Figure 2A). Although the editing at the +14 site did not affect the Drosha cleavage step, it suppressed the Dicer

cleavage step, leading to 2-fold reduction in the mature miRNA levels. By contrast, editing of pri-miR-33 at the +15 site, pri-miR-133a2 at the +20 site, and pri-miR-379 at the +10 site inhibited the Drosha cleavage step and reduced the generation of their precursor forms (pre-miRNAs) down to a 50% level as compared to their unedited controls (Figure 2B, C, F). This is likely the reason for less frequent detection of edited mature miR-379 RNAs (Figure 1B). Editing of these three pri-miRNAs had no effect on the Dicer cleavage step (Figure 2B, C, F). Interestingly, editing of pri-miR-203 at the +85 site enhanced its processing by the Drosha-DGCR8 complex, leading to 2-fold increase in generation of pre-miR-203 RNAs (Figure 2E). This is the first example of pri-miRNA editing that results in promotion of processing. Similarly, a slight increase in processing of pri-miR-197 edited at the +14 site by the Drosha-DGCR8 complex was detected (Figure 2D). Editing of both pri-miR-203 and pri-miR-197 did not affect the Dicer cleavage step (Figure 2D, E). The editing site of pri-miR-203 is located within the mature miRNA sequence (Figure 2E). Thus, although not detected in human brain investigated in this study, it is possible that edited miR-203 RNAs may be detected in the tissues where they are highly expressed. Interestingly, it has been recently reported that blockage of pri-let-7g processing by Lin28 RNA binding protein plays an important role in the miRNA-mediated differentiation in stem cells. Furthermore, a critical role played by miR-203 and miR-133a2 in differentiation and proliferation of skin progenitor cells and cardiomyocytes, respectively, has been reported. It remains to be investigated whether A→I editing of let-7g, miR-133a2, and miR-203 affect the cellular process proposed to be regulated by these miRNAs. Although only a limited number of pri-miRNA editing sites were tested, all six examples investigated were found to affect the processing of miRNAs, indicating that the major function of pri-miRNA editing is modulation of miRNA biogenesis and expression levels rather than generation of edited mature miRNAs and redirection of target gene selection. Finally, editing of pri-mRNAs may affect the processing steps not tested in this study such as nuclear export as well as degradation of edited pri-miRNAs by the ribonuclease specific to inosine-containing RNAs.

