Semi-Targeted Exposome Strategies to Measure Biomarkers of Exposure and Disease Associated with Type II Diabetes in Asian Indians

By
Sarah I. Daniels

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Health Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Martyn T. Smith, Co-Chair

Professor Luoping Zhang, Co-Chair

Professor Stephen M. Rappaport

Professor Alan Hubbard

Fall 2016

ABSTRACT

Semi-Targeted Exposome Strategies to Measure Biomarkers of Exposure and Disease Associated with Type II Diabetes in Asian Indians

By

Sarah I. Daniels

Doctor of Philosophy in Environmental Health Sciences

University of California, Berkeley

Professor Martyn T. Smith, Co-Chair

Professor Luoping Zhang, Co-Chair

The prevalence of type II diabetes (T2D) is escalating worldwide, yet incidence rates differ dramatically between ethnic groups. Some known risk factors of T2D include diet, exercise, and genetic inheritability, yet these factors alone cannot fully explain the differences observed between populations. Better approaches are needed to identify other non-genetic factors, such as toxic chemical exposures, that are related to T2D, particularly in highly-susceptible populations. The exposome is a relatively new concept used to investigate causes of disease due to endogenous and exogenous exposures. This dissertation aims to use semi-targeted exposomics to examine biomarkers of exposure and disease that are associated with T2D. Chapter 1, the introduction, discusses the broader objectives of exposomics, including ideal comparison populations for identifying underlying causes of chronic disease. In the case of T2D, Asian Indians are a population of great interest, with a 3-4-fold higher risk of T2D than European white counterparts. Chapter 2 examines blood concentrations of environmental pollutants in small volumes of plasma in Asian Indian immigrants and a low-risk comparison group, European whites. This study is the first to investigate associations between persistent organic pollutants and T2D in Asian Indians. Chapter 3 describes a method to measure sources of variability in a potential biomarker of T2D, microRNA (miRNA). This study demonstrates the importance of empirically measuring technical and biological sources of variability and using simulations to inform power calculations for new biomarkers of interest. Chapter 4 takes lessons learned from Chapter 3 and applies them to a case-control study on plasma miRNA in Asian Indians. In the concluding Chapter 5, the current state of research on T2D as it relates to blood biomarkers of environmental exposures and disease are described for Asian Indians and other highly susceptible groups.

Dedication

I dedicate this dissertation to the next generation of environmental health scientists. This body of work emphasizes the importance of scrutinizing the methods around “biomarker discovery” research. May the lessons learned here help guide the design, execution, and analysis of your future epidemiological studies.

Table of Contents

Acknowledgements

I am indebted to many people who have made this work possible and I am pleased to have the opportunity to acknowledge them here. I apologize to those whose names I have thoughtlessly omitted.

This project was born out of the need to find new strategies to identify environmental factors that contribute to chronic illnesses. In leading by example, Martyn Smith showed me how to be a pioneer in a new paradigm, and brought a fresh context to my understanding of the relationship between our environment and human health. He provided a challenge and a counter-argument, all while teaching me the art of effective risk-taking and risk-management (both scientifically and professionally). His foresight as well as his confidence in my own ability, allowed me to accomplish much more than I ever anticipated during my graduate career.

While there were stumbling blocks to overcome along the way to distract or deter me, I was able to stay on track thanks to the guidance of Luoping Zhang. She continually kept me focused, and she encouraged me to pay attention to small details in my experiments and writing style in order to maintain a defendable storyline. I thank her for helping me find the determination needed to complete a project and see the fruit of my labor. These are critical skills that I will carry with me throughout my career.

My other committee members have greatly contributed to my progress in the program as well. Stephen Rappaport, also at the forefront of this environmental health "paradigm-shift," has carefully checked over my research through the PhD process. Steve's breadth of knowledge in exposure biology, biomarker measurement, and related analyses gave me a new viewpoint to assess my own studies. Alan Hubbard offered critical review of the statistics used in this dissertation and helped me determine which analyses were most appropriate for my datasets. Alan's constructive feedback was an invaluable asset to my work. I hope to harness the tools that I have gained from both Alan and Steve in my future research.

The field-study portion of my dissertation would not have been possible without our collaborators in London, UK. Our host professors at Imperial College, Paul Elliott, Jaspal Kooner, and John Chambers, were very attentive upon our arrival to Ealing Hospital and made sure that we obtained the resources needed to carryout several months of blood collection. I also send my appreciation to the friendly staff at Ealing Hospital, including the physicians, nurses, locums, interns, and technicians within the Cardiology Department. Thank you for making us feel welcome each day and working tirelessly with us, both on weekdays and weekends.

In order to measure environmental pollutants using the latest advancements in technology, I had the privilege of working with mass-spectrometry specialists. Dr. Anthony Macherone, a head chemist at Agilent Technologies, developed a new method for small volumes of plasma and considered use of our subject samples for pilot screening. His hands-on involvement during this initial study allowed us to generate more targeted hypotheses for the follow-up study. Matthew McMullin and his colleagues at the National Medical Services (NMS) Labs conducted the additional follow-up study. Before performing the experiments, Matt kindly spent time going over the study design with me over-the-phone. After receiving the raw data,

Anthony patiently walked me through the software for peak integration and quantification (via cyberspace and phone), answering all my questions for each and every analyte.

I was fortunate to be surrounded by esteemed experts that I could consult whenever I had specific questions on topics that were outside my realm of knowledge. I am appreciative of the time and energy dedicated by Reuben Thomas, a biostatistician formerly working with our lab group, who played a major role in the third chapter of my dissertation. Thanks to Reuben, I learned the importance of prioritizing experimental design and empirically testing for sources of variability. Reuben also helped me gain more experience with the R software language. I have much gratitude for the graduate students and post-docs in the Rappaport Lab who also helped me along the way. Kelsi Perttula, Lauren Petrick, and Will Edmands gave great suggestions on how to set-up a mass-spectrometry "run" that accounts for noise, drift, and variability. This essential information was relayed to NMS Labs for successful execution of experiments for Chapter 2. For this same chapter, I am also appreciative of the thorough editing job and presentation suggestions that I received from Michele LaMerrill, professor at UC Davis.

It was a true pleasure to work with the current and former members of the Smith Lab. In particular, thank you to Fenna Sille, who was always willing to lend an ear about the technicalities of my research and helped me view my research from a new perspective. Cliona McHale gave innumerable suggestions and helpful hints in preparation for oral and written presentations of my work. Sylvia Sanchez did a tremendous job during our extended field research in London and made the blood collection portion of the study run as smoothly as possible. Rosemarie de la Rosa generously donated her time to helping write Chapter 1. Kipp Akers gave insight on particular topics related to bench work and the PhD process, in general, continuing to answer questions even after his own exit from the lab.

I was very lucky to work with stellar undergraduate students during my PhD. I'd like to thank my first student, Bin Tu, as she set a high standard of technique and detail and also gladly taught others to follow this regiment. Thank you to Alice Chang for her meticulous work in progressing the *in vitro* portion of my master's thesis and helping begin the research for Chapter 3. Brenda Yee and Audrey Goldbaum dedicated both weekdays and weekends to run experiments with me on qPCR to meet the deadline for a manuscript submission (ultimately becoming Chapter 3 of my dissertation). Brenda, Ellen Key, Kevin Zhang and Justin Kim, contributed to the qPCR data generated for the fourth chapter of my dissertation. Audrey, Ellen, and Irtaza Haider also did a terrific job in perusing PubMed and individual journal articles for the literature review table in Chapter 4 as well. I simply could not have finished the massive amount of bench work without their combined efforts and I am truly grateful for the countless hours that they devoted to the project.

My dear friends, who live both near and far, were continuous shoulders to lean on during this process. The community that I built through my time in the Berkeley Student Cooperative and at local dance studios brought a different way of thinking about certain aspects of my own research. I send my esteemed gratitude to those that I've bonded with over these past years, both in and outside of these circles.

This dissertation was facilitated by the perpetual support of my family. I thank my parents for standing by me through the journey and giving me the opportunities that I have today. My father's passion for the environmental and human health is contagious, and has led me to further my education in this field. From an early age, I appreciated him instilling in me the urgency of the environmental problems that we face. My mother's optimism has been a crucial factor in my success, as well as her demonstration of how to listen sincerely and respond thoughtfully. My sister, Jessica, has been a stronghold, sharing objective opinions that hold great weight in my decision-making. No matter the distance she and I live apart, I don't take this for granted. My grandparents provided their eloquent words of wisdom (along with heart-warming meals) during the most trying moments— a testament to their own perseverance amidst adversity. My aunt, Gaye, a constant cheerleader through thick-and-thin, knew how to rally in my favor every time we spoke. And the positive words of encouragement from my extended family in Southern California and the Netherlands have been timed just when I needed them most.

Lastly, Rinaldo Pietrantonio always reminded me that we, as humans, are driven to take a deep and contemplative look at the mechanics that steer the world around us. To do this, he never feared the humility in learning something entirely new. I watched in awe as he completed his second PhD, in a different language, in a different country, and in an entirely separate field than his first PhD. Perhaps I'll match him one day.
Repose en paix, Rinaldo.

Chapter 1: Using Exposomics to Assess Cumulative Risks and Promote Health*

Martyn T. Smith, Rosemarie de la Rosa and Sarah I. Daniels

Superfund Research Program, Division of Environmental Health Sciences,
School of Public Health, University of California,
Berkeley, California 94720-7360

*A similar version of this manuscript has been published: Martyn T. Smith, Rosemarie de la Rosa and Sarah I. Daniels. Using Exposomics to Assess Cumulative Risks and Promote Health. Environmental and Molecular Mutagenesis. 2015 Dec;56(9):715-23. doi: 10.1002/em.21985. Epub 2015 Oct 17. This chapter is printed here with acknowledgement to all co-authors and Wiley Online Library.

ABSTRACT

Under the exposome paradigm all non-genetic factors contributing to disease are considered to be 'environmental' including chemicals, drugs, infectious agents and psycho-social stress. We can consider these collectively as environmental stressors. Exposomics is the comprehensive analysis of exposure to all environmental stressors and should yield a more thorough understanding of chronic disease development. We can operationalize exposomics by studying all the small molecules in the body and their influence on biological pathways that lead to impaired health. Here, we describe methods by which this may be achieved and discuss the application of exposomics to cumulative risk assessment in vulnerable populations. Since the goal of cumulative risk assessment is to analyze, characterize, and quantify the combined risks to health from exposures to multiple agents or stressors, it seems that exposomics is perfectly poised to advance this important area of environmental health science. We should therefore support development of tools for exposomic analysis and begin to engage impacted communities in participatory exposome research. A first step may be to apply exposomics to vulnerable populations already studied by more conventional cumulative risk approaches. We further propose that recent migrants, low socioeconomic groups with high environmental chemical exposures, and pregnant women should be high priority populations for study by exposomics. Moreover, exposomics allows us to study interactions between chronic stress and environmental chemicals that disrupt stress response pathways (i.e. 'stressogens'). Exploring the impact of early life exposures and maternal stress may be an interesting and accessible topic for investigation by exposomics using biobanked samples.

The Exposome and the New Field of Exposomics

Several definitions of the exposome now exist. Wild originally defined the "exposome" as representing all environmental exposures (including those from diet, lifestyle, and endogenous sources) from conception onwards, as a quantity of critical interest to disease etiology [Wild, 2005]. His goal in doing so was to articulate the need for new tools to assess environmental exposures from all sources for studies of adverse gene-environment interactions as causative factors in chronic disease.

As toxicologists we recognize that adverse effects on the body's tissues and organs are related to the concentration of chemical agents circulating in the biofluids that bathe the tissues, notably the blood plasma and lymph. This internal dose of the chemical or drug is directly related to the toxicity and biological effects at given concentrations. Thus, when Rappaport and Smith considered how Wild's original exposome concept could be measured, they concluded that this could best be achieved by monitoring the internal chemical environment of the human body during critical windows of exposure (i.e., measuring "snapshots") [Rappaport and Smith, 2010]. They also recognized that all chemical and non-chemical stressors mediate effects on the body via signaling of small molecules that alter cellular activity and physiological processes. For example, during emotional stress our adrenal glands release adrenaline (also known as epinephrine) and other hormones into the bloodstream that increase breathing, heart rate, and blood pressure. Thus, if one wants to consider all non-genetic factors that influence health, it is reasonable to consider the "environment" as the body's internal chemical environment and "exposures" as the amounts of biologically active chemicals (small molecules) in this internal environment that stem from both exogenous and endogenous sources.

The new field of exposomics should therefore attempt to measure as many small molecules as possible in human bodily fluids. A million molecule exposome is a potential goal that is not too unrealistic. Further, it should attempt to link the presence of these small molecules with functional changes in biology leading to chronic illnesses. The internal measurements made in exposomics could be of individual chemicals, groups of chemicals or the totality of chemicals acting on a particular receptor or biological pathway in a functional assay. Hence, exposomics can be operationalized by studying all the small molecules in the body and their influence on biological pathways that lead to impaired health. This concept of exposomics fits with the revised definition of the exposome proposed by Miller and Jones that explicitly incorporates the body's response to environmental influences [Miller and Jones, 2014]. They argue that the exposome and biology are interactive and that changes in biology due to the environment may change one's vulnerability to subsequent exposures. Further, Miller and Jones argue that by studying the effects of exposures we may gain insight into past chemical exposures as they may leave a molecular fingerprint. Thus, through linking exposures to specific biological responses, exposomics could serve as an approach to gain insight into the mechanistic connections between a culmination of exposures and risk of adverse health outcomes that occur over a lifetime.

Environmental Stressors, Exposure Assessment and the Exposome

Another, entirely different approach to examine the relationships between environmental exposures and disease is to measure exposures to various environmental stressors through wearable and regional sensors and survey instruments. These are being used, for example, to measure exposure to air pollution and drinking water contaminants; to better assess the diet through smartphone capture of dietary habits; and, to evaluate exercise through pedometers and other devices. This is how measurement of the exposome was conceptualized in a NAS committee report on exposure science and was expanded to the term eco-exposome so as to include wildlife as well as humans [Committee on Human and Environmental Exposure Science in the 21st Century and Board on Environmental Studies and Toxicology, 2012]. Sensors and 21st century exposure science tools are, of course, useful for improving exposure assessment in targeted epidemiology studies of specific risk factors such as physical exercise, diet, and air pollution and for avoiding known risks through smartphone applications and other mechanisms. These exposure science tools are limited, however, in their ability to identify novel environmental causes of disease, but in combination with internal exposomics tools they could be a powerful approach to assessing an individual or community's exposome.

Towards Measurement of a Community's Exposome at the Individual and Group Levels

Measuring environmental pollutants has become a subset of an even broader initiative termed the "Public Health Exposome", coined by Juarez, which captures an assessment of risk at the community level, including the influences of the natural, built, social, and policy environment [Juarez et al., 2014]. The natural environment includes chemicals in air, water, soil, and food. The built environment includes quality of the workplace, educational centers, places of worship, and playgrounds as well as access to commercial businesses and public transportation. The social environment includes rates of discrimination, poverty, crime, unemployment in the surrounding area and moderating factors such as social networks, capital and integration. Lastly, the policy environment represents local rules and regulations that influence the quality of public health services and exposures.

This public health exposome approach incorporates exposures at the ecological-level to determine the impact on the overall health of a population within a particular region. One of the first studies on the public health exposome, included over 600 variables for counties throughout the U.S. to better understand determinants of preterm birth [Kershenbaum et al., 2014]. Interestingly, a unique clustering method distinguished between "resilient counties" with low preterm birth rates nestled within high-risk regions. Hence, identification of resilient versus susceptible sub-groups may be key in deciding optimal target populations for comparison or intervention studies in exposomics.

The public health exposome can uncover plausible sources of social determinants of health that contribute to the internal exposome. In Table 1, we expand upon this framework proposed by Juarez et al. by providing examples of biological mechanisms disrupted through various community level exposures. Exposomics would allow detection of these biological responses and, furthermore, assessment of the overall health impacts (Table 1). This may be a particularly novel approach for assessing cumulative risk in the community setting.

Using Exposomics to Assess Cumulative Exposures and Cumulative Risk

From this discussion, one can see that the health of a given community, and the individuals within it, is dependent on a variety of environmental and social factors. The EPA defines cumulative risk assessment as, "Combined risks from aggregate exposures to multiple agents or stressors, where agents or stressors may include chemical and nonchemical stressors" [US EPA, 2003]. This is essentially the exposome paradigm where all non-genetic environmental stressors are considered. Therefore, cumulative risk assessment, where the impact of all stressors on a population is assessed, could be operationalized by exposomics (Figure 1).

There has been little effort so far to examine the totality of both chemical and non-chemical stressors on a population. Initial observations of low-income, race, and other socioeconomic factors exacerbating the effects of individual chemical exposures have been reported [Shankardass et al., 2009; Vishnevetsky et al., 2015; Zota et al., 2013]. New agnostic methods can be applied to identify candidate chemicals that exacerbate disease risk via interaction with effects of the social environment. Exposomics could be used for the discovery of environmental chemicals that interfere with stress response pathways that are chronically activated by adverse social environments.

