Children's Hospital of Philadelphia

Annual Progress Report: 2012 Formula Grant

Reporting Period

July 1, 2013 – June 30, 2014

Formula Grant Overview

The Children's Hospital of Philadelphia received $3,713,220 in formula funds for the grant award period January 1, 2013 through December 31, 2016. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

Functional Follow up of Genetic Commonalities to Diabetes and Cancer – Diabetes affects 18 million adults in the United States, with approximately 90 to 95 percent of those affected having type 2 diabetes (T2D). It is clear from recent genomic study observations that there is a specific yin and yang between cancer and T2D at the genetic level, including for the most strongly associated T2D gene, TCF7L2, which was discovered by the principal investigator. This has motivated us to systematically investigate the relationship between loci uncovered by genome wide association studies (GWAS) for cancer, leading us to ultimately carry out islet proliferation studies in mice, a mechanism which still largely eludes the diabetes research community but could revolutionize the way diabetes is treated if successful.

Anticipated Duration of Project

1/1/2013 – 12/31/2016

Project Overview

The repertoire of genes already established to play a role in the pathogenesis of type 2 diabetes (T2D) has grown substantially due to recent genome wide association studies (GWAS). In 2006, the P.I. on this application discovered the strong association of variants in the transcription factor 7 like 2 (*TCF7L2*) gene with T2D. Other investigators have already independently replicated this finding in different ethnicities and, interestingly, from the first GWAS of T2D in Caucasians, the strongest association was indeed with *TCF7L2*; this is now considered the most significant genetic findings in T2D to date.

Interestingly, there is also a very strong connection between *TCF7L2* and cancer. The key 8q24 locus found to be the most strongly associated genomic region with a number of cancers through GWAS contributes to the disease pathogenesis through mutation of an upstream TCF7L2-binding element driving the transcription of the *MYC* gene. Indeed, it has been known for some years that *TCF7L2* harbors specific mutations that strongly influence colorectal cancer risk plus

genomic sequencing of colorectal adenocarcinomas identified a recurrent *VTI1A-TCF7L2* gene fusion. Furthermore, many of the T2D GWAS-derived risk conferring alleles have been shown to protect against prostate cancer; in addition, *THADA*, *JAZF1* and *TCF2* are loci that have been strongly detected in separate GWAS analyses of prostate cancer and T2D. Thus, *TCF7L2* and other T2D associated genes also appear to be key players in cancer pathogenesis; however, this mechanism is still far from understood.

We previously performed ChIP-seq with this transcription factor to elucidate its binding repertoire genome wide. Unexpectedly, and despite employing a carcinoma cell line, the genes with TCF7L2 binding sites are strongly enriched in pathway categories related to metabolic-related functions and traits, further suggesting a role for metabolism in cancer.

We are taking forward the loci that are common to T2D and cancer GWAS outcomes to investigate their impact on cell proliferation with the ultimate goal of testing their role in beta-cell proliferation in mice, a mechanism which still largely eludes the diabetes research community.

Aim 1: Oligo-pull down / mass spec – characterize the set of proteins binding across rs7903146 and any allele differences in affinity.
Aim 2: Over-expression through constructs and under-expression through siRNA of *TCF7L2*, *THADA*, *JAZF1* and *TCF2* (also known as *HNF1B*) plus two other genes (based on expected finds from subsequent literature) in selected cell lines (colorectal and prostate, including HCT116) plus murine derived colon cells in order to assess influence of proliferation. Further we aim to also carry this out in an L-cell gut cell line, the human EndoC-βH1 cell line and a rodent beta-cell line to further explore influence on proliferation.
Aim 3: Somatic Gene Targeting (SGT) of *TCF7L2* in selected cell lines, both at the single rs7903146 SNP level and at the whole gene level - carried out internally with vendor generated constructs. Subsequent RNA-seq, proteomic studies and proliferation of SGT generated cell lines to fully investigate *TCF7L2* allele and gene effects in the cell lines altered. Also assess the influence of using medium from alpha, gamma and other cells in this context.
Aim 4: siRNA experiments with key genes resulting from pathway analyses to investigate how their perturbation influences gene expression in cancer, adipocyte and pancreatic derived cell lines.
Aim 5: Generation of mouse for best gene coming out of proliferation studies in year 1-3 to assess islet proliferation in this model system.