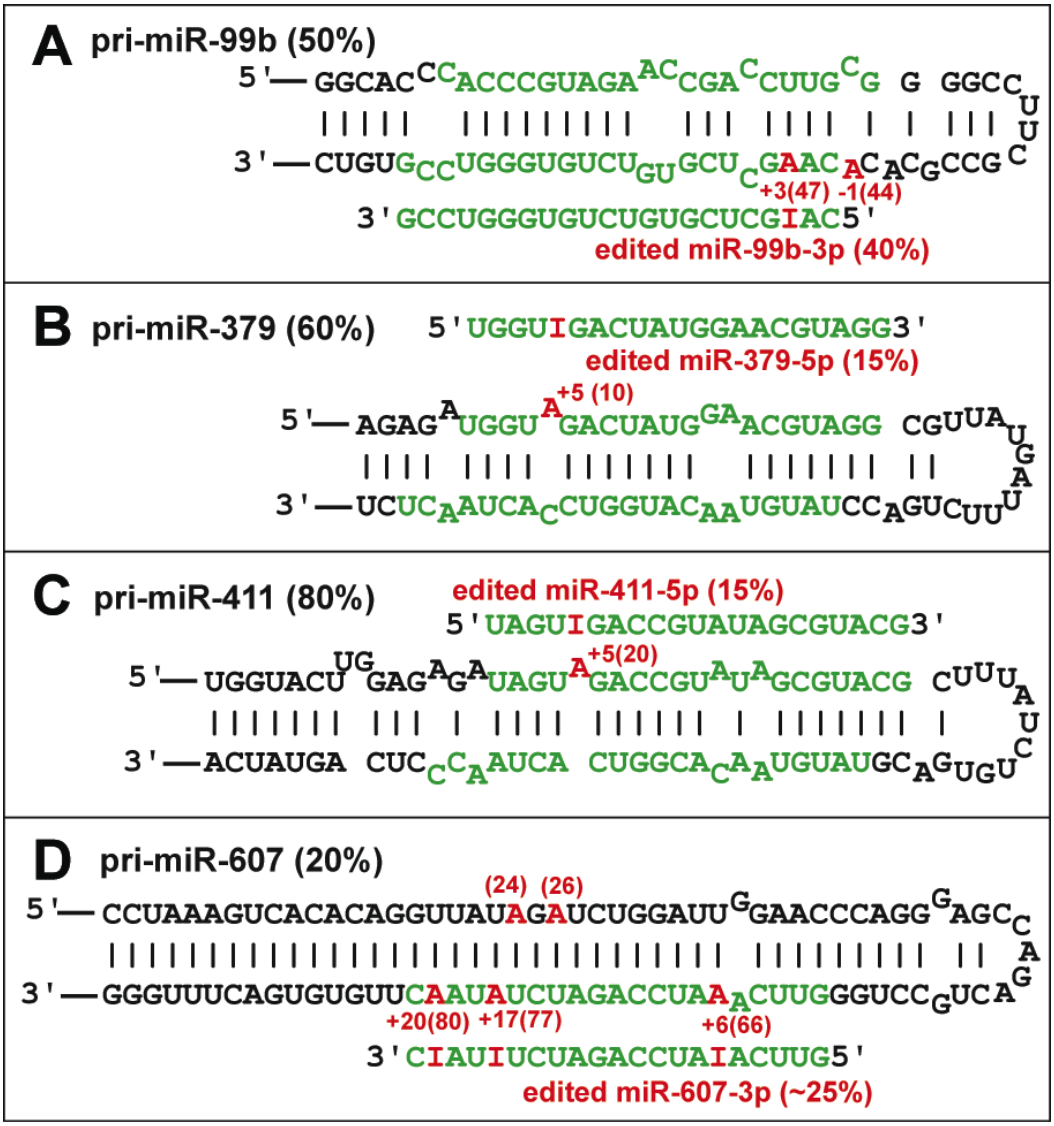

Figure 1. Identification of edited mature miRNAs. Four pri-miRNA structures and edited mature miRNAs are presented (A, B, C, D). Editing efficiency of pri-miRNA and mature miRNAs is indicated in parentheses in black and red, respectively. The 5' end of the mature miRNA is counted as +1, and the editing site is indicated by the number highlighted in red. The position of the pri-miRNA editing site is also indicated in parentheses. The 5' end of the pri-miRNA structure registered at the Sanger Center miRBase site is counted as 1.