Bruce McEwen was the first to propose that prolonged activation of these stress response pathways causes "wear and tear" on regulatory mechanisms, adjusting the homeostatic set point of various physiological systems [McEwen, 1998]. This cumulative burden on the body is referred to as the allostatic load and is quantified using a cumulative index of physiologic deregulation of the cardiovascular, inflammatory, and endocrine systems (Fig. 2). While there is evidence that increased allostatic load and stressful life-experiences enhance vulnerability to the adverse health and behavioral effects of chemicals [Shankardass et al., 2009; Vishnevetsky et al., 2015; Zota et al., 2013], it is unclear how these "natural" and "social" environments work in concert to cause disease.

An exposomics approach would quantify endogenous primary mediators found in the blood, such as cortisol and adrenaline, to obtain a measurement of "allostatic load". Cortisol, secreted by the adrenal gland in response to stress, activates the glucocorticoid receptor (GR) and has systemic effects on the endocrine, metabolic, cardiovascular, immune, reproductive, and central nervous systems [Sapolsky et al., 2000]. Environmental chemicals that mimic cortisol can disrupt stress response pathways through altered GR signaling [Odermatt et al., 2006; Odermatt and Gumy, 2008]. We define these environmental chemicals that alter stress response pathways as "stressogens."

Within the context of exposomics, it is essential to obtain a measure of the totality of stressogen burden within subject samples. Recently, by using a functional bioassay that measures glucocorticoid receptor activity, we have identified a number of stressogens, including the morning-after pill RU486, that perturb the stress response by exerting either agonistic or antagonistic effects on GR (Figure 2). We are now applying this assay to identify additional environmental chemicals that may act as stressogens and to measure the totality of chemicals acting on GR in an individuals blood plasma. Exposomic classification of stressogens and detection of endogenous stress response mediators moves us one step closer to

developing more holistic models of attributable risk factors of disease, particularly among vulnerable populations with substantial mixing of multiple sources of environmental stress.

Targeted and Untargeted Methods to Measure Snapshots of the Exposome

To measure snapshots of the exposome, we must be able to quantify exposure to and the impacts of all non-genetic factors including chemicals, drugs, dietary components and supplements, psycho-social stress, infection, and ionizing radiation during critical stages in the life course. This is clearly a major challenge but seemingly not an impossible one. By focusing on classes of chemicals with probable effects such as electrophiles and chemicals that target specific receptors we may be able to assess the impacts of many of the chemicals in commerce (Table 2). Further, modern mass spectrometry now allows us to measure pharmaceuticals, vitamins and other dietary components with relative ease and is being expanded to untargeted methods which measure thousands of molecular ions (Table 2). Psycho-social stress could be measured by various markers including telomere length, cortisol and amylase levels and activity through stress response pathways such as GR (Table 2). It is also important to measure current and prior exposures to infectious agents, as they can play an important role in chronic disease development.

There is some debate over the best strategies to use for exposomics research, given the limitations of both targeted and untargeted methods. While untargeted methods provide promise in examining thousands of molecules simultaneously, some sensitivity is sacrificed in measurement of low abundance compounds. It has been previously observed in the literature that the majority of pollutants are at 100-1000 times lower concentration than drugs and dietary components [Rappaport et al., 2014]. While this begets the need for targeted methods with improved sensitivity, it is important to incorporate both in exposomics research.

The advantage of untargeted methods is the potential for discovery of novel analytes while measuring hundreds to thousands of compounds simultaneously. This technique has been demonstrated successfully in previous cases [Wang et al., 2011a, 2011b], however, given the statistical limitations of these methods, the likelihood of obtaining reproducible findings still remains small. To improve upon characterization of “biologically active” molecules in the blood by metabolomics, the method could be paired with other assays to quantify the net potential effect of endogenous and exogenous compounds in human serum. These preliminary screening methods may allow discrimination between analytes of interest and background noise that are measured using untargeted approaches (i.e., metabolomics). An example of such methods is use of receptor-binding reporter assays in responses to chemicals in human blood samples. Currently we are using sensitive CALUX receptor-based reporter bioassays, which measure the overall net effect of both endogenous and exogenous molecules acting on a particular receptor simultaneously. (e.g., This includes the GR receptor activity from stressogens, as described earlier.) This high-throughput and inexpensive method of detecting total endocrine activity of serum against a particular receptor can be scaled-up, as previously done for purposes of chemical screening within ToxCast and Tox21.

Several methods are being explored to isolate the candidate active agonistic/antagonistic compounds from serum. For example, the serum can be fractionated by HPLC, and then the fractions can be applied separately to the receptor assays to measure

activity of endogenous hormones versus exogenous chemicals [Bonefeld-Jorgensen et al., 2011]. Another method is to use receptor affinity extraction liquid chromatography to first isolate the chemicals that bind to the column and then elute the bound chemicals for further profiling by LC-MS/MS [Hock, 2012]. This has been improved upon by immobilizing the receptor ligand binding domain, which has more stable binding affinity than the entire receptor and still maintains high sensitivity to xenobiotics. While this method was originally demonstrated with ERα [Pillon et al., 2005], it can be expanded to other binding domains as well [US EPA]. Lastly, active molecules could be identified by running the serum in tandem on both the bioassay and an HPLC-MS/MS instrument, and modeling differences in average peak sizes between comparison populations in association with reporter signals. These agnostic methods provide an exposomic approach to detect novel endogenous and exogenous exposures that influence cellular function.

Targeted methods of past and current exposures are also useful for examining chemical compounds that are known to be pervasive and/or bioaccumulative in the environment. With improved resolution of instrumentation, smaller volumes are needed than before to assess levels of these chemicals in bodily fluids. For example, Agilent Technologies has recently developed a method using a quadrupole GC-MS/MS system using only 200uL of plasma/serum to measure more than 60 POPs including PCBs, PBDEs, OCPs, PAHs, furans, and dioxins [Macherone et al., 2015]. This has potential for scale-up to measure even more compounds. Plasma is extracted using chemical denaturation, liquid-liquid extraction, solid-phase cleanup and reconstituted with isooctane. This targeted GC MS/MS method exemplifies improvements in measuring differential POPs exposure profiles over those previously used by the CDC (NHANES) and others by reducing volumes of precious blood samples by at least 10-fold. The limits of detection are 0.005–0.02 ng/mL for PCB; 0.05–0.15 ng/mL for OCP; 0.0075–0.075 ng/mL for PBDE. Targeted methods like these should be restricted to chemicals such as POPs with known persistence in the environment and association with harmful effects. Interestingly, given the long half-life of these pollutants, previous exposure and migration patterns can be chronicled, particularly among populations that have migrated from highly exposed to lower exposed areas during their life.

Including Measurement of Exposure to Infectious Agents in Exposomics Research

New advancements in detecting past and current exposure to infectious agents allows for expansion of this branch of exposomics. Recently, a screening procedure, called VirScan, has demonstrated extreme sensitivity and specificity for detecting antibodies against previous infections in just 1uL of serum [Xu et al., 2015]. The VirScan target library is based on the viral proteome sequence database within UniPro [Consortium, 2014] and includes 206 known viral species and over 1000 different strains. Several strategies have been used to discover novel non-human sequences in the human transcriptome including digital transcriptome subtraction[Feng et al., 2008]. To detect such integrated viral sequences, algorithmic methods such as VirusSeq scan either RNA Seq or whole genome data for viruses that map to a viral database [Chen et al., 2013]. More recently, “sequence-based ultrarapid pathogen identification,” SURPI, was developed to assess both known and novel bacterial, viral, fungal, and parasitic sequences in human tissue samples [Naccache et al., 2014]. Published NGS data

can also be scavenged for novel discovery of new emerging infectious agents. This was exemplified using metagenomics data from fecal samples of twins and their mothers from a public database and then extrapolating to verify findings of a new bacteriophage in over 900 samples [Dutilh et al., 2014]. These new techniques to study current and previous infections in population studies are imperative to understanding their relationship with other exposures and disease onset within the exposome.

Exposomics research relies on understanding the interactions of both past and present exposures to chemical and non-chemical agents, but there are few studies that have examined links between environmental exposures and susceptibility to new or recurrent infection. Previous work has focused on early-life exposure to individual environmental pollutants and increased incidence of viral infections. Associations have been found between early-life exposure to persistent organic pollutants such as PAHs, dioxins, and PCBs and increased risk of flu-like symptoms, and respiratory and ear infections [Winans et al., 2011]. There is also evidence of altered immune function with early-life exposures to heavy metals such as arsenic [Rager et al., 2014] and increased mortality from infection due to arsenic exposure [Smith et al., 2010]. Exposomics has the capacity to expand upon these findings by examining how interactions between numerous chemical and non-chemical stressors increase risk of disease by infectious agents.

In Which Populations Should We Do Exposomics?

If exposomics is to perform an agnostic search of many different environmental exposures, populations with the highest "totality of exposures" are of primary interest. Attention should focus on vulnerable environmentally-exposed populations, as the risks of chronic illnesses are higher than in the general population. Examples of these "at risk" groups in the U.S. are undisputedly minority populations living in urban or agricultural settings. This is exemplified by the CalEnviroScreen 2.0 [Faust et al., 2014], which maps scores by county based on the pollution burden and population characteristics of the region. Counties with the highest (most severe) scores are invariably concentrated in low-income regions of densely populated city centers or the agricultural Valleys of California. Thus, these populations could be sampled and compared to adjacent populations with lower CalEnviroScreen scores.

Another population that may be well-suited to exposomic analysis is pregnant women and their newborn infants. Bio-banked samples of mid-pregnancy maternal blood, cord blood and Guthrie card blood spots could be used for exposomic analyses in relation to fetal growth, pre-term delivery, birth defects and other early life outcomes. Methods for the rapid analysis of these biobanked samples should be developed and applied in well-controlled epidemiological studies.

Immigrant populations are another exemplary group for exposomics research. These populations were exposed to different environmental and non-chemical stressors in early-life and may have made changes in behavior due to acculturation as compared to the native populations in their new and former residences. This leads to profound differences in disease incidence rates that could be driven by both environmental and genetic factors. Given the unique conditions of immigrant populations, several strategies in study design could be employed to help parse apart environmental from genetic factors. For instance, some

populations continually immigrate to the same region for generations, making it possible to measure the exposomics profiles associated with the number of years since emigration as compared to first-generation, non-immigrant, and non-emigrating populations, all with similar genetic background. Trans-generational effects on the immigrant population can be explored as well. Furthermore, differences in exposomic profiles between countries or regions of emigration could also be used to map genetic and non-genetic contributions to disease onset.

Using the exposomics approach to conduct cumulative risk assessments would be an excellent opportunity to examine differences in disease onset in immigrant populations. For example, this approach may help to resolve enigmas such as the "Hispanic Paradox," which is described as similar rates of health outcomes (including infant mortality, life-expectancy, and mortality from CVD and major types of cancer) among immigrant Hispanic populations compared to whites, despite lower socioeconomic status [Markides and Coreil, 1986]. This effect dissipates with acculturation [Burgos et al., 2005]. Taking an exposome approach would incorporate previous observations of differences in early-life nutrients, chemical exposures, stressogens, and non-chemical stressors into a single study, providing a more comprehensive assessment of exposure.

Another exemplary population for exposomics is the "South Asian Phenotype" of diabetes. This group is deserving of further investigation as South Asians are at 4-fold higher risk of type 2 diabetes (T2D) as compared to Caucasian populations and begin to obtain insulin resistance at a relatively lower BMI and younger age of onset than Caucasians (reviewed in [Bakker et al., 2013]). While there has been individual studies to examine effects of low-birth weight, diet, chemical exposure, the *in utero* environment , and even mitochondrial activity in relation to T2D (reviewed in [Bakker et al., 2013]), an exposomics approach would take all these factors into account to explain this unique phenotype in these immigrant populations.

We are currently pursuing exposomic studies in both Hispanic and Indian populations. Specifically, our two study populations of interest are 1) a case-control subset of foreign-born and native Mexican American females from the San Francisco Bay Area Breast Cancer Study, comprised of 5,000 Hispanics, African-Americans, and non-Hispanic whites 2) a cross-sectional study of Asian Indian immigrants and native European whites residing in Greater London and nested within a continuing cohort, called the London Life Sciences Prospective Population (LOLIPOP) Study. While distinct outcomes (breast cancer versus type II diabetes) and populations are considered in these two studies, the exposomics methodology is similar for both. Improved understanding of the role of endogenous and exogenous compounds on endocrine response is imperative for both breast cancer and diabetes. We will examine hormone receptor activation of all small molecules in the serum using luciferase reporter bioassays. Then we will profile subjects with extremes of activity by untargeted high-resolution mass spectrometry (HRMS) of small molecules in the serum to determine which chromatogram peaks may be responsible for the widely-differing levels of receptor activity and are associated with disease onset.

For both of these studies, significant inheritable findings have already led to exciting progress in the respective disease fields. For the Latina population study, a protective SNP variant was identified 5' of the estrogen receptor 1 gene in those of Indigenous American descent [Fejerman et al., 2014]. For the LOLIPOP cohort, six unique genetic variant loci in six

separate genes reported specifically for Asian Indians— three genes which were directly linked to insulin sensitivity and pancreatic beta-cell function [Kooner et al., 2011]. Given these strong inheritable components of disease within these susceptible sample populations, Genome x Exposome interactions will be of great interest once we obtain exposomics data.

Use of Strategies for Genomics Analysis to Inform Exposomics Analysis

Overall, as genomics is the oldest and most advanced omics field, similar strategies employed for GWAS studies could be applied to exposomics. For example, the ease of genome sequencing today has facilitated the study of pleitrophy, defined as a single locus being responsible for multiple phenotypic traits. This is an important concept in studying inheritability of complex diseases (e.g., mental disorders, metabolic syndrome, and cancers)(reviewed in [Yang et al., 2015]). Moreover, pleitrophy in genomics can lend to novel findings of differences in environmental exposures. In a recent study using VARIMED (VARiants Informing MEDicine), a manually curated database of disease-SNP associations, an association was found between gene variants in three genes, gastric cancer and serum magnesium levels [Li et al., 2014]. In a follow-up assessment of medical records, the magnesium levels were altered 1-year prior to gastric diagnosis. We must consider how individual chemicals can have multiple targets in the body simultaneously and can increase risk of multiple phenotypic outcomes. Improved databases, such as ToxCast, that provide evidence of the relationships between chemical exposures and phenotypic traits will help guide the direction of appropriate chemical analysis in exposomics research.

As most chronic illnesses are multi-factorial, it is expected that multiple exposures may be involved with disease onset. The idea that particular exposures can be “inherited” together is an important concept that is likely to be specific per population. In the analysis phase it is important to consider the similarities of particular chemicals in structure and mechanisms of actions against a given biological target, thus simplifying the combined effects of many exposures. Patel et al demonstrated correlations between particular exposures and the importance of recognizing these clusters [Chirag J Patel and Arjun K Manrai, 2014]. The paper draws an analogy to linkage disequilibrium of the genetic code, and how we must not think of every SNP as unique. This comparison could be expanded upon in consideration of other traits at the community level, including social determinants of health as those described by Juarez et al. [Juarez et al., 2014]

Conclusions and Recommendations

Under the exposome paradigm all non-genetic factors contributing to disease are considered to be ‘environmental’ including industrial chemicals, drugs, infectious agents and psycho-social stress. It is perhaps best to consider these as environmental stressors. Exposomics is the comprehensive analysis of exposure to all environmental stressors and should yield a more thorough understanding of chronic disease development. Since exposomics can be performed at the individual as well as the population level it could have a broad impact on personalized preventative medicine, policy changes, and our understanding of disease mechanisms (Figure 3).

Exposomics can also be used in the context of cumulative risk assessment. Since the goal of cumulative risk assessment is to analyze, characterize, and quantify the combined risks to health or the environment from exposures to multiple agents or stressors, it seems that exposomics is perfectly poised to advance this important area of environmental health science. We should therefore develop and apply exposomics to issue of cumulative risk and support development of tools for exposomic analysis. We should also begin to engage impacted communities and develop the public health exposome concept of Juarez and others. A first step may be to apply exposomics to vulnerable populations already studied by more conventional cumulative risk approaches. Moreover, inferences made from these exposomics studies within the context of cumulative risk assessment may be translated to policymakers for promoting change in environmental exposure regulations.

Exposomics allows us to study interactions between chronic stress and environmental chemicals and to discover environmental chemicals that may disrupt stress response pathways. We have named such chemicals 'stressogens' as they have the ability to influence how our bodies respond to stress. For example, exploring the role of environmental exposures and chronic stress in pre-term delivery may be an interesting topic for investigation by an exposomic approach. We further conclude that susceptible groups (migrants, low socioeconomic groups with high environmental exposures, pregnant women) should be the study populations of interest for exposomics. Physicians who work with these populations nationwide and worldwide can use exposomics to work towards earlier identification of high-risk individuals and communities and ultimately disease prevention.