Principal Investigator

Struan F.A. Grant, PhD
Associate Professor, Department of Pediatrics
Children's Hospital of Philadelphia
Rm 1216F, 3615 Civic Center Boulevard
Philadelphia PA 19104-4318

Other Participating Researchers

Andrei Thomas-Tikhonenko, PhD; Adam Resnick, PhD; Qianghua Xia, PhD; Vanessa Guy, MS- employed by Children's Hospital of Philadelphia
Klaus Kaestner, PhD; Morris Birnbaum, MD, PhD; Tia Banks-Bernard; Jia Zhang, PhD – employed by University of Pennsylvania

Expected Research Outcomes and Benefits

The T2D *TCF7L2* association has an advantage, as the causal variant is widely thought to be identified, i.e., the work with multiple ethnicities has distilled down the association to a single variant, rs7903146, in intron 3. We will be applying the cutting edge approaches of somatic gene targeting in pre-selected cell lines and oligo pull-down combined with mass spectrophotometry in order to elucidate which proteins are binding across the SNP (plus look for allelic-specific binding differences) and determine what happens when you perturb the gene and even the actual base specifically in a cell model setting – this will in turn shed light on the mechanism by which TCF7L2 exerts its effect on T2D risk, giving us crucial insights in to the genetic architecture of the disease.

siRNA experiments with key genes resulting from pathway analyses related to Akt will be used to investigate how their perturbation influences gene expression in cancer, adipocyte and pancreatic derived cell lines to understand more precisely the classical pathways for cancer and diabetes.

Over-expression of *TCF7L2*, *THADA*, *JAZF1* and *TCF2* in key cell lines will help us determine which influence proliferative potential which in turn will inform us as to which gene(s) to take forward in to a mouse model to check for beta-cell proliferation. If successful this would have a fundamental impact on the way type 2 diabetes is both treated and prevented in the future. Furthermore, it will inform us on how metabolism contributes to the etiology of cancer.

Summary of Research Completed

MILESTONE(S) FOR 7/1/2013-6/30/2014:

Aim 1. Oligo-pull down / mass spec will have characterized allele differences in affinity of characterized proteins binding across rs7903146 in *TCF7L2*
We investigated if there was allele specific preferential binding for some of the proteins detected across rs7903146 in order to gain further insight in to the regulation of the transcription factor 7-like 2 (*TCF7L2*) gene To test our hypothesis, we performed "two-way" oligo pull-down experiments. The nuclear extracts from the 'light' cells that were isotopically labeled in 'stable isotope labeling by amino acids' (SILAC) medium were used in the C allele oligo pull-down while the nuclear extracts from the 'heavy' cells that were isotopically labeled in SILAC medium were used for the pull-down of the T allele oligo. The SILAC labeled extracts were then switched for a reverse experiment in order to assess the reliability and reproducibility of the approach.

Cell culture, nuclear extracts preparation and subsequent oligo-pull down were described in last year's report. In order to characterize allele differences in affinity the 5' Dual Biotin modified oligonucleotides for each allele (C/T) for rs7903146 were synthesized by Integrated DNA Technologies, Inc. The sample was digested with trypsin and analyzed with nanoLC/mass spectrometry (MS) at the University of Pennsylvania Proteomics Core. The data were analyzed with Sequest. Scaffold was used to validate MS/MS based peptide and protein identifications. For the SILAC experiments, Zoom Scan was added to the MS analytical method to monitor the labeled and unlabeled peptides. The quantification of the SILAC peaks was performed with ProteoIQ with customized modification so that the software could use the zoom scanned SILAC peaks for quantification.

We identified two less abundant proteins, X-ray repair cross-complementing protein 5 (XRCC5; also known as Ku80) and replication protein A 70kDa DNA-binding subunit (RP-A p70), that were preferentially binding to the T allele over the C allele (Figure 1). This work was recently published online - *Q. Xia, S. Deliard, C.X. Yuan, M.E. Johnson and S.F.A. Grant: Characterization of the transcriptional machinery bound across the widely presumed type 2 diabetes causal variant, rs7903146, within TCF7L2. European Journal of Human Genetics [Epub ahead of print] March 2014.*

Aim 2. Assessment of influence on proliferation through over-expression through constructs and under-expression through siRNA of *TCF7L2*, *THADA*, *JAZF1* and *TCF2* (also known as HNF1B) plus two other genes (based on expected finds from subsequent literature) in six human cancer cell lines (colorectal and prostate, including HCT116) plus murine derived colon cells. *TCF7L2, THADA*, *HNF1B* and *JAZF1* are all genes that have been implicated in type 2 diabetes (T2D) genome wide association studies (GWAS). Additionally, *TCF7L2* has been implicated in colorectal cancer and *THADA*, *HNF1B* and *JAZF1* have all been implicated in prostate cancer. Before manipulating these genes, we chose to assess levels of their expression in two colorectal cancer cell lines (HCT116 and CACO2) and two prostate cancer cell lines (LNCaP and PC3). *TCF7L2* is highly expressed in both HCT116 and CACO2 cells. However, expression of *THADA*, *HNF1B* and *JAZF1* is nearly undetectable in both of these colorectal cancer cell lines. PC3 cells showed adequate expression of *TCF7L2* and *THADA* but showed little expression of *HNF1B* and JAZF1. LNCaP cells showed nearly undetectable expression of *TCF7L2* but showed high expression of the T2D/prostate cancer implicated target genes.