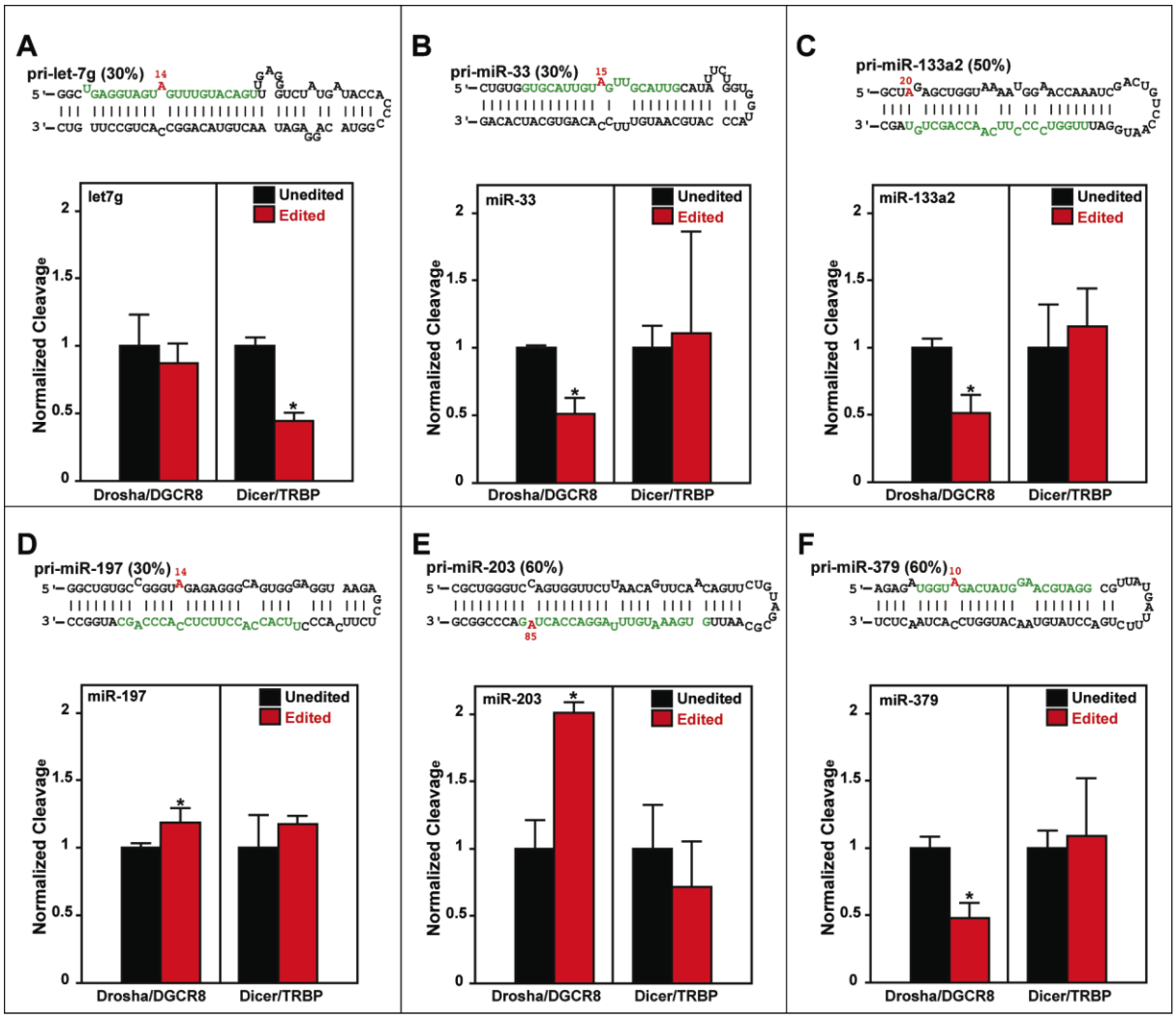

Figure 2. *In vitro* processing of edited pri-miRNAs. *In vitro* cleavage of pri-miRNAs by Drosha/DGCR8 complexes and pre-miRNAs by Dicer/TRBP were determined separately (A, B, C, D, E, F). Editing efficiency of pri-miRNA is indicated in parentheses. The editing site is indicated by the number highlighted in red. The 5' end of the pri-miRNA structure registered at the Sanger Center miRBase site is counted as 1. Three independent assays were done (n=3). Mann-Whitney *U*-test, *p<0.05.

Research Project 4: Project Title and Purpose

Diagnosis of Lung Cancer from Peripheral Blood Using Genomics - Lung cancer is the most common worldwide cause of cancer mortality. It is the second most prevalent cancer in both men and women and will account for more than 162,000 cancer deaths in the United States per year – more than from breast, prostate and colorectal cancers combined. Treatment for lung cancer remains very limited and the prognosis is strongly correlated to the stage of disease at presentation. Unfortunately, at the time of diagnosis only 15% of patients with lung cancer have

localized disease. Given these facts, one of the most important long-term goals in improving lung cancer survival is to detect malignant tumors in high-risk patients (i.e., smokers and ex-smokers) who represent the majority of all lung cancer cases, at an early stage, while they are still surgically resectable.