Finally, we highlight the importance of not "reinventing the wheel" when it comes to analysis of large amounts of data that will clearly be generated by exposomics studies. Collaboration with bioinformaticists and biostatisticians skilled in analyzing genomics data and other patterns will be essential. This is an exciting time for scientific collaboration across disciplines, and using exposomics research may be transformative in our understanding of the causes of adverse health outcomes in human populations.

Acknowledgments
We thank Sylvia Sanchez, Fenna Sille, Laura Fejerman, Anthony Macherone, and Martin Kharrazi for important discussion and insights. Supported by NIH grant P42 ES004705 from the National Institute of Environmental Health Sciences and award 21UB-8009 from the California Breast Cancer Research Program.

Statement of Authors Contributions

All authors contributed to the conception and writing of this article.

Conflicts of Interest Statement

The authors declare no conflicts of interest for the opinions expressed in this article.

Tables and Figures

Figure 1. Cumulative Risk Framework including the Exposome

(based on Morello-Frosch & Shenassa, *EHP* 114, 1150, 2006)

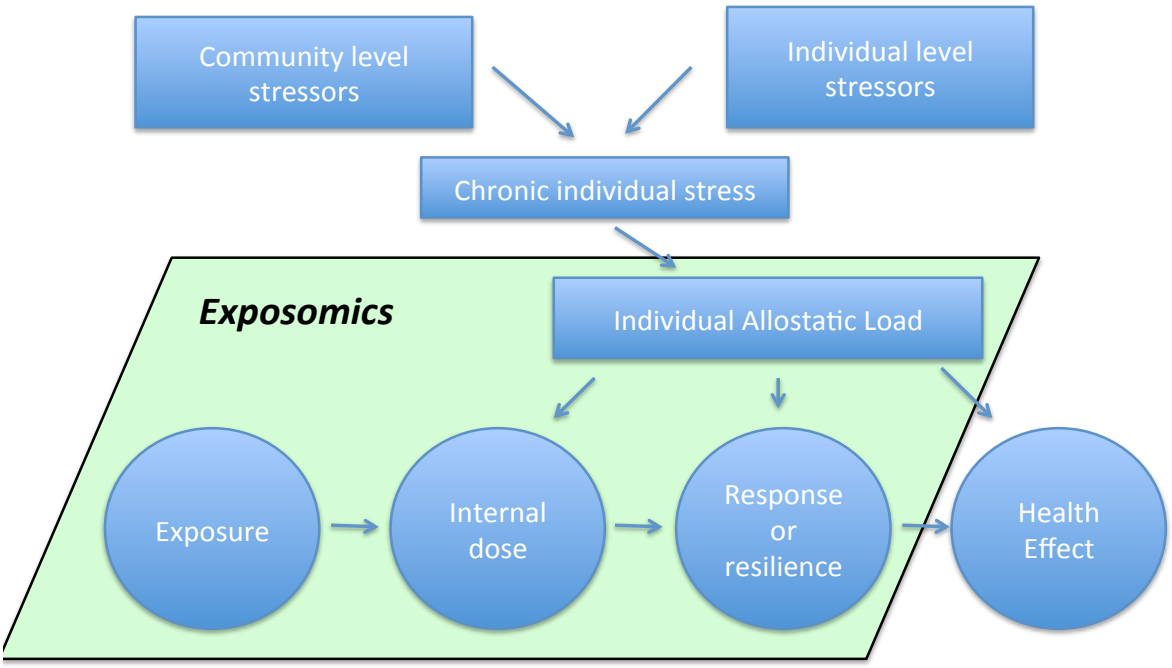


Figure 2. **Bioassay modeling disrupted glucocorticoid receptor (GR) signaling**. Cortisol is a GR agonist. RU486 is a chemical antagonist that inhibits cortisol activation of GR. The box represents the endogenous cortisol range of 193-690nM. Environmental stressogens may act as an agonist like cortisol or as antagonists like RU486

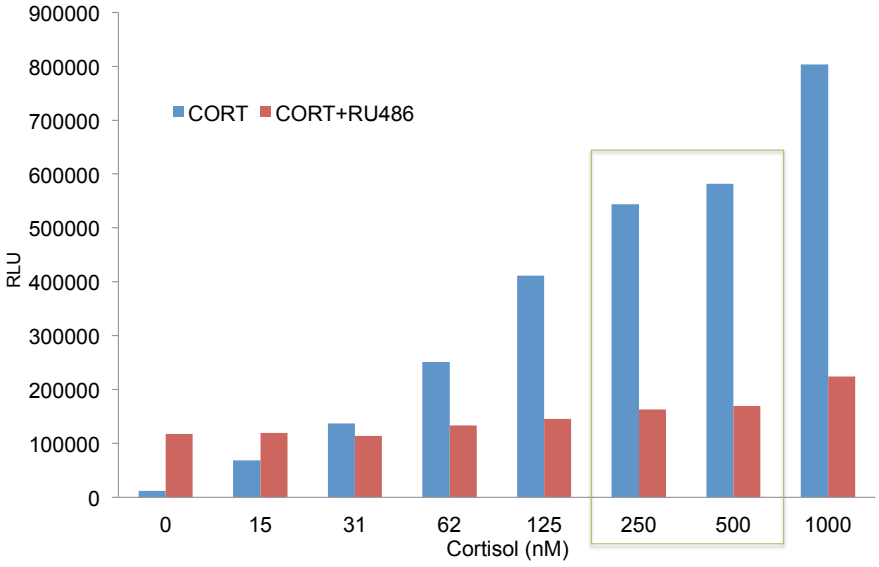


Figure 3. How Exposomics Could Contribute to Disease Prevention

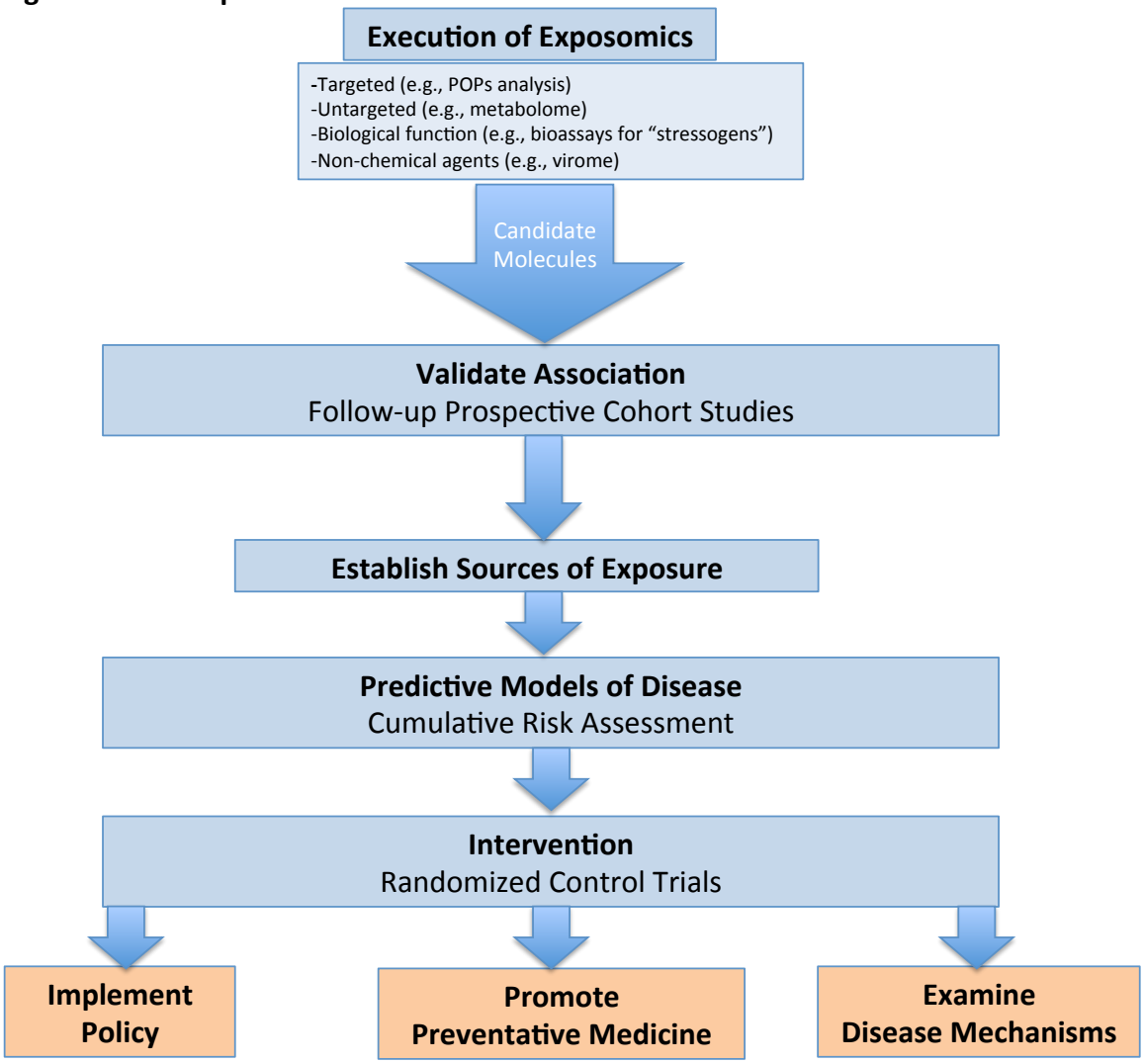


Table 1. Exposomics in the context of the "Public Health Exposome" [Juarez et al., 2014]

Environment Type	Examples	Biological Response	Health Impact
Natural	• Quality of air, water, soil, food • Chemical contamination	• Inflammation, reactive oxygen species, protein/DNA adducts, • Methylation and gene expression changes	Chronic diseases including cancer and diabetes
Built	• Quality of workplace and housing • Presence of educational centers, places of worship, playgrounds • Access to fresh produce, commercial businesses, public transportation, greenery • Proximity to roadways	• Increased responsiveness to cortisol and "stressogen" on the glucocorticoid receptor • Changes in sex hormone levels and receptor responses	Stress and chronic health issues induced by poor living quality, and lack of resources and social interaction
Social	• Rates of discrimination, poverty, crime, violence, unemployment, gentrification, de facto segregation • Access to capital, loans, social services, law enforcement, education, and health care	• Increased adrenaline, resting heart rate, and blood pressure (vasoconstriction) • Altered brain function, structure and plasticity • Increased pro-inflammatory cytokine secretion	Psychological effects due to unsafe settings and turbulent activities near the home coupled with a lack of economic and community support
Policy	• impacts of state and federal regulations and laws • restrictive city ordinances • local rules • voting rights • housing laws • evident corruption • voice within town council	• Changes in concentrations of neurotransmitters (i.e. dopamine, serotonin, GABA)	Emotional insecurity and feelings of hopelessness due to inequality, disenfranchisement and lack of political representation

Table 2. Current techniques for exposomics

- Metabolomics: ~30,000 small molecules in untargeted analysis; Targeted analysis of 100-500 compounds.
- Targeted mass spectrometry: Can measure low levels of environmental pollutants
- Adductomics: Measures electrophiles binding to blood proteins
- Hormone receptor activation in cell based assays: Measure endocrine disruptors
- AhR cell based assay: Measures totality of persistent organic pollutants (POPs) and short-term transient activators
- Mass spectrometry and speciation of metals : ~20 easily measured
- Antibody arrays and subtractive sequencing: Measures current and past exposure to infectious agents
- Assays of telomere length, telomerase activity, CD28 cells, cortisol, amylase: Measures stress
- Oxidative stress markers: isoprostanes etc. (Panel)
- Markers of inflammation: cytokines, C-reactive protein (Panel)
- Early biomarkers of response/resilience: transcriptome, methylome, cellular immune response, etc.

References

Bakker LEH, Sleddering MA, Schoones JW, Meinders AE, Jazet IM. 2013. Pathogenesis of type 2 diabetes in South Asians. Eur. J. Endocrinol. Eur. Fed. Endocr. Soc. 169:R99–R114; doi:10.1530/EJE-13-0307.

Bonefeld-Jorgensen EC, Long M, Bossi R, Ayotte P, Asmund G, Krüger T, Ghisari M, Mulvad G, Kern P, Nzulumiki P, et al. 2011. Perfluorinated compounds are related to breast cancer risk in Greenlandic Inuit: a case control study. Environ. Health Glob. Access Sci. Source 10:88; doi:10.1186/1476-069X-10-88.

Burgos AE, Schetzina KE, Dixon LB, Mendoza FS. 2005. Importance of generational status in examining access to and utilization of health care services by Mexican American children. Pediatrics 115:e322–330; doi:10.1542/peds.2004-1353.

Chen Y, Yao H, Thompson EJ, Tannir NM, Weinstein JN, Su X. 2013. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. Bioinformatics 29:266–267; doi:10.1093/bioinformatics/bts665.

Chirag J Patel, Arjun K Manrai. 2014. Development of exposome correlation globes to map out environment-wide associations. In *Biocomputing 2015*, pp. 231–242, WORLD SCIENTIFIC.

Committee on Human and Environmental Exposure Science in the 21st Century, Board on Environmental Studies and Toxicology. 2012. *Exposure Science in the 21st Century: A Vision and a Strategy*. The National Academies Press, Washington D.C.

Consortium TU. 2014. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 42:D191–D198; doi:10.1093/nar/gkt1140.

Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, et al. 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. 5; doi:10.1038/ncomms5498.

Faust J, August L, Alexeeff G, Bangia K, Cendak R, Cheung-Sutton E, Cushing L, Kadir T, Leichty J, Milanes C, et al. 2014. California Communities Enivronmental Health Screening Tool, Version 2.0 (CalEnviroScreen 2.0) Guidance and Screening Tool.

Fejerman L, Ahmadiyeh N, Hu D, Huntsman S, Beckman KB, Caswell JL, Tsung K, John EM, Torres-Mejia G, Carvajal-Carmona L, et al. 2014. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. Nat. Commun. 5; doi:10.1038/ncomms6260.

Feng H, Shuda M, Chang Y, Moore PS. 2008. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. Science 319:1096–1100; doi:10.1126/science.1152586.

Hock B. 2012. *Bioresponse-Linked Instrumental Analysis*. Springer Science & Business Media.

Juarez PD, Matthews-Juarez P, Hood DB, Im W, Levine RS, Kilbourne BJ, Langston MA, Al-Hamdan MZ, Crosson WL, Estes MG, et al. 2014. The Public Health Exposome: A Population-Based, Exposure Science Approach to Health Disparities Research. Int. J. Environ. Res. Public. Health 11:12866–12895; doi:10.3390/ijerph111212866.

Kershenbaum AD, Langston MA, Levine RS, Saxton AM, Oyana TJ, Kilbourne BJ, Rogers GL, Gittner LS, Baktash SH, Matthews-Juarez P, et al. 2014. Exploration of Preterm Birth Rates Using the Public Health Exposome Database and Computational Analysis

Methods. Int. J. Environ. Res. Public. Health 11:12346–12366; doi:10.3390/ijerph111212346.
Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, Been LF, Chia K-S, Dimas AS, Hassanali N, et al. 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat. Genet. 43:984–989; doi:10.1038/ng.921.
Li L, Ruau DJ, Patel CJ, Weber SC, Chen R, Tatonetti NP, Dudley JT, Butte AJ. 2014. Disease Risk Factors Identified through Shared Genetic Architecture and Electronic Medical Records. Sci. Transl. Med. 6:234ra57; doi:10.1126/scitranslmed.3007191.
Macherone A, Daniels S, Maggitti A, Churley M, McMullin M, Smith MT. 2015. Measuring a slice of the exposome: Targeted GC-MS/MS analysis of persistent organic pollutants (POPs) in small volumes of human plasma.
Markides KS, Coreil J. 1986. The health of Hispanics in the southwestern United States: an epidemiologic paradox. Public Health Rep. 101: 253–265.
McEwen BS. 1998. Stress, Adaptation, and Disease: Allostasis and Allostatic Load. Ann. N. Y. Acad. Sci. 840:33–44; doi:10.1111/j.1749-6632.1998.tb09546.x.
Miller GW, Jones DP. 2014. The nature of nurture: refining the definition of the exposome. Toxicol. Sci. Off. J. Soc. Toxicol. 137:1–2; doi:10.1093/toxsci/kft251.
Morello-Frosch R, Shenassa ED. 2006. The Environmental "Riskscape" and Social Inequality: Implications for Explaining Maternal and Child Health Disparities. Environ. Health Perspect. 114:1150–1153; doi:10.1289/ehp.8930.
Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger AL, Luk K-C, Enge B, et al. 2014. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res. 24:1180–1192; doi:10.1101/gr.171934.113.
Odermatt A, Gumy C. 2008. Glucocorticoid and mineralocorticoid action: why should we consider influences by environmental chemicals? Biochem. Pharmacol. 76:1184–1193; doi:10.1016/j.bcp.2008.07.019.
Odermatt A, Gumy C, Atanasov AG, Dzyakanchuk AA. 2006. Disruption of glucocorticoid action by environmental chemicals: potential mechanisms and relevance. J. Steroid Biochem. Mol. Biol. 102:222–231; doi:10.1016/j.jsbmb.2006.09.010.
Pillon A, Boussioux A-M, Escande A, Aït-Aïssa S, Gomez E, Fenet H, Ruff M, Moras D, Vignon F, Duchesne M-J, et al. 2005. Binding of estrogenic compounds to recombinant estrogen receptor-alpha: application to environmental analysis. Environ. Health Perspect. 113: 278–284.
Rager JE, Yosim A, Fry RC. 2014. Prenatal Exposure to Arsenic and Cadmium Impacts Infectious Disease-Related Genes within the Glucocorticoid Receptor Signal Transduction Pathway. Int. J. Mol. Sci. 15:22374–22391; doi:10.3390/ijms151222374.
Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. 2014. The blood exposome and its role in discovering causes of disease. Environ. Health Perspect. 122:769–774; doi:10.1289/ehp.1308015.
Rappaport SM, Smith MT. 2010. Epidemiology. Environment and disease risks. Science 330:460–461; doi:10.1126/science.1192603.