Since *TCF7L2* is highly expressed in both colorectal cancer cell lines and has been previously implicated in colorectal cancer, we decided to deplete its expression using distinct small interfering (si) RNA sequences in order to assess its contribution to cell viability. Cells were transfected with siRNAs at a final concentration of 50 nmol/L using Lipofectamine RNAiMAX (Invitrogen). Nontargeting control, TCF7L2 smartpool, TCF7L2 #6, THADA smartpool, and THADA #10 siRNAs were purchased from Dharmacon.

Both the *TCF7L2* smartpool siRNA and the *TCF7L2* individual siRNA #6 efficiently depleted expression of *TCF7L2*. Additionally, both siRNAs resulted in reduced cell viability as measured by a WST-1 assay.

In order to assess the influence of *THADA*, *HNF1B* and *JAZF1* on cell viability, we decided to

manipulate the expression of these genes in prostate cancer cell lines (since all three genes have been implicated in prostate cancer rather than colorectal cancer). Although we were not able to reduce expression of any of these genes in LNCaP cells using a variety of transfection reagents, we were able to substantially reduce expression of *THADA* well below baseline levels in PC3 cells (Figure 2A). However, this reduction in *THADA* expression did not result in any significant change in cell viability (Figure 2B). Since expression of *HNF1B* and *JAZF1* were nearly undetectable in PC3 cells, we decided to assess the impact of these genes using transient ectopic expression. *JAZF1* and *HNF1B* were cloned into pcDNA3.1D/V5-His-TOPO expression vectors (Invitrogen). Transient transfection of expression vectors was performed using Lipofectamine 2000 (Invitrogen).

As shown in Figure 2C, we were able to successfully induce ectopic expression of *HNF1B, JAZF1* and *LACZ* (as a control). We found that low levels of expression of *HNF1B* and *JAZF1* did not significantly alter cell viability; however, higher levels of expression of both *HNF1B* and *JAZF1* significantly reduced cell viability of PC3 cells (Figure 2D).

Aim 3. Completed Somatic Gene Targeting (SGT) of *TCF7L2* in HCT116 cells, both at the single rs7903146 SNP level and at the whole gene level

The single nucleotide polymorphism (SNP) rs7903146 within the third intron of *TCF7L2* is the strongest association with T2D, widely considered to be the causal variant. However, the underlying mechanism by which *TCF7L2* is impacted by this genomic region is unknown. Recent technological advancements in gene targeting, including transcription activator-like effector nucleases (TALEN) or clustered regularly interspaced short palindromic repeats-cas9 (CRISPR-cas9) mediated genome modification, have made it feasible to edit the genomic element precisely in somatic cells. To study the role of this variant within *TCF7L2*, we aimed to edit the genomic region harboring the SNP rs7903146 in HCT116 cells, where *TCF7L2* is abundantly expressed. Targeted sites were predicted and the corresponding sequences were cloned into the delivery constructs. The targeting constructs were transfected into HCT116 cells. To optimize targeting efficiency, several constructs targeting different sites were tested. The targeted single clones were isolated for genotyping. Sanger sequencing results revealed that no homozygous recombination clones were obtained, however, 100bp or 1.5kb surrounding the SNP rs7903146 were successfully deleted.

To study the effects of this genomic alteration on TCF7L2 expression, we isolated protein and messenger ribonucleic acid (mRNA) from the targeted cells. We compared TCF7L2 protein levels between the untargeted and targeted cells using quantitative real time polymerase chain reaction (PCR). rs7903146 is located between exons 4 and 5. The TCF7L2 isoform containing the exon 4 is less abundant. We chose two pairs of primers for PCR amplification. One amplified exons 4 thru 5 and the other amplified exon 5 alone. As shown in Figure 3A, *TCF7L2* mRNA levels were significantly reduced by deleting the genomic region corresponding to the SNP in HCT116 cells. Importantly, Western blot results confirmed that TCF7L2 protein levels were also significantly reduced in the targeted cells (Figure 3B), perfectly consistent with mRNA levels. It should be noted that the reduction of protein appeared to be general and not exon specific, suggesting that the region corresponding to the SNP may act as an enhancer element to regulate *TCF7L2* expression.

Aim 4. Completed siRNA experiments with key genes, giving us insights into candidate intervention points in the key Akt pathway that has been classically central to the biology of both T2D and cancer.