Duration of Project

1/1/2009 - 6/30/2010

Project Overview

Non-small cell lung cancer (NSCLC) is the most common cause of premature death from malignant disease in western countries. Early detection followed by surgery remains the most effective treatment. The diagnosis of patients who have lung nodules of unknown etiology identified on screening chest X-rays or computerized axial tomography (CT) scans is a common and important clinical problem for internists, surgeons, and pulmonologists. A blood test that could more accurately determine the risk of malignant disease in patients with lung nodules would be extremely valuable and have very important economic implications by reducing unnecessary surgery, biopsies, Positron Emission Tomography (PET) scans, and/or repeated CT scans. *We hypothesize that a specific and diagnostic gene expression profile for NSCLC can be identified in the peripheral blood mononuclear cells (PBMC) of lung cancer patients*. Preliminary studies have shown clear differences in gene expression profiles among patients with early stage adenocarcinomas of the lung (AC) and the appropriate high risk control group of smokers and ex-smokers matched by age, gender, and race. The goal of this project is to expand these preliminary studies to validate our preliminary results using a larger study population in a statistically rigorous manner.

Specific Aim I. Identify gene expression profiles on *microarrays* that accurately distinguish early stage NSCLC patients from the at-risk controls; assess the effects of gender, race and smoking history on classification; and evaluate changes in the cancer signature as a function of surgical removal of the cancer by comparing pre- and post-surgery PBMC profiles.

Specific Aim II. Validate the study using samples from additional locations and reassess biomarker panel generated on new samples collected over the grant period.

Principal Investigator

Louise C. Showe, PhD
Professor
The Wistar Institute
3601 Spruce St.
Philadelphia, PA 19104

Other Participating Researchers

Andrew Kossenkov, PhD, Michael K. Showe, PhD - employed by The Wistar Institute

Expected Research Outcomes and Benefits

Presently, early detection of tumors and their subsequent removal is the only effective treatment for lung cancer. However, unlike breast, colon and prostate cancer, no screening test exists for the detection of early stage lung cancer. For this reason the number of deaths due to lung cancer has not declined in decades. If one could develop a minimally invasive, inexpensive test that was easily accessible to individuals at risk for development of lung cancer, this could provide the same benefits of early detection that mammograms, colonoscopy and the PSA test have provided for breast, colon and prostate cancer. There are very large numbers of smokers and ex-smokers who are at increased risk for developing lung cancer. A simple blood test that could be included as part of a routine office visit or regular check-up that could detect the presence of an early stage lung cancer, with reasonable specificity and sensitivity, would provide a way to monitor these at-risk individuals for the development of lung cancer. This test would have the added advantage of not having to expose individuals to radiation associated with routine chest x-rays or CT scans. If a significant fraction of lung cancers can be detected early enough for surgical intervention, this would have a major effect on both mortality and health care costs.

Summary of Research Completed

<u>Aim 1. Using the Illumina arrays, we will optimize our diagnostic signatures for early stage lung cancer.</u>

1a. Assess the accuracy of a global diagnostic signature to identify patients with Non-Small Cell Lung Cancers (NSCLCs) of various cell types, and compare the accuracy of a more global diagnostic to cell type specific diagnostic signatures.

A publication documenting the accuracy of our peripheral blood mononuclear cells (PBMC) derived signature for early stage lung cancer appeared in Cancer Research in December 2009, Showe et al. In summary, we identified a 29 gene expression signature using Illumina gene expression arrays that distinguished NSCLC patients from an at-risk control population of smokers and ex-smokers with non-malignant lung disease with 91% sensitivity and 80% specificity. We also identified a number of significant pathways that distinguished patients from controls that provide some information as to the effects of the NSCLC on immune pathways. These are shown in *Figure 1*.

During the past six months, we have focused on obtaining sufficient numbers of samples from different locations to allow us to have the power to assess and validate gene expression signatures for cancer subtypes as well as smoking status. We have now obtained almost 400 additional samples for analysis, and these are being processed for array studies. In particular, we have focused on recruitment of additional samples from under-represented classes of patients, including lung squamous cell cancer (LSCC), and from never-smokers. In order to more easily standardize sample collection at different sites, we tested and validated a new blood collection technique that would facilitate moving this test from the research bench to the clinic. We convinced collaborators at the University of Pennsylvania, New York University Medical College, and Brigham and Women's Hospital to collect samples into PAXgene tubes from