Sapolsky RM, Romero LM, Munck AU. 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. Endocr. Rev. 21:55–89; doi:10.1210/edrv.21.1.0389.

Shankardass K, McConnell R, Jerrett M, Milam J, Richardson J, Berhane K. 2009. Parental stress increases the effect of traffic-related air pollution on childhood asthma incidence. Proc. Natl. Acad. Sci. 106:12406–12411; doi:10.1073/pnas.0812910106.

Smith AH, Marshall G, Yuan Y, Liaw J, Ferreccio C, Steinmaus C. 2010. Evidence From Chile That Arsenic in Drinking Water May Increase Mortality From Pulmonary Tuberculosis. Am. J. Epidemiol. kwq383; doi:10.1093/aje/kwq383.

US EPA. 2003. Framework for Cumulative Risk Assessment.

US EPA ERC. 2007 Annual Report | Development of Receptor- to Population-Level Analytical Tools for Assessing Endocrine Disruptor Exposure in Wastewater-Impacted Estuarine Systems. Available: http://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.highlight/abstract/7889/report/2007 [accessed 30 July 2015].

Vishnevetsky J, Tang D, Chang H-W, Roen EL, Wang Y, Rauh V, Wang S, Miller RL, Herbstman J, Perera FP. 2015. Combined effects of prenatal polycyclic aromatic hydrocarbons and material hardship on child IQ. Neurotoxicol. Teratol. 49:74–80; doi:10.1016/j.ntt.2015.04.002.

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, et al. 2011a. Metabolite profiles and the risk of developing diabetes. Nat. Med. 17:448–453; doi:10.1038/nm.2307.

Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, et al. 2011b. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472:57–63; doi:10.1038/nature09922.

Wild CP. 2005. Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. Cancer Epidemiol. Biomarkers Prev. 14:1847–1850; doi:10.1158/1055-9965.EPI-05-0456.

Winans B, Humble MC, Lawrence BP. 2011. Environmental toxicants and the developing immune system: a missing link in the global battle against infectious disease? Reprod. Toxicol. Elmsford N 31:327–336; doi:10.1016/j.reprotox.2010.09.004.

Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung’u T, Ruxrungtham K, Sanchez J, Brander C, Chung RT, et al. 2015. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348:aaa0698; doi:10.1126/science.aaa0698.

Yang C, Li C, Wang Q, Chung D, Zhao H. 2015. Implications of pleiotropy: challenges and opportunities for mining Big Data in biomedicine. Front. Genet. 6:229; doi:10.3389/fgene.2015.00229.

Zota AR, Shenassa ED, Morello-Frosch R. 2013. Allostatic load amplifies the effect of blood lead levels on elevated blood pressure among middle-aged U.S. adults: a cross-sectional study. Environ. Health Glob. Access Sci. Source 12:64; doi:10.1186/1476-069X-12-64.

Chapter 2: Elevated Levels of Organochlorine Pesticides in Asian Indians May Partially Explain their Increased Risk of Type II Diabetes

Sarah I. Daniels[1], John Chambers[2], Sylvia Sanchez[1], Michele A. LaMerrill[3], Alan E. Hubbard[1], Anthony Macherone[4], Matthew McMullan[5], Luoping Zhang[1], Paul Elliott[2], Jaspal Kooner[2], Martyn T. Smith[1]

(Authorship order to be determined)
Associated Institutions
[1]Division of Environmental Health Sciences, School of Public Health, University of California,, Berkeley, California
[2]Department of Epidemiology and Biostatistics, Imperial College London, UK; MRC-PHE Centre for Environment and Health, Imperial College London, London, UK; Ealing Hospital NHS Trust, Middlesex, UK; Imperial College Healthcare NHS Trust, London, UK
[3]Department of Environmental Toxicology, University of California, Davis, California
[4]Agilent Technologies, Inc., Santa Clara, CA 95051
[5]NMS Labs, Willow Grove, PA 19090

ABSTRACT
Context: Rates of type II diabetes (T2D) have increased dramatically throughout the world in the past 30 years, yet there is a disproportionately higher prevalence of T2D in Asian Indians. Unique heritable traits, dietary choices, and behavioral changes in this population only partially contribute to their increased risk of T2D. Environmental exposures have not been studied in Asian Indians as a plausible explanation despite growing evidence of an association between chemical exposures, particularly organochlorine pesticides (OCPs), and T2D.
Hypothesis: Levels of OCPs detected in blood are higher in Asian Indian immigrants than European whites residing in Greater London, UK, and these increased OCP concentrations contribute to elevated T2D risk.
Design: A nested case-control study of Asian Indians and European whites enrolled in the London Life Sciences Population (LOLIPOP) Study cohort.
Setting and Participants: Tamils mostly from Sri Lanka, Telugus originally from Southern India, and European whites from England, were recruited for blood collection at Ealing Hospital in 2012. Biometric, clinical, and survey data was also collected.
Main Outcome Measure: Subjects with fasting-blood glucose ≥126 mg/dL (≥7 mmol/L) were considered diabetic.
Results: Tamils had approximately 3-9-fold higher levels of OCPs and Telugus had 9-30-fold higher levels as compared to European whites. Odds of exposure to higher levels of *p,p'*-Dichlorodiphenyldichloroethylene (*p,p'*-DDE) was significantly greater in Asian Indians with T2D than controls, OR= 7.00 (2.22, 22.06 95% CI). Similarly, odds of higher β-hexachlorhexane (β-HCH) levels were significantly greater in the Tamils with T2D than controls, OR= 9.35 (2.43, 35.97 95%CI).
Conclusions: Asian Indians have a significantly higher body burden of OCPs than European whites and high *p,p'*-DDE and β-HCH concentrations are associated with T2D in this population.

Introduction

There are almost 70 million adults living with diabetes in India (9-10% prevalence), and approximately 90% of these cases are type II diabetes (T2D) [1]. Rates of T2D are rising in Asian Indian diaspora populations too, including those in the UK [2–4]. As compared to European white populations, Asian Indians have 2-3-fold higher rates of T2D [2–4]. Diabetes develops in Asian Indians at relatively lower body weight, blood lipid level, and age compared to other racial groups, yet known risk factors cannot fully explain these differences in risk. While there is a proposed familial inheritance of glucose dysregulation and insulin resistance in Asian Indians [5], little T2D risk has been attributed to genetic polymorphisms in Indians native to India and within the diaspora population in the UK [6,7]. Furthermore, no major differences have been found in SNPs associated with T2D in Asian Indians versus Europeans [8]. Other non-genetic factors, such as effects of rapid urbanization, malnutrition *in utero*, and psychosocial stress, have been considered as partially responsible for the accelerated incidence of T2D in Indians [9]. Environmental exposures, such as a number of persistent organic pollutants (POPs), have been associated with T2D [10], but have not been explored as risk factors of T2D in Asian Indians and may shed light on the high uncharacterized susceptibility of this population.

The current case-control study examines differences in exposure to persistent organic pollutants (POPs) within diabetic and healthy Asian Indian immigrants as compared to European whites. The London Life Sciences Prospective Population Study (abbreviated as "LOLIPOP"), a cohort that includes Asian Indian immigrants and European whites residing in West London, is used as our study population. Several genetic and epigenetic studies of T2D susceptibility have been performed previously on the LOLIPOP cohort [6,11,12], however, environmental exposures associated with T2D have not yet been examined in this population. Given the unusually high risk to insulin resistance in Asian Indian immigrants in the UK and the potentially high level of exposure to POPs while living in India, investigating the relationship between POPs blood levels and T2D is warranted.

We describe a novel method and strategy to assess exposure to specific POPs of interest within Asian Indian immigrants that may be related to T2D. As numerous POPs have been associated with insulin resistance and T2D in other populations [10], it is difficult to determine *a priori* which compounds are of greatest importance to measure with respect to T2D risk in a given sub-group. Therefore, we first analyzed 66 POPs, representative of six different chemical classes, using small volumes of human plasma samples, to broaden our search for environmental exposures associated with T2D in Asian Indians. This semi-targeted method can help to define a population's "exposure fingerprint" for tens to hundreds of analytes and help choose the best candidates for further analysis. In our larger follow-up study, we made comparisons of candidate POPs between Asian Indian immigrants and European whites with- and without- clinically diagnosed T2D. We observed higher exposure to organochlorine pesticides (OCPs) in Asian Indian immigrants compared to European whites and found strong associations between T2D and OCP concentrations in Asian Indians.

Materials and Methods

Human Subjects

The LOLIPOP study was established in 2002 and is a prospective cohort comprised of >24,000 Asian Indians and European whites living in West London. Details of the study are outlined previously [13]. Adult (>21 years) volunteers, of mostly Telugu or Sri Lankan Tamil descent, were newly recruited for the LOLIPOP study in 2012, per request by Corriell's 1000 Genome Project [http://www.1000genomes.org/]. Tamils immigrated an average of 20 years ago, while Telugus immigrated an average of 12 years ago. For purposes of comparison, we also obtained blood samples from European white volunteers of the LOLIPOP study (n= 6 cases and n= 72 controls). Diabetic cases were defined as subjects with fasting-blood glucose $\geq$7 mmol/L at the time of blood collection. In our initial pilot study, 66 POPs were screened in plasma samples from 24 Asian Indian T2D cases and 25 healthy controls, frequency matched on age, sex, proportion of Telugus, smoking status, and WHR (Table S1). Differences observed between cases and controls and between ethnic groups were followed-up in a larger study that included; the pilot samples (N=49), more Tamil and Telugu controls (N=71), European white controls (N=72), and European white cases (N=6). There were a limited number of T2D cases in this study due to the cross-sectional nature of our original sampling method. Therefore, the number of healthy controls was increased to maximize the statistical power to detect differences between the diseased and healthy sub-populations. The healthy controls in the follow-up study were frequency-matched for age and sex, both between the European whites and Asian Indians as well as between the Telugus and Tamils in order to make comparisons in POPs concentrations between these different groups.

Mass Spectrometry for Persistent Organic Pollutants

A targeted method was developed for the quantitative analysis of 66 POPs (analytes) in 200 µL of plasma. Traditional methods for POPs measurements require larger volumes (5-10 mL) of plasma [14]. However using a sensitive quadrupole GC-MS/MS system we reduced these volume requirements and still obtained comparable detection limits. In the pilot, 66 POPs (15 polycyclic aromatic hydrocarbons, 12 dioxin-like polychlorinated biphenyls, 11 polybrominated diphenylethers, 18 organochlorine pesticides, 5 dioxins and 5 furans) were measured in plasma samples from 49 subjects. For the follow-up study, a subset of DDT compounds, HCH compounds, and PCBs were measured. This reduced list included: α-hexachlorohexane (α-HCH), β-HCH, γ-HCH, *o,p'*- Dichlorodiphenyldichloroethane (DDD), *o,p'*-Dichlorodiphenyldichloroethylene (DDE), *o,p'*- dichlorodiphenyltrichloroethane (DDT), *p,p'*-DDD, *p,p'*-DDE, *p,p'*-DDT, polychlorinated biphenyl-105 (PCB-105), PCB-114, PCB-118, PCB-123, PCB-156, PCB-157, PCB-167 and, PCB-189 (AccuStandard, Inc., New Haven, CT). Plasma was extracted in four batches of 50 samples each, using chemical denaturation, liquid-liquid extraction, solid-phase cleanup and reconstitution with hexanes. To account for sample variability, a pooled reference sample was extracted in each batch, alongside the other subject samples, and measured at regular time points throughout the analyses. An Agilent (Santa Clara, CA) 7890B GC coupled to an Agilent 7000C GC-triple quadrupole mass spectrometer operated in electron impact (EI) multiple reaction monitoring (MRM) mode, was used for the analyses. System performance and precision were monitored with calibrators at 7 concentration levels

for each analyte. Further details regarding sample preparation, extraction, and measurement by MS/MS for the pilot and follow-up are described in Supplementary Methods.

Quantitative analysis was performed using Agilent MassHunter software and calibration curves were constructed using the relative analyte response (calculated area of analyte / calculated area of internal standard). These normalized peak-areas were converted to concentration (ng/mL) using the known concentrations of the standard curve calibrators from each extraction batch. Concentrations (ng/mL) were converted to lipid-adjusted values by calculating total lipids using the available clinical lipid profile measurements and the following formula: Total lipids = (2.27 × total cholesterol) + triglyceride + 0.623 [15]. The MDL/ √2 value was substituted when the analyte concentration was <MDL, [16].

Statistical Analysis

To assess the association of individual POPs with T2D, exposure status was divided into two groups based on the 50^{th}-percentile of each POP concentration for the Asian Indian subjects (N=120) and white subjects (N=78). Logistic regression models were performed to obtain unadjusted and adjusted odds ratios of T2D given exposure above the 50^{th} percentile for each POP. At times this was not possible because all cases of T2D fell above the 50^{th} percentile. For analyte concentrations in ng/mL units, triglycerides and cholesterol were included as covariates in the regression model. (Other adjustments considered in multivariable models include the variables for age, WHR, SBP, sex, smoking status, and alcohol use. Given the small sample size of cases, these were not reported.) P-values for the 2x2 contingency tables of exposed versus diseased were obtained using Fisher's exact t-test. When appropriate, this same analysis was completed for Telugu (N=47) and Tamil (N=73) subsets.

Within the whole sample population, POPs levels were correlated with each other. Spearman correlation coefficients were calculated for each pair of POPs using the entire N=198. All statistical analyses were performed using R software.

Results

We used a targeted approach to screen 66 POPs from five chemical classes in a pilot study of Asian Indian immigrants (N=49). A total of 27 of the 66 POPs were detected in the plasma of at least 10 subjects in the pilot, with limits of detection as low as 0.1 ng/mL (Supplementary Table S2). We initially observed higher *p,p'*-DDE levels in T2D cases versus controls and higher β-HCH levels in Telugus versus Tamils. PCB concentrations were relatively constant across comparison groups, and were therefore designated as the negative control analyte (not associated with case versus control status). A follow-up study in a larger sample size was conducted on three chemical classes (DDTs, HCHs and PCBs).

A total of 198 Asian Indian immigrant and European whites were selected for follow-up from 375 of the blood samples collected in this cross-sectional study. This included repeated measures from the original pilot, additional Asian Indian control subjects (both Tamils and Telugus), and European white subjects as a comparison group. The baseline characteristics for these groups are described in Table 1, subdivided into T2D cases and controls. Of note, 79% of the Asian Indians and 65% of the whites had "abdominal obesity", defined by ethnicity-specific cut-offs of waist circumference [17]. Age, WHR, triglycerides, cholesterol, LDLs and systolic

blood pressure (SBP) were higher in the Indian T2D cases versus the healthy controls (with similar trends seen for the European whites). These differences in physical characteristics were considered as covariates in downstream analyses. (The final adjusted models did not include SBP as this variable was not a significant covariate and did not greatly change the OR for individual POPs exposures.)

The POPs with the highest detected concentrations from each chemical class are reported here. The LODs, LOQs, and CVs (based on a pooled reference sample) for *p,p'*-DDT, *p,p'*-DDE, β-HCH, and PCB-118 were calculated, as well as the median and ranges of concentrations for each analyte within the entire sample population (in ng/mL) (Table S3). Differences in POP concentrations were observed between healthy Asian Indian immigrants versus European whites (Supplementary Figure S1). The median concentrations of *p,p'*-DDE and *p,p'*-DDT in control subjects were over 8-fold higher in healthy Asian Indians (median; 535.87 and 17.65 ng/g-lipid for *p,p'*-DDE and *p,p'*-DDT, respectively) than in healthy whites (median; 61.26 and 2.08 ng/g-lipid for *p,p'*-DDE and *p,p'*-DDT, respectively) while PCB-118 concentrations were similar in the two groups (median; 4.51 for Asian Indians and 3.94 ng/g-lipid for whites). β-HCH concentrations in healthy controls were 3-fold higher in Tamil control subjects (36.73 ng/g-lipid) and 30-fold higher in Telugus (365.32 ng/g-lipid) than in white control subjects (12.86 ng/g-lipid). Similar fold-change differences were also found in ng/mL units..