To begin, we conducted pathway analysis on T2D genes implicated by GWAS. This pathway analysis revealed a great deal of overlap between T2D genes and cancer genes. Specifically, the most consistently implicated T2D genes that were also implicated in pathways in cancer were cyclin-dependent kinase inhibitor 2A (*CDKN2A*), cyclin-dependent kinase inhibitor 2B (*CDKN2B*) and peroxisome proliferator-activated receptor gamma (*PPARG*). Before manipulating these genes, we chose to assess levels of their expression in metabolically relevant tissues including colon (HCT116 - carcinoma; NCM460 - colonocytes), pancreatic beta-cells (EndoC-βH1) and pre-adipocytes (SGBS). Expression of *CDKN2B* was undetectable in any of these cell lines, and the expression of *PPARG* did not exhibit any interesting tissue variability. However, *CDKN2A* showed dramatically higher expression in the pancreatic β-cells compared to any other cell type assayed.

Since CDKN2A encodes the cell cycle inhibitor $p16^{INK4A}$, we decided to investigate the effects of this gene's expression in controlling pancreatic β-cell viability. As shown in Figure 4A, both the *CDKN2A* smartpool siRNA and the *CDKN2A* individual siRNA #8 efficiently depleted expression of $p16^{INK4A}$. However, this reduction in $p16^{INK4A}$ expression did not result in any significant change in cell viability (Figure 4B).

Note: Aim 5 not due to start until 7/1/2015 (Rodent-related work)

Figure 1: Relative DNA binding affinity of XRCC5 and RP-A p70 between C and T allele of rs7903146. Forward labeling experiment shown as an example.

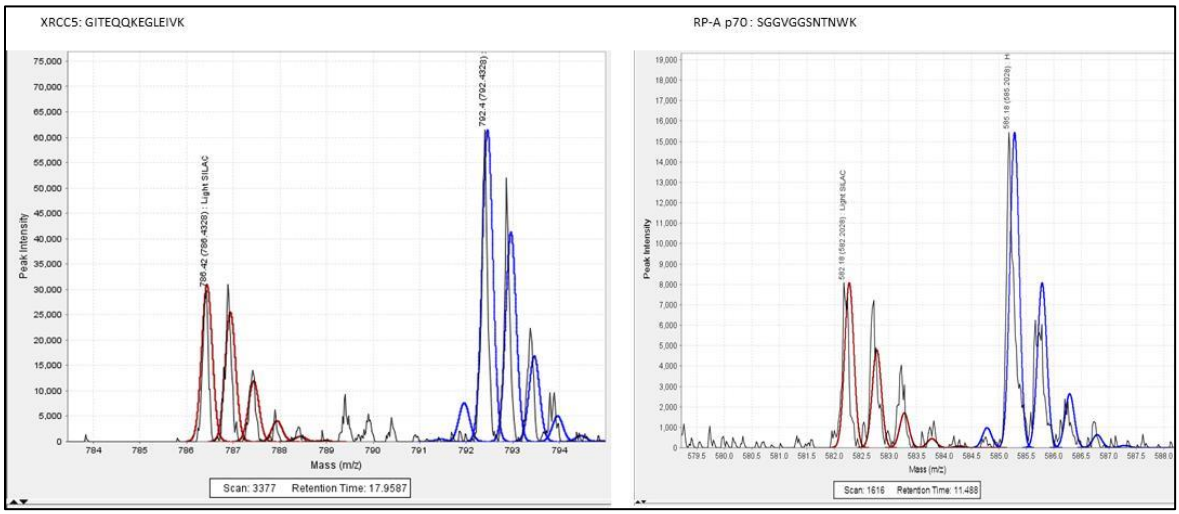


Figure 2: Manipulation of T2D/prostate cancer. **A.** PC3 cells were transfected with non-targeting control siRNA (Cnt), THADA smartpool siRNA or THADA siRNA #10 sequences. After 72 hours, cells were lysed, and lysates were analyzed by western blotting as indicated. **B.** Same as A, except that cells were analyzed by WST-1 after 2 hours of exposure to the reagent. **C.** PC3 cells were transfected with the indicated amounts of LACZ-V5, HNF1B-V5, or JAZF1 expression constructs. After 24 hours, cells were lysed, and lysates were analyzed by western blotting as indicated. **D.** Same as C, except that cells were analyzed by WST-1 48 hours post-transfection after 2 hours of exposure to the reagent.