PreAnalytics. This allows for immediate stabilization of the ribonucleic acid (RNA) and no further purification of PBMC is required. *Figure 2* is a comparison of samples collected and rapidly processed to purify PBMC and samples collected in PAXgene where the RNA was purified and processed on arrays four months later. The paired PAXgene and PBMC-derived samples were hybridized at the same time to Illumina gene expression arrays, and the gene expression in paired samples was compared. We tested five samples from the same patient, collected weekly with both CPT and PAXgene tubes. Out of 16,319 genes detected by either collection method, there was a set of 1,145 genes detected by either PAXgene or CPT with fold change >2. Correlation analysis between PAXgene and CPT of the five corresponding pairs of samples without those differentially detected genes showed an average r=0.969, while pair-wise comparison of just the CPT samples showed an average r=0.988. *Figure 2* shows a representative plot of CPT vs PAXgene expression values for one of these time points and indicates high concordance between two collection methods. A recent paper further supports this similarity. This simplified collection system has made it significantly easier to recruit new clinicians to samples collection.

1b. Determine if accuracy can be improved by developing specific signatures that take into account cancer subtype and stage and, in particular, develop high accuracy in distinguishing benign from malignant nodules.

The current noninvasive diagnostic tool for lung cancer, computerized axial tomography (CT), efficiently detects small lung nodules (>1 mm). However, it has disadvantages that create substantial need for alternative and universally accessible diagnostic modalities. First, very few small nodules detected by CT prove to be malignant upon biopsy. Sub-centimeter nodules, ~2,000,000 of which are detected annually in the U.S., require additional sequential CT examinations to determine whether they grow. Furthermore, although CT scans are readily available in urban areas, they remain unaffordable in many regions and in countries with very large at-risk populations. Therefore, our proposed alternative approach of a blood-based diagnostic test, that could be made more readily available in typical clinical settings, would provide alternative, complementary information to that obtained from CT scans. Combining the proposed genomic data with clinical imaging data may facilitate development of more reliable algorithms for cancer diagnosis and will likely reduce unnecessary surgery or biopsies and their associated expense and risks. Our previous studies suggested that more accurate diagnosis of nodules as benign or cancer will require the development of a diagnostic signature specific for these cases. In addressing the problems with nodules, we have focused recruitment on this important class of "at-risk" individuals as described below. We expect that this focused study of malignant vs benign lung tumors will allow us to develop a robust signature that distinguishes the 2 classes of tumors. Our preliminary studies suggest this is possible.

1c. Evaluate changes in diagnostic signatures present in pre-surgery and post-therapy samples to identify signatures that may be used to assess the probability of recurrence in a long term goal.

Effect of Tumor Removal on Individual Classification Scores

Additional pre-post studies (Kossenkov et al 2010 paper in preparation): During the past funding period, we further analyzed the differences in gene expression in the peripheral immune system

before and after removal of NSCLC. These studies have begun to provide information that suggests the importance of pursuing these studies in a much larger sample size. To further assess the differences between pre- and post-surgery samples, we directly compared these 2 classes of samples. Because of the unique relationship of our cancer and control cases, expression of only four genes are required for 100% accurate discrimination of the two classes. A comparison of the post-surgery samples and patients with non malignant lung disease, with and without the presence of benign nodules, indicates that they form a separate class in a hierarchy of relationships. Even though all the four classes of patients share many characteristics of lung disease, their relative relationships to NSCLCs based on gene expression are: NSCLC>benign nodules>post surgery samples>samples with other lung diseases. Significant alterations are found post-surgery in pathways associated with the innate immune response.

We also analyzed miRNA transcription profiles that differed in pre- vs. post-surgery samples. We found expression of five to be universally *upregulated* in pre- compared to post-surgery samples, and then computationally identified targets of those miRNAs within the differentially expressed genes *downregulated* in pre- compared to post-surgery samples. miRNA targets were significantly enriched for transcription factors, including several key factors in immune regulation.

We have now established collaboration with a lung cancer surgeon at Brigham and Women's Hospital in Boston. He is carefully collecting samples in PAXgene tubes pre- and post-surgery for both NSCLCs and benign nodules. We have applied for NCI funding to further pursue these studies. In addition to gene expression, we will also carry out microRNA expression studies with the intent to develop a gene expression signature which includes markers for both coding and non-coding RNAs, which may provide a more robust classifier.