Differences in POPs concentrations were also observed within the Asian Indian sub-populations. β-HCH was higher in Telugus than Tamils in the pilot study, and was confirmed follow-up study, with approximately 10-fold higher levels of β-HCH observed in Telugu than Tamils (365.32 and 36.73 ng/g-lipid, respectively). The *p,p'*-DDT, *p,p'*-DDE, and PCB-118 concentrations were more similar between Tamils and Telugus (Supplementary Figure S1 and Supplementary Table S4a and S4b). DDT concentrations were 2-fold higher in Telugus than Tamils (median; 27.01 and 13.00 ng/g-lipid, respectively), and the PCB-118 concentrations were 2-fold higher in Tamils than Telugus (6.06 and 3.23 ng/g-lipid, respectively).

The associations between the exposure to individual POPs and T2D status are also reported. Within each group strata (i.e., Asian Indians or whites), exposure to each POP analyte was dichotomized (above versus below median value), and logistic regression was performed to estimate the crude or adjusted ORs for exposure to POPs in T2D subjects versus controls (Table 2a and 2b). In Asian Indians, we observed an increased odds of *p,p'*-DDE serum concentrations above the 50^{th} percentile in T2D subjects, OR= 7.00 (2.22, 22.06 95% CI). In whites, all T2D cases had *p,p'*-DDE concentrations above the 50^{th} percentile in ng/mL units (Supplementary Figure S2). As for PCB-118, the OR of high exposure in Asian Indian T2D cases versus controls was only significant in the unadjusted model OR = 2.99 (1.13, 7.88).

We assessed associations between each POP analyte and T2D after stratifying the Asian Indians into Tamil and Telugu ethnic groups. An increased odds, OR= 9.35 (2.43, 35.97 95% CI), of β-HCH concentrations above the 50^{th} percentile was observed in Tamils with T2D compared to healthy subjects (Table 3a) and similar a OR was observed when using ng/mL units (Table 3b). Five of the Tamils were from mainland India (and had relatively higher concentrations of β-HCH compared to other Tamils). Upon removal of these Tamils and repeating the analysis, the odds of relatively high exposure to β-HCH in Tamils with T2D increased dramatically compared to controls, OR_{adj}= 22.94 (3.0, 175.33). Odds ratios could not be determined for the Telugu

population because all T2D cases were in the high-exposure group (this was also true for whites in the lipid-unadjusted ng/mL units) (Supplementary Figure S2). When modeling the relationship between p,p'-DDT, p,p'-DDE or PCB-118 and T2D in Tamils or Telugus alone, no significant OR was found after adjusting for covariates in the model (Supplementary Table S4a and S4b).

Some of the POPs in this study were highly correlated with each other. Levels of *p,p'*-DDT and *p,p'*-DDE were highly correlated ($r=0.77$, $p<0.001$); the β-HCH levels were correlated with the DDT compounds ($r=0.83$, $p<0.001$ for DDT, $r=0.72$, $p<0.001$ DDE); and yet PCB-118 was poorly correlated with the OCPs (Supplementary Figure S3).

Discussion

Methods were implemented to screen 66 POPs and characterize the "exposure fingerprint" of Asian Indian immigrants in a pilot study. The distributions of POPs levels in T2D cases versus controls and among Asian Indian immigrant ethnic groups could not be predicted *a priori* because studies had not previously been conducted on this population. Based on the pilot results, we were able to power our follow-up study to examine differences in POPs of interest between cases versus controls and among Tamils, Telugus and whites in a larger sample population.

In the follow-up study, we observed Asian Indian immigrants in London had much higher concentrations of OCPs than their European white counterparts. DDT and DDE concentrations were over 8-fold higher in Asian Indians than European whites living in West London. Additionally, β-HCH concentrations were 3-fold higher in Tamils and up to 30-fold higher in Telugus compared to European whites. Concentrations observed in blood samples of Asian Indians immigrants collected from this study are lower than previously reported in India [18], yet still higher in comparison to European white populations in the UK [19]. This is the first study to show differences in exposure levels are sustained even 10-20 years after immigration to a new, relatively low-exposure environment.

Asian Indians have been exposed to OCPs for longer periods and at higher concentrations than other populations in Western Europe, where these legacy compounds were largely phased out in the 1970s and 80s. Unregulated DDT spraying occurred throughout India for agricultural purposes and control of mosquito-borne diseases [18,20] until India ratified the Stockholm Convention in 2006. Still, India has the highest use of DDT in the world [21]. Even though the predicted half-life of DDE in human blood is approx. 6-7 years, while the half-life of the parent compound, DDT, is 2 years [22], both compounds were above the LOD for almost all of the subject samples tested in this study. Until 2012 another pesticide, Lindane (gamma-HCH), was heavily applied as a pesticide alternative to DDT to control for locusts in India [18,20]. Lindane is fairly short-lived in the environment [20], but the β-isomer (β-HCH), an impurity formed during production of Lindane, has a half-life of 7 years in humans [23]. Lindane was only detected in 75% of our subject samples and the signal was 32-fold lower than β-HCH in the MS/MS data. Moreover, the β-HCH levels varied greatly, depending on country of origin (i.e., Sri Lanka versus India).

Within Asian Indian ethnic groups Telugus had 2-fold higher levels of *p,p'*-DDT and a 8-10-fold higher levels of β-HCH than Tamils. This higher concentration of *p,p'*-DDT in Telugus may

be due to their more recent exposure, as this sub-group immigrated to the UK 10 years after Tamils (on average). The higher concentration of β-HCH in Telugus is likely due to the more prevalent use of Lindane in India compared to Sri Lanka. Blood POP levels are known to vary considerably among individuals living in different regions of India [18], however no studies have made comparisons between Indian (Telugu) and Sri Lankan (Tamil) immigrants.

Significant associations between T2D and p,p'-DDE and β-HCH exposures were found in Asian Indians. The measures of effect in this case-control study, OR= 7.00 (2.22, 22.06 95% CI) for p,p'-DDE in Asian Indians and OR= 9.35 (2.43, 35.97 95% CI) for β-HCH in Tamils, are even higher than some previously reported in different populations from around the world [10]. Smaller associations between T2D and *p,p'*-DDE have been observed in cross-sectional studies in Swedes [24,25], American Indians [26], Americans [27–31], Koreans [32], Spanish [33] and Slovakians [34] as well as in longitudinal studies in Americans and Swedes [35,36]. A link between β-HCH and pre-existing T2D has also been reported in cross-sectional studies conducted on Mexican-Americans [29], the entire 1999-2004 NHANES sample population [28], Slovakians [34], Koreans [32], Spanish [33], Saudi Arabians [37], and Norwegians [38]. Insulin resistance has been associated with OCP exposure in humans as well [39]. Given the high levels of OCPs found in Indian immigrants, future prospective studies confirming the relationship between T2D and OCPs in this population may help explain Asian Indians' greater susceptibility to T2D.

Animal and tissue-culture models provide additional evidence of mechanisms for glucose dysregulation and reduced insulin sensitivity from OCP exposure. Associations between DDT exposure and blood glucose levels were initially found in rats [40] and mice [41] over forty years ago. More recent studies in mice have shown acute exposure to DDE increases fasting-blood glucose levels and body weight for 7-21 days post-treatment [42]. Another study in female mice exposed to DDT on a high-fat diet showed increased fat mass and insulin resistance, as well as reduced thermogenesis [43]. Further, pancreatic beta-cells chronically-exposed to *p,p'*-DDT or *p,p'*-DDE *in vitro* decreased expression of proteins involved with a hyperglycemia stress response [44]. Additionally, glucose dysregulation has been observed under acute treatment of Lindane in animal and cell models, yet the opposite effects were seen *in vitro* vs. *in vivo* [45]. Individual OCPs have been shown to affect glucose metabolism in experimental models, yet the combined effect of mixtures on metabolic changes needs further elucidation. OCPs were highly correlated in this sample population as expected from other studies [34].

This is the first large study comparing POP levels of Asian Indians to European whites residing in a Western city, representing both current and prior exposure. This study also uses a novel method of measuring representative POPs in relatively small volumes of plasma. We report significant associations of T2D with OCPs using both lipid-adjusted (ng/g-lipid) and unadjusted (ng/mL) units. Even with small sample sizes of T2D cases, relatively higher concentrations of *p,p'*-DDE and β-HCH in diabetics versus healthy control subjects were observed in each subpopulation. This effect was not seen with PCB-118 after adjustment for covariates (data not shown), which had similar concentrations across all Asian Indians and whites, and was used as a negative control analyte. Our finding exemplifies the importance of using a semi-targeted exposomic approach to first screen for plausible chemical exposures

associated with T2D in environmentally-exposed populations, and then follow-up on these findings with a larger sample size.

Yet there were still several limitations to this study. Firstly, the small number of cases did not allow for rigorous statistical analysis in the Telugu and white sub-groups. However, the observed trends suggest that diabetics have higher levels of OCPs in T2D cases versus controls. High correlation was observed with several OCPs, and so these compounds may not contribute independently to T2D risk. While there is both human and experimental evidence that only single classes of POPs, including PCBs, organochlorine pesticides, and PBDEs, are associated with T2D [10], *in situ* models suggest that all POPs could also act through similar mechanisms [46]. Moreover, the risk of T2D may be more dependent on the timing and dose of cumulative POPs levels (particularly OCPs) as opposed to current measurements of single analytes or chemical classes. In the future, cohort studies on banked blood samples from the LOLIPOP and other studies, such as the Mediators of Atherosclerosis in South Asians Living in America (MASALA) cohort of Indian migrants to the US, could be used to demonstrate exposure-disease temporality.

Despite these limitations, this study adds to the growing literature of positive epidemiological associations between OCPs and T2D. There have been few biomonitoring studies of OCPs in Asian Indian immigrants to date, and to our knowledge, no studies examining the relationship between their high rates of T2D and OCP exposure. Future prospective studies on OCPs in Indians should focus on native and immigrant Asian Indians, who historically had high and low exposures to multiple pesticides. Asian Indians comprise a substantial proportion of the world population, thus, confirmation of the associations we found between OCPs and T2D could have public health implications on a global scale.

Tables

Table 1. Sample Population Characteristics

	Whites*					Asian Indians				
	Controls N= 72		Cases N=6			Controls N=96		Cases N=24		
					p-value					p-value
Males (%)	36 (50)		6 (100)		0.03	53 (55.2)		17 (70.8)		0.25
Smoke (%)	12 (16.7)		0		0.58	4 (4.17)		3 (12.5)		0.14
Drink (%)	43 (59.7)		3 (50)		0.69	25 (26.0)		5 (20.8)		0.79
	Mean	**SD**	**Mean**	**SD**	**p-value**	**Mean**	**SD**	**Mean**	**SD**	**p-value**
Age (years)	48.49	6.65	62.15	9.64	**<0.01**	48.32	8.38	56.10	9.57	**<0.01**
SBP (mmHg)	122.00	15.10	138.30	7.92	**0.03**	123.90	13.50	131.75	15.15	**0.02**
DBP (mmHg)	76.70	10.62	84.60	8.73	0.07	78.56	9.13	81.17	7.79	0.15
BMI (kg/m^2)	26.43	4.55	29.25	5.29	0.13	26.52	3.37	26.14	3.53	0.63
WHR	0.89	0.08	1.02	0.05	**<0.01**	0.93	0.07	0.97	0.07	**0.01**
HDL (mmol/L)	1.54	0.46	1.24	0.34	0.10	1.32	0.32	1.23	0.36	0.28
LDL (mmol/L)	3.02	0.84	2.59	0.63	0.19	3.36	0.84	2.16	1.08	**<0.01**
Glucose (mmol/L)	4.97	0.34	7.68	0.51	**<0.01**	4.83	0.33	9.02	1.63	**<0.01**
HbA1c (%)	5.41	0.30	7.62	0.99	**<0.01**	5.62	0.40	8.20	1.29	**<0.01**
Chol (mg/dl)	195.30	34.20	188.50	26.20	0.69	204.22	35.35	165.82	50.94	**<0.01**
Trig (mg/dl)	99.40	59.90	199.10	91.40	**<0.01**	120.31	56.42	167.78	83.12	**0.01**
Years lived in the UK*						16.33	11.28	23.955	23.96	12.06
Body fat (%)	30.88	8.70	32.43	9.76	0.68	32.34	7.75	29.61	7.89	0.14

*All white control subjects were born in the UK. Three white T2D cases were born outside of the UK.
(One subject immigrated as an infant, and two subjects immigrated when they were adults.)
P-value differences for whites were determined by a permutation exact test.

Table 2a. Exposure characteristics for diabetics versus controls within ethnicity groups (ng/g-lipid)

			Controls		Cases			
Compound (ng/g-lipid)	**Population**	**Exposure Status**	**N**	**Median (Range)**	**N**	**Median (Range)**	**Odds Ratios (95% CI)***	***p*-value**
p,p'-DDE	White	<65.12	38	39.97 (17.65,63.37)	0	N/A	(Inf)	**0.03**
		>65.12	33	93.93 (65.12,353.30)	6	149.91 (116.92,421.03)		
	Asian Indians	>710.87	56	318.00 (26.82, 705.10)	4	208.34 (141.38,552.80)	**7.00 (2.22, 22.06)**	**<0.001**
		<710.87	40	1282.48 (736.62, 25143.80)	20	1698.55 (716.627,6212.58)		
p,p'-DDT	White	<2.14	38	1.48 (0.64,2.14)	1	1.73	5.59 (0.62, 50.25)	0.20
		>2.14	34	3.33 (2.15, 70.97)	5	4.55 (3.06, 5.92)		
	Asian Indians	<17.61	47	11.12 (3.91,17.57)	13	10.03 (6.24, 16.05)	0.8 (0.33, 1.99)	0.82
		>17.61	49	30.91 (17.65, 316.50	11	28.91 (17.66, 194.90)		
PCB-118	White	<4.56	36	3.17 (0.89,4.52)	2	4.37 (4.30, 4.45)	5.61 (0.62, 50.48)	0.67
		>4.56	34	6.45 (4.60,13.38)	4	7.36 (5.92,22.93)		
	Asian Indians	<4.36	53	2.66 (0.81,4.33)	7	2.76 (2.03, 4.22)	**2.99 (1.13, 7.88)**	**0.04**
		>4.36	43	7.33 (4.40, 34.21)	17	6.32 (4.52,27.34)		

*Adjustment for age, WHR, sex, smoking status and alcohol use did not change the effect size or significance levels except for PCB-118 for Asian Indians, OR_{adj}=2.56 (0.80, 8.16). (The adjusted ORs are likely inaccurate estimates because of the small number of cases and are not reported in the table.)

Table 2b. Exposure characteristics for diabetics versus controls within ethnicity groups (ng/mL)

			Controls		Cases			
Compound (ng/mL)	**Population**	**Exposure Status**	**N**	**Median (Range)**	**N**	**Median (Range)**	**Odds Ratios (95% CI)***	***p*-value**
p,p'-DDE	White	<0.38	38	0.24 (0.08,0.37)	0	N/A	(Inf)	**0.03**
		>0.38	33	0.59 (0.38,2.18)	6	1.07 (0.71,3.04)		
	Asian Indians	<3.82	54	1.86 (0.16,3.76)	6	2.49 (1.18,3.79)	**5.01 (1.40, 17.99)**	**0.01**
		>3.82	42	7.54 (3.84,145.85)	18	12.11 (3.92,52.99)		
p,p'-DDT	White	<0.01	39	0.01 (0.004,0.1)	0	N/A	(Inf)	**0.03**
		>0.01	33	0.02 (0.01,0.32)	6	0.03 (0.01,0.04)		
	Asian Indians	<0.11	47	0.07 (0.02,0.11)	13	0.06 (0.03, 0.10)	0.94 (0.29, 3.08)	0.82
		>0.11	49	0.19 (0.11,2.17)	11	0.24 (0.11, 1.34)		
PCB-118	White	<0.03	37	0.02 (0.006,0.02)	1	0.02	5.87 (0.45, 76,63)	0.20
		>0.03	33	0.04 (0.03,0.10)	5	0.05 (0.03,0.17)		
	Asian Indians	<0.03	49	0.02 (0.006,0.02)	11	0.02 (0.01,0.02)	1.60 (0.52, 4.97)	0.82
		>0.03	47	0.05 (0.03,0.24)	13	0.05 (0.03,0.12)		

*Only adjusted for triglycerides and cholesterol levels. Further adjustment for age, WHR, sex, smoking status and alcohol use did not widely change the effect size nor significance levels. (The adjusted ORs are likely inaccurate estimates because of the small number of cases and are not reported in the table.)