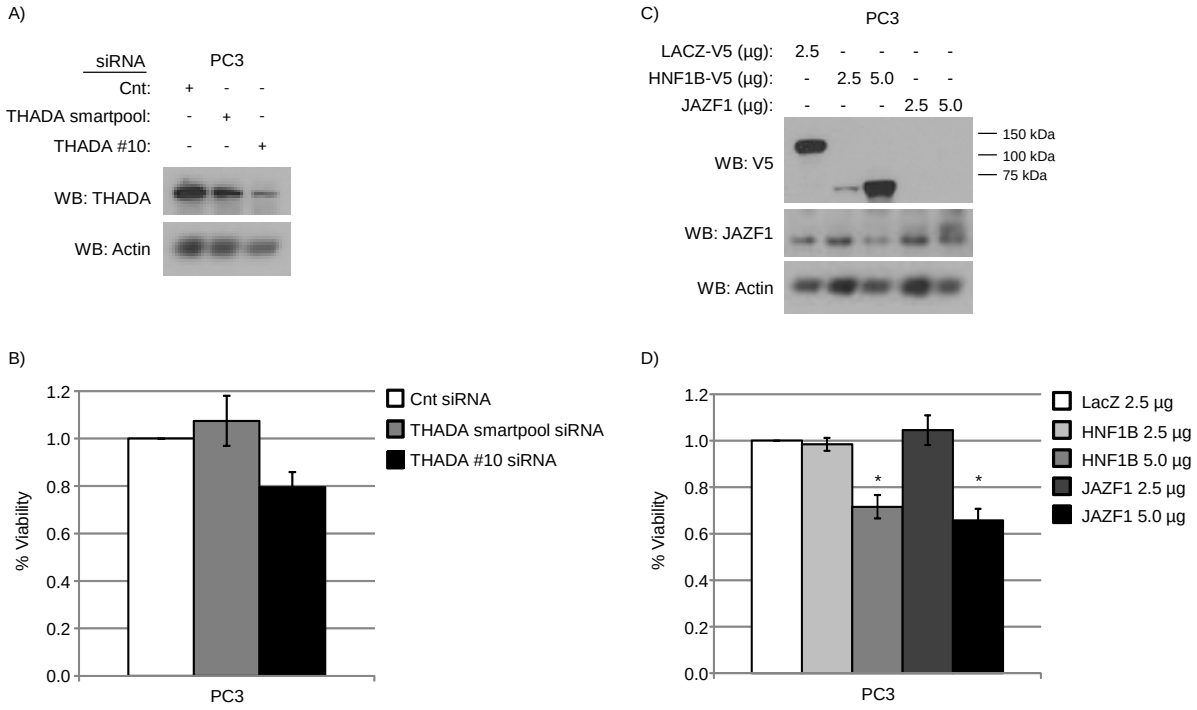


Figure 3: A. *TCF7L2* mRNA levels were significantly reduced by the targeting deletion of then genomic element crossing the SNP rs7903146. Fold changes were normalized by GAPDH level. B. TCF7L2 protein levels were significantly reduced by the targeting deletion of the genomic element spanning the genomic region corresponding to SNP rs7903146. β-actin was used as the loading control.

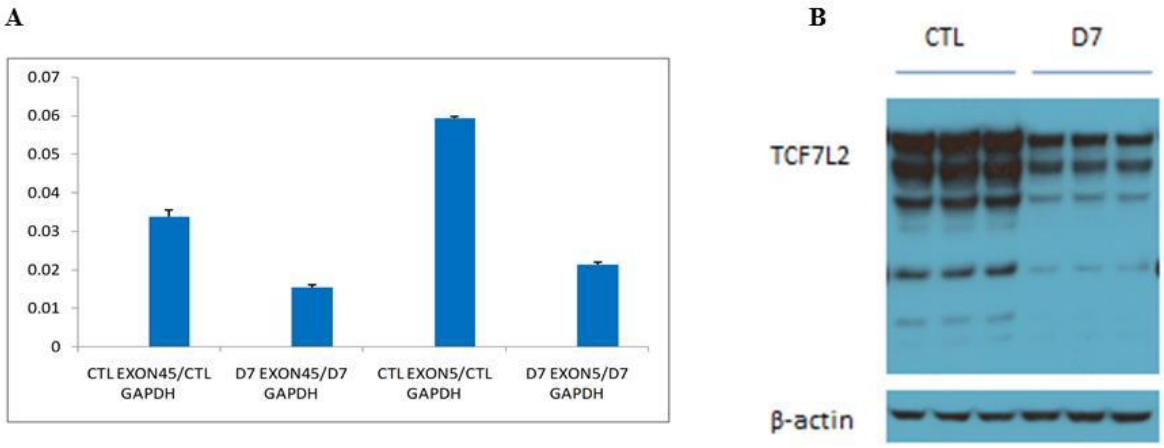
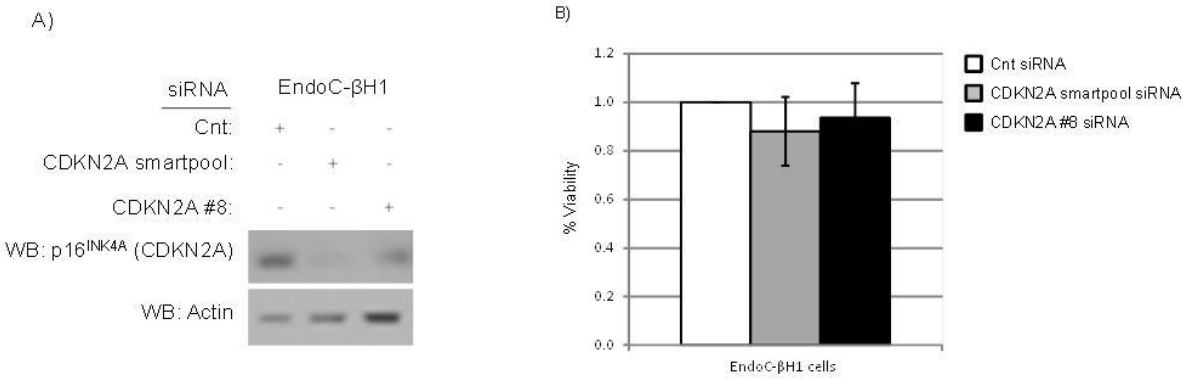