Outcome assessment using the gene expression data. Given the passage of time of more than five years for the majority of the patients that we assessed, we have now collected outcome data on 108 of the 137. We are in the process of completing this analysis, and a manuscript based on this study is in preparation. (Vachani et al 2010)

Aim 2. Validate the generalizability of our signatures on subsets of case and control samples from external validation sites not used to develop the diagnostic signature.

Validation of the 29-Gene Classifier on Independent Samples

Validation studies are continuing with samples being collected at various locations. Because of the large differences in the quality of the PBMC samples acquired from several different locations, we are continuing our studies by analyzing samples in whole blood collected in the more easily standardized PAXgene system. We have demonstrated the similarities in gene expression in PAXgene and immediately processed RNA in *Figure 2*. PAXgene tubes require the collection of only 2.5 mls of blood, which is immediately stabilized by a reagent in the collection tube. As it does not require sample processing in a timely manner at the site of sample collection, this approach facilitates highly controlled sample collection at sites that lack equipment or expertise to process samples at the time of collection. This approach has greatly reduced the burden on the collection and has increased our ability to collect many more samples. We have now received ~400 PAXgene samples from New York University Medical College and

50 samples from the University of Pennsylvania, and we have also set up collaborations with investigators at Brigham and Women's Hospital who are collecting pre- and post-surgery samples from patients and controls in PAXgene. A grant application to study these samples has been submitted. We are presently processing the samples from New York University Medical College and the University of Pennsylvania.

We continue to collect samples for both cases and controls from the University of Pennsylvania and New York University Medical College. We identified a source of pre/post surgery samples for NSCLCs and benign nodules at Brigham and Women's Hospital. To date, we have collected approximately 400 samples from New York University Medical College, and approximately 300 have been processed on arrays for analysis.