Table 3a. Exposure characteristics for β-HCH diabetics versus controls within ethnicity groups (ng/g-lipid)

			Controls		Cases			
Compound (ng/g-lipid)	**Population**	**Exposure Status**	**N**	**Median (Range)**	**N**	**Median (Range)**	**Odds Ratios (95% CI)**	***p*-value**
β-HCH	White	<12.98	37	8.82 (3.18,12.97)	1	12.76	5.61 (0.62, 50.48)	0.20
		>12.98	33	17.6 (13.00,36.44)	5	33.63 (18.91,60.01)		
	Tamil	<50.58	33	27.12 (4.63,48.98)	3	49.30 (35.61,49.89)	**9.35 (2.43, 35.97)**	**<0.001**
		>50.58	20	84.61 (50.58,541.70)	17	95.35 (52.03,499.20)		
	Telugu	<369.30	23	272.81 (96.86,365.42)	0	N/A	(Inf)	0.11
		>369.30	20	461.41 (369.34,714.45)	4	535.66 (374.28,627.60)		

*Further adjustment for age, WHR, sex, smoking status and alcohol use did not widely change the effect size nor significance levels except for Tamils, OR_{adj}=7.01 (1.44, 34.0). (The adjusted ORs are likely inaccurate estimates because of the small number of cases in this study and are not reported in the table.)

Table 3b. Exposure characteristics for β-HCH diabetics versus controls within ethnicity groups (ng/mL)

			Controls		Cases			
Compound (ng/mL)	**Population**	**Exposure Status**	**N**	**Median (Range)**	**N**	**Median (Range)**	**Odds Ratios (95% CI)***	***p*-value**
β-HCH	White	<0.08	38	0.05 (0.02,0.07)	0	N/A	(Inf)	**0.03**
		>0.08	32	0.12 (0.08,0.26)	6	0.18 (0.09,0.43)		
	Tamil	<0.31	31	0.16 (0.03,0.30)	5	0.26 (0.17,0.31)	**8.46 (1.65, 43.48)**	**0.02**
		>0.31	22	0.51 (0.31,3.04)	15	0.50 (0.32,2.11)		
	Telugu	<2.56	23	1.78 (0.56,2.50)	0	N/A	(Inf)	0.11
		>2.56	20	3.13 (2.56,4.50)	4	3.83 (3.15,4.10)		

Only adjusted for triglycerides and cholesterol levels. Further adjustment for age, WHR, sex, smoking status and alcohol changed the effect size value and the significance level for Tamils, OR_{adj}=5.72 (0.72, 45.70). (The adjusted ORs are likely inaccurate estimates because of the small number of cases and are not reported in the table.)*

References

1. Cavan D, de Rocha Fernandes J, Makaroff L, Ogurtsova K, Webber S. IDF Diabetes Atlas, 7th Edition [Internet]. International Diabetes Federation; 2015. Available from: http://www.diabetesatlas.org/resources/2015-atlas.html
2. Barnett AH, Dixon AN, Bellary S, Hanif MW, O'hare JP, Raymond NT, Kumar S. Type 2 diabetes and cardiovascular risk in the UK south Asian community. Diabetologia. 2006 Oct;49(10):2234–2246. PMID: 16847701
3. Tillin T, Hughes AD, Godsland IF, Whincup P, Forouhi NG, Welsh P, Sattar N, McKeigue PM, Chaturvedi N. Insulin resistance and truncal obesity as important determinants of the greater incidence of diabetes in Indian Asians and African Caribbeans compared with Europeans: the Southall And Brent REvisited (SABRE) cohort. Diabetes Care. 2013 Feb;36(2):383–393. PMCID: PMC3554271
4. Gujral UP, Pradeepa R, Weber MB, Narayan KV, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. Ann N Y Acad Sci. 2013 Apr;1281(1):51–63. PMCID: PMC3715105
5. Zabaneh D, Chambers JC, Elliott P, Scott J, Balding DJ, Kooner JS. Heritability and genetic correlations of insulin resistance and component phenotypes in Asian Indian families using a multivariate analysis. Diabetologia. 2009 Sep 10;52(12):2585–2589.
6. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, Been LF, Chia K-S, Dimas AS, Hassanali N, Jafar T, Jowett JBM, Li X, Radha V, Rees SD, Takeuchi F, Young R, Aung T, Basit A, Chidambaram M, Das D, Grundberg E, Hedman AK, Hydrie ZI, Islam M, Khor C-C, Kowlessur S, Kristensen MM, Liju S, Lim W-Y, Matthews DR, Liu J, Morris AP, Nica AC, Pinidiyapathirage JM, Prokopenko I, Rasheed A, Samuel M, Shah N, Shera AS, Small KS, Suo C, Wickremasinghe AR, Wong TY, Yang M, Zhang F, DIAGRAM, MuTHER, Abecasis GR, Barnett AH, Caulfield M, Deloukas P, Frayling TM, Froguel P, Kato N, Katulanda P, Kelly MA, Liang J, Mohan V, Sanghera DK, Scott J, Seielstad M, Zimmet PZ, Elliott P, Teo YY, McCarthy MI, Danesh J, Tai ES, Chambers JC. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011 Oct;43(10):984–989. PMCID: PMC3773920
7. Holliday EG. Hints of Unique Genetic Effects for Type 2 Diabetes in India. Diabetes. 2013 May 1;62(5):1369–1370. PMID: 23613552
8. Hassanali N, Silva NMGD, Robertson N, Rayner NW, Barrett A, Bennett AJ, Groves CJ, Matthews DR, Katulanda P, Frayling TM, McCarthy MI. Evaluation of Common Type 2 Diabetes Risk Variants in a South Asian Population of Sri Lankan Descent. PLOS ONE. 2014 Jun 13;9(6):e98608.
9. Pandit K, Goswami S, Ghosh S, Mukhopadhyay P, Chowdhury S. Metabolic syndrome in South Asians. Indian J Endocrinol Metab. 2012;16(1):44–55. PMCID: PMC3263197
10. Taylor KW, Novak RF, Anderson HA, Birnbaum LS, Blystone C, DeVito M, Jacobs D, Köhrle J, Lee D-H, Rylander L, Rignell-Hydbom A, Tornero-Velez R, Turyk ME, Boyles AL, Thayer KA, Lind L. Evaluation of the Association between Persistent Organic Pollutants (POPs) and Diabetes in Epidemiological Studies: A National Toxicology Program Workshop Review. Environ Health Perspect. 2013 May 7;121(7):774–783.

11. Li H, Kilpeläinen TO, Liu C, Zhu J, Liu Y, Hu C, Yang Z, Zhang W, Bao W, Cha S, Wu Y, Yang T, Sekine A, Choi BY, Yajnik CS, Zhou D, Takeuchi F, Yamamoto K, Chan JC, Mani KR, Been LF, Imamura M, Nakashima E, Lee N, Fujisawa T, Karasawa S, Wen W, Joglekar CV, Lu W, Chang Y, Xiang Y, Gao Y, Liu S, Song Y, Kwak SH, Shin HD, Park KS, Fall CHD, Kim JY, Sham PC, Lam KSL, Zheng W, Shu X, Deng H, Ikegami H, Krishnaveni GV, Sanghera DK, Chuang L, Liu L, Hu R, Kim Y, Daimon M, Hotta K, Jia W, Kooner JS, Chambers JC, Chandak GR, Ma RC, Maeda S, Dorajoo R, Yokota M, Takayanagi R, Kato N, Lin X, Loos RJF. Association of genetic variation in FTO with risk of obesity and type 2 diabetes with data from 96,551 East and South Asians. Diabetologia. 2012 Apr;55(4):981–995. PMCID: PMC3296006
12. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, Wahl S, Elliott HR, Rota F, Scott WR, Zhang W, Tan S-T, Campanella G, Chadeau-Hyam M, Yengo L, Richmond RC, Adamowicz-Brice M, Afzal U, Bozaoglu K, Mok ZY, Ng HK, Pattou F, Prokisch H, Rozario MA, Tarantini L, Abbott J, Ala-Korpela M, Albetti B, Ammerpohl O, Bertazzi PA, Blancher C, Caiazzo R, Danesh J, Gaunt TR, de Lusignan S, Gieger C, Illig T, Jha S, Jones S, Jowett J, Kangas AJ, Kasturiratne A, Kato N, Kotea N, Kowlessur S, Pitkäniemi J, Punjabi P, Saleheen D, Schafmayer C, Soininen P, Tai E-S, Thorand B, Tuomilehto J, Wickremasinghe AR, Kyrtopoulos SA, Aitman TJ, Herder C, Hampe J, Cauchi S, Relton CL, Froguel P, Soong R, Vineis P, Jarvelin M-R, Scott J, Grallert H, Bollati V, Elliott P, McCarthy MI, Kooner JS. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015 Jul;3(7):526–534. PMCID: PMC4724884
13. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, Balding D, Scott J, Kooner JS. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet. 2008 Jun;40(6):716–718. PMID: 18454146
14. Turner W. Laboratory Procedure Manual: PCBs and Persistent Pesticides [Internet]. National Center for Environmental Health, CDC; 2006 [cited 2016 Aug 25]. Available from: http://198.246.124.22/nchs/data/nhanes/nhanes_99_00/lab28poc_met_dioxin_pcb.pdf
15. Phillips DL, Pirkle JL, Burse VW, Bernert JT, Henderson LO, Needham LL. Chlorinated hydrocarbon levels in human serum: effects of fasting and feeding. Arch Environ Contam Toxicol. 1989 Aug;18(4):495–500. PMID: 2505694
16. Hornung RW, Reed LD. Estimation of Average Concentration in the Presence of Nondetectable Values. Appl Occup Environ Hyg. 1990 Jan 1;5(1):46–51.
17. International Diabetes Federation. IDF worldwide definition of the metabolic syndrome [Internet]. [cited 2016 Aug 2]. Available from: http://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf
18. Sharma BM, Bharat GK, Tayal S, Nizzetto L, Cupr P, Larssen T. Environment and human exposure to persistent organic pollutants (POPs) in India: a systematic review of recent and historical data. Environ Int. 2014 May;66:48–64. PMID: 24525153
19. Thomas GO, Wilkinson M, Hodson S, Jones KC. Organohalogen chemicals in human blood from the United Kingdom. Environ Pollut. 2006 May;141(1):30–41.

20. Sang S, Petrovic S, Cuddeford V. Lindane—a review of toxicity and environmental fate. World Wildl Fund Can [Internet]. 1999 [cited 2016 Jul 22];1724. Available from: http://www.pops.int/documents/meetings/poprc/request/Comments_2006/wwf/WWF%20canada.pdf
21. van den Berg H. Global Status of DDT and Its Alternatives for Use in Vector Control to Prevent Disease. Environ Health Perspect. 2009 Nov;117(11):1656–1663. PMCID: PMC2801202
22. Ritter R, Scheringer M, MacLeod M, Schenker U, Hungerbühler K. A Multi-Individual Pharmacokinetic Model Framework for Interpreting Time Trends of Persistent Chemicals in Human Populations: Application to a Postban Situation. Environ Health Perspect. 2009 Aug;117(8):1280–1286.
23. Jung D, Becher H, Edler L, Flesch-Janys D, Gurn P, Konietzko J, Manz A, Päpke O. Elimination of beta-hexachlorocyclohexane in occupationally exposed persons. J Toxicol Environ Health. 1997 May;51(1):23–34. PMID: 9169059
24. Rylander L, Rignell-Hydbom A, Hagmar L. A cross-sectional study of the association between persistent organochlorine pollutants and diabetes. Environ Health. 2005 Nov 29;4:28. PMCID: PMC1318465
25. Rignell-Hydbom A, Rylander L, Hagmar L. Exposure to persistent organochlorine pollutants and type 2 diabetes mellitus. Hum Exp Toxicol. 2007 May 1;26(5):447–452. PMID: 17623770
26. Codru N, Schymura MJ, Negoita S, Rej R, Carpenter DO. Diabetes in Relation to Serum Levels of Polychlorinated Biphenyls and Chlorinated Pesticides in Adult Native Americans. Environ Health Perspect. 2007 Oct;115(10):1442–1447. PMCID: PMC2022671
27. Everett CJ, Frithsen IL, Diaz VA, Koopman RJ, Simpson Jr. WM, Mainous III AG. Association of a polychlorinated dibenzo-p-dioxin, a polychlorinated biphenyl, and DDT with diabetes in the 1999–2002 National Health and Nutrition Examination Survey. Environ Res. 2007 Mar;103(3):413–418.
28. Everett CJ, Matheson EM. Biomarkers of pesticide exposure and diabetes in the 1999-2004 national health and nutrition examination survey. Environ Int. 2010 May;36(4):398–401. PMID: 20299099
29. Cox S, Niskar AS, Narayan KMV, Marcus M. Prevalence of Self-Reported Diabetes and Exposure to Organochlorine Pesticides among Mexican Americans: Hispanic Health and Nutrition Examination Survey, 1982–1984. Environ Health Perspect. 2007 Dec;115(12):1747–1752. PMCID: PMC2137130
30. Turyk M, Anderson HA, Knobeloch L, Imm P, Persky VW. Prevalence of diabetes and body burdens of polychlorinated biphenyls, polybrominated diphenyl ethers, and p,p'-diphenyldichloroethene in Great Lakes sport fish consumers. Chemosphere. 2009 May;75(5):674–679.
31. Eden PR, Meek EC, Wills RW, Olsen EV, Crow JA, Chambers JE. Association of type 2 diabetes mellitus with plasma organochlorine compound concentrations. J Expo Sci Environ Epidemiol. 2016 Mar;26(2):207–213.

32. Son H-K, Kim S-A, Kang J-H, Chang Y-S, Park S-K, Lee S-K, Jacobs Jr. DR, Lee D-H. Strong associations between low-dose organochlorine pesticides and type 2 diabetes in Korea. Environ Int. 2010 Jul;36(5):410–414.
33. Arrebola JP, Pumarega J, Gasull M, Fernandez MF, Martin-Olmedo P, Molina-Molina JM, Fernández-Rodríguez M, Porta M, Olea N. Adipose tissue concentrations of persistent organic pollutants and prevalence of type 2 diabetes in adults from Southern Spain. Environ Res. 2013 Apr;122:31–37.
34. Ukropec J, Radikova Z, Huckova M, Koska J, Kocan A, Sebokova E, Drobna B, Trnovec T, Susienkova K, Labudova V, Gasperikova D, Langer P, Klimes I. High prevalence of prediabetes and diabetes in a population exposed to high levels of an organochlorine cocktail. Diabetologia. 2010 Feb 25;53(5):899–906.
35. Turyk M, Anderson H, Knobeloch L, Imm P, Persky V. Organochlorine Exposure and Incidence of Diabetes in a Cohort of Great Lakes Sport Fish Consumers. Environ Health Perspect. 2009 Jul;117(7):1076–1082.
36. Rignell-Hydbom A, Lidfeldt J, Kiviranta H, Rantakokko P, Samsioe G, Agardh C-D, Rylander L. Exposure to p,p ’-DDE: A Risk Factor for Type 2 Diabetes. PLOS ONE. 2009 Oct 19;4(10):e7503.
37. Al-Othman A, Yakout S, Abd-Alrahman SH, Al-Daghri NM. Strong Associations Between the Pesticide Hexachlorocyclohexane and Type 2 Diabetes in Saudi Adults. Int J Environ Res Public Health. 2014 Sep;11(9):8984–8995. PMCID: PMC4199001
38. Rylander C, Sandanger TM, Nøst TH, Breivik K, Lund E. Combining plasma measurements and mechanistic modeling to explore the effect of POPs on type 2 diabetes mellitus in Norwegian women. Environ Res. 2015 Oct;142:365–373.
39. Park SK, Son HK, Lee SK, Kang JH, Chang YS, Jacobs DR, Lee DH. Relationship between serum concentrations of organochlorine pesticides and metabolic syndrome among non-diabetic adults. J Prev Med Public Health Yebang Ŭihakhoe Chi. 2010 Jan;43(1):1–8. PMID: 20185977
40. Kacew S, Singhal RL. Adaptive response of hepatic carbohydrate metabolism to oral administration of p,p’-1,1,1 -trichloro-2,2-bis (p-chlorophenyl)ethane in rats. Biochem Pharmacol. 1973 Jan 1;22(1):47–57.
41. Yau ET, Mennear JH. The inhibitoty effect of DDT on insulin secretion in mice. Toxicol Appl Pharmacol. 1977 Jan 1;39(1):81–88.
42. Howell GE, Meek E, Kilic J, Mohns M, Mulligan C, Chambers JE. Exposure to p,p’-dichlorodiphenyldichloroethylene (DDE) induces fasting hyperglycemia without insulin resistance in male C57BL/6H mice. Toxicology. 2014 Jun 5;320:6–14. PMCID: PMC4098932
43. Merrill ML, Karey E, Moshier E, Lindtner C, Frano MRL, Newman JW, Buettner C. Perinatal Exposure of Mice to the Pesticide DDT Impairs Energy Expenditure and Metabolism in Adult Female Offspring. PLOS ONE. 2014 Jul 30;9(7):e103337.
44. Pavlikova N, Smetana P, Halada P, Kovar J. Effect of prolonged exposure to sublethal concentrations of DDT and DDE on protein expression in human pancreatic beta cells. Environ Res. 2015 Oct;142:257–263.