Figure 4: Manipulation of CDKN2A ($p16^{INK4A}$) in pancreatic β-cells. A. EndoC-βH1cells were transfected with non-targeting control siRNA (Cnt), CDKN2A smartpool siRNA or CDKN2A siRNA #8 sequences. After 96 hours, cells were lysed, and lysates were analyzed by western blotting as indicated. B. Same as A, except that cells were analyzed by WST-1 after 4 hours of exposure to the reagent.



Research Project 2: Project Title and Purpose

Comparative Effectiveness of Developmental-Behavioral Screening Instruments in Children – Evidence of the advantages of using the Survey of Wellbeing of Young Children (SWYC) to screen for childhood behavioral-developmental disorders is accumulating. Findings are limited to English speaking, well-educated, predominantly U.S.-born families. Whether the findings are generalizable to other children from two groups rapidly growing in size in Pennsylvania-- Hispanic families lacking English language proficiency and African-born Black women-- is unclear. Consequently we propose to recruit a sample of Hispanic children and a sample of children of African-born Black women and conduct a systematic comparison study of the SWYC to other instruments widely used to identify behavioral-developmental disorders in children.

Anticipated Duration of Project

1/1/2013 – 12/31/2016

Project Overview

The overarching objective of the proposed project is to build upon existing research which systematically compares the accuracy, feasibility, and effectiveness of pediatric behavioral and developmental screening instruments. In brief, we will extend and deepen this research by recruiting families from linguistically/culturally distinct populations (of particular importance to pediatric providers and public health officials in Pennsylvania) which heretofore have *not* been included in this research. The first specific aim, is to assess the accuracy (specificity; sensitivity; positive and negative predictive values) of the Ages and Stages Questionnaires (ASQ-SE and ASQ-3), M-CHAT and the SWYC set of screening instruments for the above mentioned sub-populations. The second specific aim is to integrate all screening data collected with relevant, supplemental self-reported survey data (behavioral, socioeconomic and psychosocial) in order to *develop and evaluate a predictive model related to the detection of developmental and behavioral disorders in the study population(s*). The essential components of the research design are as follows: 1) recruit parent/child pairs across 9 child age categories (9 to 60 months) at pediatric/prenatal providers and community centers throughout the city of Philadelphia; 2) obtain informed consent, administer the Ages and Stages, M-CHAT, and the SWYC screening instrument components and collect sociodemographic, psychosocial and behavioral data; 3) conduct complete clinical behavioral and developmental evaluations on at least 356 of the enrollees providing the 'gold standard' for assessing the accuracy of the screening instruments described above; and 4) use advanced statistical techniques on all data (screening, clinical evaluation and survey), in order to identify the best predictive model of pediatric behavioral and developmental disorders.

Principal Investigator

Jennifer F Culhane, PhD, MPH
Senior Investigator
The Children's Hospital of Philadelphia
34th and Civic Center Blvd
CHOP North-Suite 880
Philadelphia, PA 19104

Other Participating Researchers

Sandra R Wadlinger, MS, RRT; Marsha Gerdes, PhD; Leny Mathew, MS; David Webb, PhD; Ximena Cuellar, BS, Karen B. Friedman, PhD; Sherry Winston, BA; Delitza Hernandez, BS – employed by The Children's Hospital of Philadelphia
Irma Elo, PhD; Emilio Parrado, PhD – employed by The University of Pennsylvania

Expected Research Outcomes and Benefits

The number of professional organizations and state governments that are recommending or mandating systematic screening for pediatric developmental and behavioral problems is growing. The use of standardized screening instruments to detect these disorders is thus rapidly increasing in pediatric care settings. The proportion of pediatricians who reported using a standardized instrument to screen for developmental problems rose from 23% to 47.7% over a five year period between 2002 and 2009. Unfortunately, because of limited research pediatricians make decisions about screening from a fairly weak base of evidence to select an appropriate tool that is inexpensive, brief, and easy to administer and score. However, financial considerations and ease of use should not trump considerations of the accuracy and validity of screening tools, since these choices will have significant implications for children and their families. Additionally, screening tool selection could have profound implications for public health costs and policies as well as individual child life-course outcomes. If only a small fraction of pediatricians were to use a screen that has slightly better accuracy, the number of children correctly identified with a developmental-behavioral disability would increase considerably, facilitating early intervention and this improved prognosis, while the number of incorrectly identified children (false positives) would be greatly diminished.