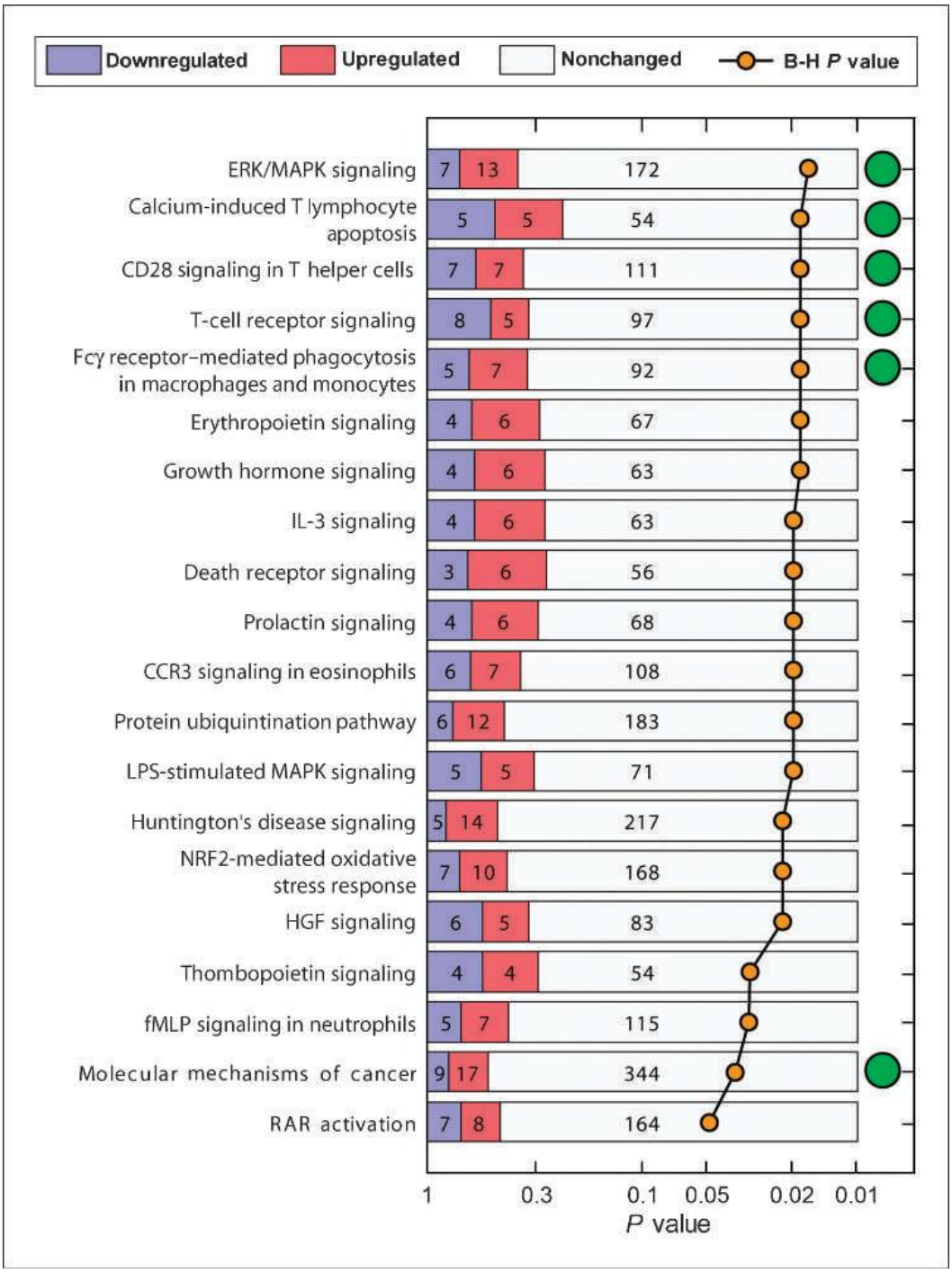

Figure 1. Significantly enriched canonical pathways from Ingenuity Pathway Analysis of the genes differentially regulated between NSCLC and NHC samples. Numbers in the bars, the number of genes in the pathway significantly higher in cancer(red) or lower in cancer (blue). B-H,Benjamini-Hochberg multiple testing correction. Green circles, pathways that were also enriched in NSCLC versus benign nodule comparison.

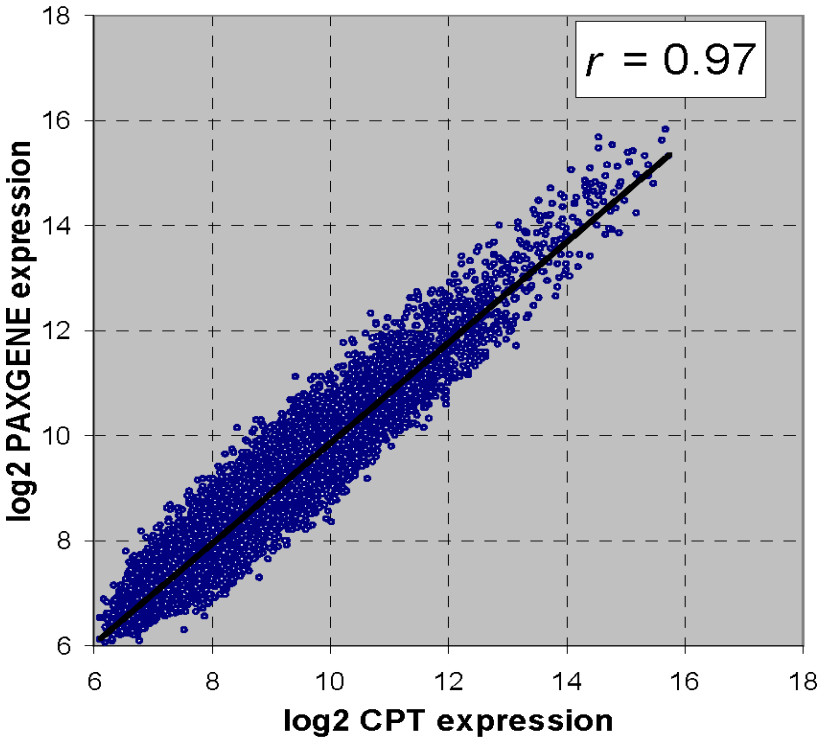

Figure 2: Correlation of expression of 15174 genes from matched CPT and PAXgene samples