45. López-Aparicio P, Recio MN, Prieto JC, Pérez-Albarsanz MA. Role of lindane in membranes. Effects on membrane fluidity and activity of membrane-bound proteins. Biosci Rep. 1994 Jun 1;14(3):131–138. PMID: 7530499
46. Ruiz P, Perlina A, Mumtaz M, Fowler BA. A Systems Biology Approach Reveals Converging Molecular Mechanisms that Link Different POPs to Common Metabolic Diseases. Environ Health Perspect [Internet]. 2015 Dec 18 [cited 2016 Jul 22];124(7). Available from: http://ehp.niehs.nih.gov/15-10308

Supplementary Materials

Table S1. Study design for pilot study of Asian Indian T2D cases and controls

	Controls (n=25)	Cases (n=24)	p-value
Glucose	4.88	9.38	
SexM(%)	0.56	0.72	0.38
Age	51.82	55.67	0.31
Telugu(%)	0.25	0.11	0.72
Smoker (%)	0.04	0.12	0.6
Drinker (%)	0.24	0.2	0.99
BMI	26.4	26.57	0.87
WHR	0.95	0.97	0.38
Hypertension (%)	0.28	0.48	0.24
HDL	1.32	1.22	0.33
TG	1.6	1.86	0.28
LDL	3.24	2.16	0.0004
Chol	5.29	4.35	0.005

Cases were defined as fasting blood glucose ≥7.0 mmol/L. Cases were frequency matched based on age, sex, WHR, ethnicity, and smoking status.

Table S2. Frequency of POPs detected in pilot study (N=49)

PCBs	Freq
PCB-118	48
PCB-123	48
PCB-126	47
PCB-156	43
PCB-169	42
PCB-114	33
PCB-167	24
PCB-105	17
PCB-157	0
PCB-77	0
PCB-81	0
PCB-189	0

BDE	Freq
BDE-99	49
BDE-47	49
BDE-100	40
BDE-153	4
BDE-154	1
BDE-66	1
BDE-49	1
BDE-28 and BDE-33	1
BDE-15	0
BDE-75	0
BDE-155	0

PAHs	Freq
Acenaphthene	49
Anthracene	49
Fluoranthene	49
Fluorene	49
Naphthalene	49
Phenanthrene	49
Pyrene	49
Benzo(g,h,i)perylene	19
Indeno(1,2,3-cd)pyrene	10
Benzo(a)pyrene	6
Benz(a)anthracene and Chrysene	4
Acenaphthylene	2
Benzo(b)fluoranthene	2
Dibenz(a,b)anthracene	2
Benzo(k)fluoranthene	0

Furans	Freq
Octachlorodibenzofuran	1
1234678-Heptachlorodibenzofuran	0
2,3,7,8-Tetrachlorodibenzofuran	0
1,2,3,7,8-Pentachlorodibenzofuran	0
1,2,3,4,7,8-hexachlorodibenzofuran	0

Organocholorine	Freq
Endosulfane II	49
p,p'-DDE	49
p,p'-DDT	48
p,p'-DDD	46
Aldrin	44
Methoxychlor	11
beta-HCH	48
delta-HCH	0
alpha-HCH	0
gamma-HCH	0
Heptachlor	0
Endosulfane I	0
Endosulfane sulphate	0
Endrin.Aldehyde	0
Heptachlor Epoxide	0
Dieldrin	0
Endrin	0

Dioxin	Freq
Octachloro.Dibenzodioxin	2
1234678.Heptachloro.Dibenzodioxin	0
123478.Hexachloro.Dibenzodioxin	0
12378.Pentachloro.Dibenzodioxin	0
2378.Tetrachloro.Dibenzodioxin	0

Table S3. Detection limits and precision of small-volume method to detect POPs

						Based on Pooled Ref Sample			
Compound Name	MDL (fg on column)	LOD (ng/mL)	LOQ (ng/mL)	Samples <LOD	Samples <LOQ	Mean Interbatch CV (%)	Mean Intrabatch CV (%)	Total Number Detected (of N=198)	Median, ng/mL (Range, ng/mL)
Alpha-HCH	8.427	0.004	0.042					**140**	
Beta-HCH	9.452	0.005	0.047	0	16	6.2%	11.2%	196	0.2175 (0.01662, 4.5020)
Gamma-HCH	9.892	0.005	0.049					**153**	
o,p-DDE	1.394	0.001	0.007					88	
p,p-DDE	9.032	0.005	0.045	0	0	11.0%	12.7%	197	1.6810 (0.0776, 145.80)
o,p-DDD	7.603	0.004	0.038					**191**	
PCB123	9.075	0.005	0.045					16	
PCB118	18.665	0.009	0.093	12	188	13.6%	14.2%	196	0.02617 (0.00636, 0.2359)
p,p-DDD	56.595	0.028	0.283					198	
o,p-DDT	1.492	0.001	0.007					198	
PCB114	18.060	0.009	0.090					121	
PCB105	12.449	0.006	0.062					**174**	
p,p-DDT	12.788	0.006	0.064	9	107	7.6%	1.8%	198	0.0577 (0.0042, 2.1710)
PCB167	8.317	0.004	0.042					**162**	
PCB156	11.703	0.006	0.059					**182**	
PCB157	16.486	0.008	0.082					86	
PCB189	16.473	0.008	0.082					72	

Table S4a. Exposure characteristics for diabetics versus controls within ethnicity groups (ng/g-lipid)

			Controls		Cases			
Compound (ng/g-lipid)	**Population**	**Exposure Status**	**N**	**Median (Range)**	**N**	**Median (Range)**	**Odds Ratios (95% CI)***	***p*-value**
p,p'-DDE	Tamil	<776.90	31	332.4 (26.82, 763.3)	5	208.5 (141.38, 716.6)	4.23 (1.34, 13.35)	0.02
		>776.90	22	1430.0 (776.90,25143.8)	15	1587.0 (810.60,4786.0)		
	Telugu	<600.90	23	301.83 (72.09,541.58)	0	N/A	Inf	0.11
		>600.90	20	864.34 (600.95,5676.31)	4	2151.00 (1553.72, 6212.58)		
p,p'-DDT	Tamil	<13.82	27	8.64 (4.14, 13.00)	9	8.92 (6.24,13.11)	1.27 (0.45, 3.56)	0.80
		>13.82	26	18.68 (13.88,190.36)	11	23.27 (13.82,79.84)		
	Telugu	<28.24	22	18.32 (3.91,27.01)	1	24.33	3.14 (0.30, 32.65)	0.61
		>28.24	21	44.68 (28.24,316.45)	3	131.41 (64.67,194.89)		
PCB-118	Tamil	<5.97	25	2.73 (0.94, 5.66)	11	4.80 (2.03, 5.78)	0.73 (0.26, 2.05)	0.61
		>5.97	28	8.04 (5.97,34.00)	9	9.81 (6.32, 27.34)		
	Telugu	<3.20	20	2.07 (0.81, 3.15)	3	2.49 (2.35, 3.20)	0.29 (0.03, 3.01)	0.35
		>3.20	23	3.89 (3.20, 34.21)	1	3.87		

*Adjustment for age, WHR, sex, smoking status and alcohol use did not change the effect size or significance levels except for *p,p'*-DDE in Tamils, OR_{adj}=3.41 (0.91, 12.80). (Adjusted ORs are likely inaccurate estimates because of the small number of cases and are not reported in the table.)

Table S4b. Exposure characteristics for diabetics versus controls within ethnicity groups (ng/mL)*

			Controls		Cases			
Compound (ng/mL)	**Population**	**Exposure Status**	**N**	**Median (Range)**	**N**	**Median (Range)**	**Odds Ratios (95% CI)***	***p*-value**
p,p'-DDE	Tamil	<4.62	29	1.90 (0.17,3.06)	7	2.86 (1.18,3.92)	2.53 (0.63, 10.09)	0.20
		>4.62	24	8.10 (4.62,145.85)	13	10.20 (4.70,27.78)		
	Telugu	<3.69	23	1.65 (0.45,3.59)	0	N/A	Inf	0.11
		>3.69	20	6.21 (3.69,43.92)	4	15.42 (10.03, 52.99)		
p,p'-DDT	Tamil	<0.08	25	0.05 (0.02,0.08)	11	0.06 (0.03,0.07)	1.07 (0.28, 4.11)	0.61
		>0.08	28	0.12 (0.08,1.10)	9	0.13 (0.08,0.46)		
	Telugu	<0.19	23	0.12 (0.02,0.19)	0	N/A	Inf	0.11
		>0.19	20	0.30 (0.19,2.17)	4	0.61 (0.24, 1.34)		
PCB-118	Tamil	<0.04	25	0.019 (0.006,0.038)	11	0.021 (0.013,0.038)	1.95 (0.47, 8.16)	0.61
		>0.04	28	0.050 (0.039, 0.236)	9	0.060 (0.047, 0.116)		
	Telugu	<0.02	21	0.015 (0.006, 0.020)	2	0.018 (0.017, 0.019)	15.42 (0.26, 920.59)	0.99
		>0.02	22	0.026 (0.020,0.225)	2	(0.026) (0.020,0.031)		

*Only adjusted for triglycerides and cholesterol levels. Adjustment for age, WHR, sex, smoking status and alcohol use did not widely change the effect size nor significance level. (Adjusted ORs are likely inaccurate estimates because of the small number of cases and are not reported in the table.)

Figure S1. Distribution of POPs concentrations

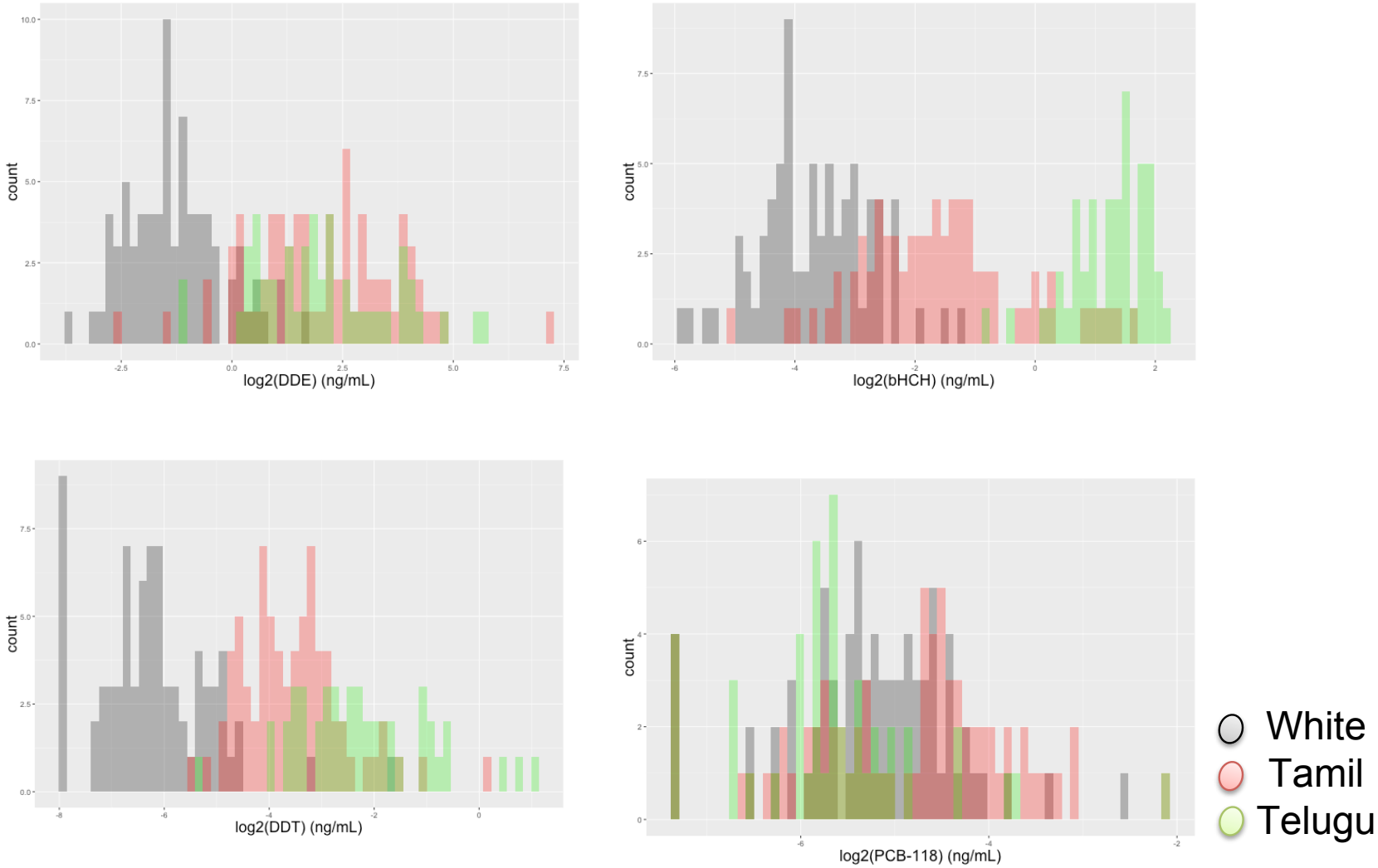


Figure S2. Exposure concentrations vs. glucose concentrations in cases versus controls for each ethnicity (dotted line indicates median concentration value)