Summary of Research Completed

Project Milestones (for period 7/1/13-6/30/14):
1. Achieve total enrollment of 420 subjects
2. Hold Executive Steering Committee Data Quality Meeting

<u>Specific Aim 1</u>

Specific Aim 1 is to assess the accuracy (specificity; sensitivity; positive and negative predictive values) of the Ages and Stages Questionnaire-Social-Emotional (ASQ-SE) and Ages and Stages Questionnaire-Third Edition (ASQ-3), the Modified Checklist for Autism in Toddlers (M-

CHAT) and the Survey of Well-Being of Young Children (SWYC) screening instruments in a sample of Hispanic children and children of African-born Black women. Progress towards this aim was achieved during this project period by completion of all regulatory activities required to start and continue participant recruitment; hiring and successful training of key personnel; and actively recruiting, screening, and evaluating children of Spanish speaking and African-born Black parents. In addition, we formed an Executive Data Quality Committee that met regularly to review Study progress and establish data quality control procedures. A detailed description of the activities associated with each of these achievements is provided below.

Completion of all Regulatory Activities

In August 2013 an amendment including updates to the protocol, outreach materials, Study survey and two instruments necessary to conduct the second visit evaluations was submitted to the Institutional Review Board (IRB) at The Children's Hospital of Philadelphia (CHOP) and approved. The instruments were the Bayley Scale of Infant Development-III (BSID III) and the Differential Abilities Scale-II (DAS 2). In October 2013 a second amendment was submitted notifying the CHOP IRB of new Study personnel. A third amendment was sent for CHOP IRB review and approval in November 2013. This amendment included updates to the protocol and informed consent form, requested authorization to videotape second visit assessments and the addition of two instruments to be administered during the second visit evaluations. The two instruments submitted and subsequently approved were the Screening Tool for Autism (STAT) and The Childhood Autism Rating Scale (CARS 2). Finally, in December of 2013 a mandated Annual Continuing Review was submitted to the CHOP IRB and approved.

Hiring and Training of Key Personnel

Two bilingual Psychology Associates (one full time and one part time) were hired to conduct the second visit assessments for all children who, based on the initial screening results, were designated for full clinical evaluations. Their training was conducted and their ongoing work proceeded under the direct supervision of a licensed psychologist (Dr. Marsha Gerdes). Their initial training involved:

- review of all Study materials with emphasis on the Study protocol and informed consent documents
- review of assessments and their companion manuals (BSID-III, DAS-2, STAT, CARS-2, Infant Toddler Social Emotional Assessment (ITSEA), Social Communication Questionnaire (SCQ) and Preschool Age Psychiatric Assessment (PAPA)
- direct observation of assessment procedures conducted by experienced CHOP clinicians
- item by item review of each test parameter in both English and Spanish
- practice sessions directly supervised by Dr. Gerdes
- review of video-recordings of live clinical evaluations in order to ensure reliability and validity of the assessment, which continued until criteria of 95% or above agreement between associates and CHOP clinicians was obtained.

In addition, the Psychology Associates were trained on establishing best practices and scripted language to optimize subject second visits by:

- establishing rapport with infants, toddlers and preschoolers

- maximizing attention and participation of toddlers and preschoolers
- maximizing parent comfort and responses on questionnaires
- role modeling to ensure that feedback to families was appropriate and effective

A bi-lingual Research Assistant was hired and finished extensive training that included the proper way to approach potential Study candidates, the approved procedure to obtain informed consent for Study participation, the adherence to all Good Clinical Practice (GCP) guidelines for data collection activities, all aspects of project implementation and outreach efforts in the community.

Recruitment, Screening, and Evaluation of Children

Active recruitment of Study subjects began in October 2013 at Casa Monarca, a social service agency in a primarily Hispanic (Mexican) community in South Philadelphia. In April 2014 we started recruitment at additional sites serving primarily Hispanic families: the Visitation Center located at 2625 B Street in North Philadelphia, Little School Daycare located in Northeast Philadelphia, Daisy's Family Daycare located in North Philadelphia and Esperanza Health Center. In addition, we approached and began recruitment of subjects at several different African Churches located in West Philadelphia and Upper Darby, PA. Since the onset of active recruitment through June 30, 2014, the Study Team has approached 428 potential families. Among those, 51 (12%) were ineligible (based on language and/or child age requirements) and 52 (12%) indicated they did not wish to participate. The remaining 325 families were consented in writing, enrolled, and completed the appropriate first visit screening instruments which include the ASQ-SE, ASQ-3, SWYC, and M-CHAT. This number does fall somewhat short of our project milestone for this period, which was to consent and enroll 420 participants. Most of this short-fall can be accounted for by two factors. One is the inclement weather in February 2014, when only 19 participants were enrolled. Secondly, as seen in Figure 1, the average number of families being enrolled per month for the last three months (April-June) of this reporting period has reached more than 80, reflecting the successful expansion of recruitment to include additional sites, as described previously. Consequently, at the current pace of recruitment we fully expect to meet subsequent milestones for recruitment in the next project period. The project is viewed very positively by the agencies where recruitment is currently being performed, and is very well-received by those families who have been approached to participate. There have been no complaints from either the agencies or the families themselves, regarding either the overall purpose of the project or its implementation. Among the 52 parents of eligible children who chose not to participate, the overwhelming majority cited concerns related to being "too busy" to commit to a possible lengthy follow-up evaluation visit, and none raised an objection to nature of the project itself. The research protocol proposes the completion of a full clinical behavioral and developmental evaluation on at least 356 of the enrollees, which represents the 'gold standard' for assessing the accuracy of the screening instruments as described earlier in this report. During this project period, 117 participants were invited for a second study visit (full evaluation), of which 85 completed the evaluation. At this time, 18 people have been scheduled for a clinical evaluation**.** The remaining 14 participants who were invited and agreed to the clinical evaluation could not attend for various reasons. Given these results, and considering the increasing rate of enrollment, we fully expect to meet our goal of 356 full clinical evaluations by

June of 2016, as stated in the original proposal.

Establishment of the Executive Data Quality Committee

A Committee comprised of Dr. Jennifer Culhane (Principal Investigator), Dr. David Webb (Co-Investigator), Ximena Cuellar (Project Manager) and Leny Matthew (Statistician and Database Manager) was assembled and convened biweekly to review recruitment progress, data quality and review data collection strategies. Prior to each meeting a report was generated that included up to date information pertaining to the number of families approached, number of families eligible, number who refused, number who were consented and enrolled, and of those who were enrolled, the number of subjects scheduled and finally the number that had completed full clinical evaluations. This Committee reviewed these reports and discussed progress relative to project milestones and Study objectives and, when necessary, made recommendations for changes in procedures to ensure that the Project was on track to meet them. Additional reports that were generated and reviewed by the Committee included the number and percentage of children who screened positive for behavioral problems and/or developmental delays, and a detailed comparison of positive and negative screening rates across different assessment instruments. Detailed data quality reviews are regularly conducted by the Database Manager and reviewed by the Committee. The Research Electronic Data Capture (REDCap) system was designed with built-in data verification procedures including range-checks and safeguards to confirm that data skip patterns are appropriately followed. The data form entered into this electronic system was routinely checked for data entry errors using statistical software. The total scores and pass/fail status of the screening forms were verified against the total values of the individual items. Summary measures and response frequencies for the individual items, as well as for the supplemental survey questions were also checked for inconsistencies and data entry errors. The Database Manager reported the results of these data quality check procedures to the Committee at each of the bi-weekly meetings and, when needed, identified appropriate strategies to improve the data accuracy and quality.

Specific Aim 2

The second Specific Aim is to integrate all screening data collected with relevant, supplemental self-reported survey data (behavioral, socioeconomic and psychosocial) in order to *develop and evaluate a predictive model related to the detection of developmental and behavioral disorders in young children who enrolled in the study*. Significant progress toward this aim was achieved by successfully completing supplemental interviews which were administered to enrolled families, as well as establishing, improving, and maintaining a REDCap database system. This database system warehouses all data and provides for the eventual linkages of all data collected during the course of this project with other relevant datasets. As of June 30, 2014, all data from 323 of the 325 subjects currently enrolled has been entered into the REDCap database. The remaining two enrollees indicated that they preferred not to complete the survey. The REDCap database was designed to assign unique identification numbers to all families and children for data collection activities across all venues and components of the project. Consequently, the supplemental information from the surveys can be easily integrated with the screening data for behavioral and developmental disorders obtained at the initial visit, in addition to the assessment data from the full clinical evaluations obtained at the second visit. In brief, the database

infrastructure was established and fine-tuned in a manner consistent with the second aim of the project—integrating all of the data collected during the course of the Study in order to explore, develop and evaluate a statistical model that will prove to be useful in terms of reliably and accurately identifying developmental delays and behavioral problems in the Study population. This was accomplished through the creation and modification of a REDCap data system which provided a secure IRB-approved environment within the CHOP research and Information Technology (IT) divisions.

Figure 1. Number of Families Enrolled during Project period