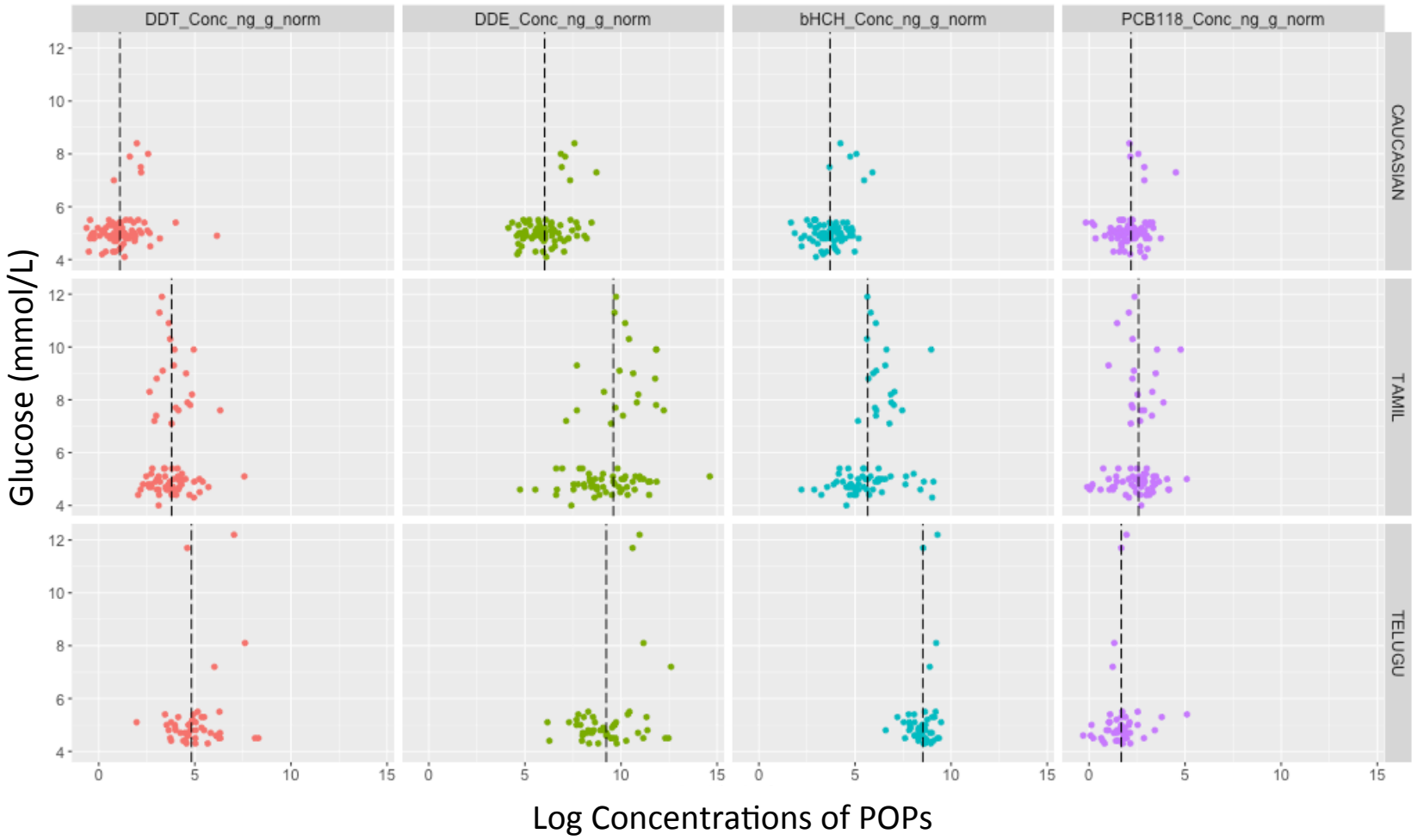
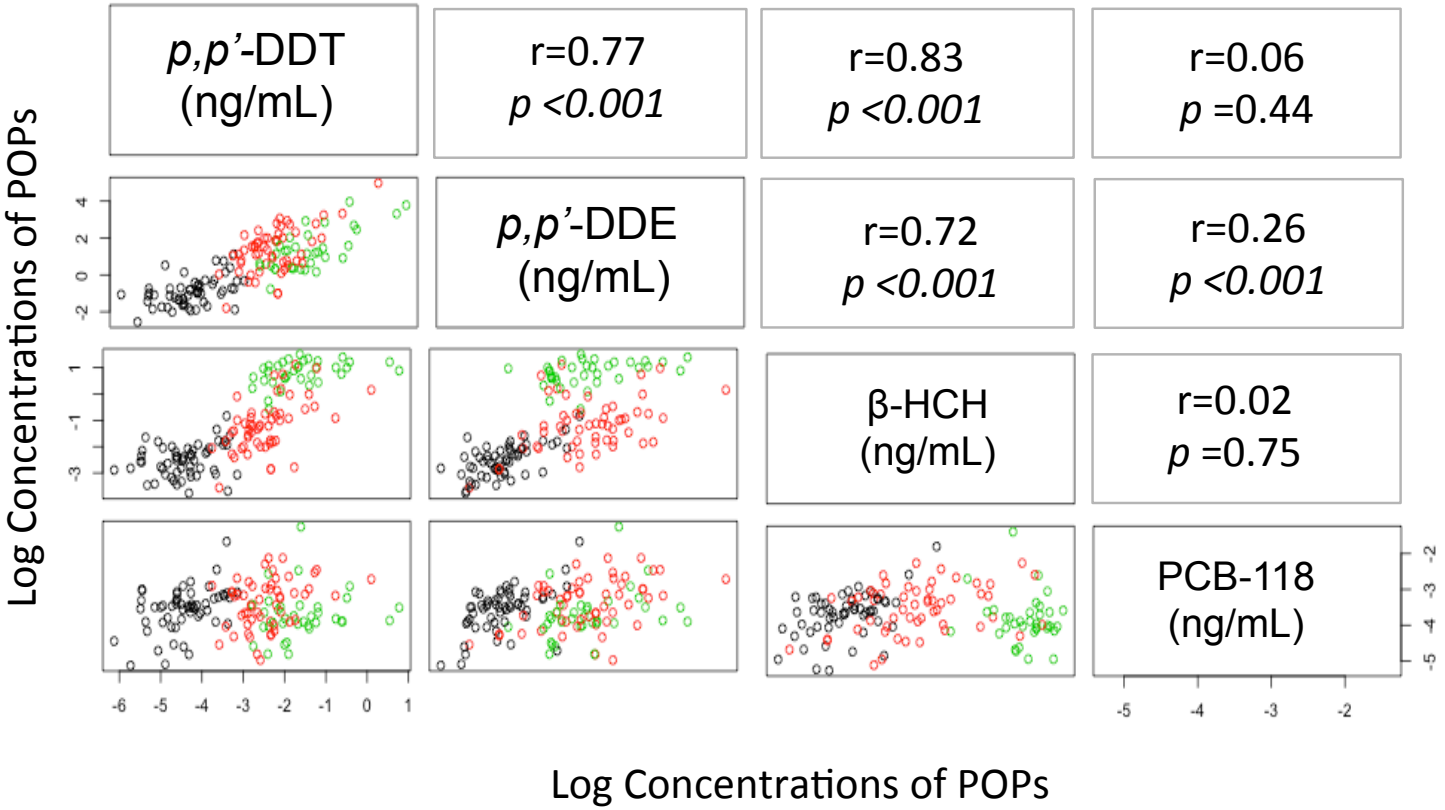


Figure S3. Correlation of POPs Concentrations



Supplementary Methods

Mass Spectrometry for Persistent Organic Pollutants continued:

We designed each experimental run on the mass spectrometer to include 1) standard curve calibration for each analyte, 2) quality controls to monitor any analyte drift during the run, and 3) internal standards spiked into each sample to normalize analyte peak area. Calibrators were prepared by adding working standard mixtures containing all analytes to charcoal stripped fetal bovine serum (ThermoFischer Scientific Waltham, MA). The concentration of the calibrators ranged from 0.01 ng/mL - 10 ng/mL for *p,p*'-DDD; *p,p*'-DDE; *p,p*'-DDT and 0.001 ng/mL - 1.0 ng/mL for all other analytes. For *p,p*'-DDD; *p,p*'-DDE; *p,p*'-DDT, quality controls (QC's) were prepared in charcoal stripped fetal bovine serum at 1.0 ng/mL and 10.0 ng/mL. All other QC's were prepared at 0.1 ng/mL and 1.0 ng/mL in charcoal stripped fetal bovine serum. (For the pilot study, only 3 calibrators and 2 QCs used for the pilot study were only prepared in isooctane as serial dilutions from a compound mixture.) The isotopically labeled internal standards were added to each calibrator, QC, pooled reference sample and unknown plasma sample. The internal standards consisted of $^{13}C_{12}$-*p,p*'-DDT, $^{13}C_6$-β-HCH (Cambridge Isotopes Tewksbury, MA) and fluorinated internal standards representative of penta-chlorinated (5'-F-PCB-105), hexa-chlorinated (5'-F-PCB-156) and hepta-chlorinated (5'-F-PCB-190) PCB homologues (Chiron AS Trondheim, Norway).

Plasma was extracted in four batches of 50 samples each, using chemical denaturation, liquid-liquid extraction, solid-phase cleanup and reconstituted with hexanes. The procedure was as follows:

- Aliquot 200 µL plasma
- Add 1mL 10M urea
- Add 1mL 10% Propanol/Water, 1mL MeOH, 6mL petroleum ether
- Centrifuge, transfer the organic layer
- Filter through 0.25g sodium sulfate layered on top of 1g florisil
- Elute with methyl-t-butyl ether / petroleum ether
- Evaporate to dryness
- Reconstitute with 50 µL hexanes prior to injection

An Agilent (Santa Clara, CA) 7890B GC coupled to an Agilent 7000C GC-triple quadrupole mass spectrometer operated in electron impact (EI) multiple reaction monitoring (MRM) mode, was used for the analyses. A DB-5MS (15 m x 250 µm x 0.25 µm) column (Agilent J&W 122-5512, Santa Clara, CA) was used with helium (He) carrier gas at a constant flow of 1.2 ml/minute. (For the pilot study, an Agilent 7890B/7010 GC-MS/MS instrument and a 30 m HP-5MS column were used.) Two microliter injections were made in pulsed split-less mode at 290 °C. The oven program ranged from 60 °C to 310 °C. The transfer line temperature and EI source were set at 290 °C and the quadrupole temperatures were 180 °C. Nitrogen collision gas was set at 1.5 mL/min and Helium gas at 2.25 mL/min was used to quench metastable helium and reduce neutral noise. Internal standards were used to account for signal attenuation for each class of compounds. For each compound, one quantifying MRM and one qualifying MRM was

defined. All calibrators, QC's, pooled reference samples and plasma samples were injected in duplicate batches, however, only the first injection was used for downstream analysis.

Minimum detection limits (MDL) were determined by equation 1. Limits of detection (LOD) were defined as the MDL converted to ng/mL and limits of quantitation (LOQ) were defined as 10xLOD.

Equation 1. $MDL_{(n-1,\ 99\%\ confidence)} = t_{\alpha}(\%RSD)(mass\ injected)\left(\frac{1}{100}\right)$

Where the percent relative standard deviations (%RSD) = [(standard deviation / mean)100] and t_{α} is the statistical confidence factor given in the in the one-sided Student t-distribution table for n-1 degrees of freedom at the 0.99 confidence level. In these determinations n=7, was used for 2,4'-DDE, Alpha-HCH b-HCH, 2,4'-DDD PCB-123 PCB118, PCB-114, PCB-157 and n=8 for all others. A five-times MDL (5xMDL) acceptance criterion was applied to each MDL determination and given in equation 2. Table S1 illustrates calculated MDLs, LODs and LOQs (10 x LOD).

Equation 2. MDL < mass of analyte injected on column < 5 x MDL.

Chapter 3: Improving Power to Detect Changes in Blood miRNA Expression by Accounting for Sources of Variability in Experimental Designs*

Sarah I. Daniels, Fenna C. M. Sillé, Audrey Goldbaum, Brenda Yee, Ellen Key, Luoping Zhang, Martyn T. Smith, Reuben Thomas

Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, USA

*A similar version of this manuscript has been published: Sarah I. Daniels, Fenna C. M. Sillé, Audrey Goldbaum, Brenda Yee, Ellen Key, Luoping Zhang, Martyn T. Smith, Reuben Thomas. Improving Power to Detect Changes in Blood miRNA Expression by Accounting for Sources of Variability in Experimental Designs. 2014 Dec;23(12):2658-66. doi: 10.1158/1055-9965. This chapter is printed here with acknowledgement to all co-authors and AACR.

ABSTRACT

Background: Blood microRNAs (miRs) offer a new promising area of disease research, but variability in miR measurements may limit detection of true-positive findings. Here, we measured sources of miR variability and determine whether repeated measures can improve power to detect fold-change differences between comparison groups.

Methods: Blood from healthy volunteers (*N*=12) was collected at three time points. The miRs were extracted by a method predetermined to give the highest miR-yield. Nine different miRs were quantified using different qPCR assays and analyzed using mixed models to identify sources of variability. A larger number of miRs from a publicly-available blood miR microarray dataset with repeated measures was used for bootstrapping to investigate effects of repeated-measures on power to detect fold-changes in miR expression for a theoretical case-control study.

Results: Technical variability in qPCR replicates was identified as a significant source of variability ($p<0.05$) for all nine miRs tested. Variability was larger in the TaqMan qPCR assays (*SD* = 0.15-0.61) versus the qScript qPCR assays (*SD* = 0.08-0.14). Inter- and intra- individual and extraction variability also contributed significantly for two miRs. The bootstrapping procedure demonstrated that repeated measures (20-50% of *N*) increased detection of a 2-fold change for ~10-45% more miRs.

Conclusion: Statistical power to detect small-fold changes in blood miRs can be improved by using repeated measures and choosing appropriate methods to minimize variability in miR quantification.

Impact: This study demonstrates the importance of including repeated measures in experimental designs for blood-miR research.

Introduction

The use of microRNAs (miRs) as blood-based biomarkers is a new field of research for diagnostic and preventative medicine. A limitation of this field is the lack of statistical power to detect true differences between comparison groups, which can contribute to difficulties in validating results. Accounting for sources of variability in the experimental design may increase power in blood-biomarker studies. Previously, we demonstrated that by controlling for technical variability in preparation of blood RNA for microarray analysis, we were able to improve power to detect small, yet significant, fold-changes in blood transcriptomic data (1). Here, we assess sources of inter- and intra-individual and technical variability for miRs found in blood samples and predict how repeated measures can improve power to detect differences in miR expression.

MicroRNAs have been widely studied as biomarkers for a number of diseases. These small non-coding transcripts regulate translation of RNA by binding to the 3' untranslated region of target RNA. Overall, miRs regulate 30-60% of RNA translation to protein, usually by down-regulation of the transcript (2,3). Disease status, chemical exposures, and life-style factors have been linked to differences in expression of miRs between individuals (discussed in 4–6). However, as most reported miR expression fold-changes are small (~1.5-2-fold), it is difficult to replicate findings and discover true associations. Therefore, it is critical to control for important sources of variability in the experimental design.

Variability in RNA transcription within subjects over time has seldom been discussed in the literature, particularly for microRNAs (miRs). Several transcriptomic studies have shown limited fluctuation in blood RNAs when measured from healthy individuals over weeks to months (7–11). The proportion of transcripts with high intra-individual variability was attributed to a small number of immunological genes (i.e., immunoglobulin) (9) or could not be differentiated from technical variability due to poor experimental design (10,11). This evidence from transcriptomics suggests that there may be similarly small intra-individual variability for miR transcription, however, this has not been previously measured.

Other overlooked sources of variability include methods for miR quantification and extraction. For example, competing platforms for miR microarray and qPCR analysis have shown differences in sensitivity(12–14), which suggests that some variance in miR measurements may be due to technical variability. For processing of fresh blood samples, miRs studied in specific blood-partitions (i.e., plasma, red blood cells, platelets and leukocytes) have attributed certain miR expression in plasma and serum to contamination of red blood cells and platelets (15–18). Extraction of miRs can also introduce variability as systematic differences can depend on the particular method or manufacturer (14,19–23). Most of these previous studies focused on samples obtained from cell lines and did not thoroughly compare miR yield obtained from primary cells.

Here, we hypothesize that there are important sources of inter-, intra- and technical variability in miRs extracted from primary human peripheral blood mononuclear cells (PBMCs). We calculated the contributions of these sources of variability using experimental data obtained by qPCR and compared them to estimates obtained from a previously published study. As PBMCs are a popular and non-invasive sample-type and can be affected in early

stages of disease, it is important to improve methods of pre-analytical processing of PBMC biomarkers for future disease-related research.

Materials and Methods:

Comparison of RNA extraction kits

Four kits were compared to each other for miR yield: miRNeasy kit (Qiagen, Valencia, CA), mirVana kit (Ambion / Life Technologies, Grand Island, NY), ZR-*duet* (Zymo Research Corporation, Irvine, CA), and Trizol (Life Technologies, Grand Island, NY), with the addition of 25 nmoles-250 nmoles of *c. elegans* oligos spike-ins (cel-39 and cel-54) to each sample during the cell lysis step. A total of nine miRNAs (seven endogenous miRNAs and two exogenous spike-ins) were compared using these 4 extraction kits, with samples from 4 individuals and two technical qPCR replicates for each individual (included in the residual variability). In this fractional factorial design, the variance components are shown below for each of the tested miRNA:

$$\sigma_i^2 = \text{ variability between individuals}$$
$$\sigma_j^2 = \text{ variability between extraction kits}$$
$$\sigma_e^2 = \text{ residual variability}$$

(The AllPrep kit, comprised of on-column extraction of both DNA and RNA, was also compared to the miRNeasy kit. Details are included in the Supplementary Methods.)

Blood sample collection from volunteers: In order to calculate sources of inter- and intra-individual and technical variability, we measured miR expression in PBMCs of healthy volunteers over an 8-month time period. A sample size of *N*=12 healthy subjects were included in the study (exclusion criteria for volunteer subjects were chronic illness or pregnancy at the time of blood draws). Samples were obtained at three time points, roughly 2-4 months apart. On the day of collection, blood samples were processed to isolate PBMCs. Data collection for this study was approved by the Internal Review Board within University of California Berkeley's Human Research Protection Program. Informed consent was obtained from all participants.

PBMCs were isolated from fresh whole blood collected in EDTA tubes using the standard Ficoll gradient protocol (24). Upon isolation of the PBMCs, they were immediately washed in PBS, pelleted, and resuspended in aliquots of RNAProtect Cell reagent (Qiagen, Valencia, CA) and frozen at -80°C until further use. At a later time, PBMC samples were thawed and RNA was extracted by the miRNeasy kit (Qiagen) as described in the Supplementary Methods.

Real-time PCR quantification

Probe-based miRNA TaqMan Assays (LifeTechnologies, Grand Island, NY), or SYBR Green-based qScript - PerfeCTa microRNA Assays (Quanta BioSciences, Gaithersburg, MD) were used to quantify miR targets of interest. Reaction volumes were proportionately scaled-down from the initial protocols (see Supplementary Methods). Six miRs (miR-30d, let-7d, miR-185, miR-130a, miR-451, miR-342-3p) were chosen based on overlap between miRs expressed in the miR microarray dataset (used for some of the simulation studies) (25) and miRs differentially expressed in PBMCs of Type II diabetics (26). One miR (cel-39) was used as an exogenous control for elution variability. Two small RNAs (SNU6 and RNU48) and miR-16, frequently used for normalization of miR expression, were included as well.

Statistical analysis of qPCR results:

To assess the contributions of different sources of variability in miRNA expression of the volunteer blood samples, we used the following mixed-effects model :

$$Y_{ijklmn} = \beta_0 + \beta_1 + \beta_{0i} + \beta_{0j} + \beta_{0k} + \beta_{0l} + \beta_{0ijm} + \varepsilon_{ijklmn} \quad (1)$$

The random effects in model (1) are each normally distributed, with zero means and variances given by the respective variance components listed below:

Random Effect	Variance Component	Description
β_{0i}	${\sigma_i}^2$	variability between individuals (biological)
β_{0j}	${\sigma_j}^2$	variability over time (biological)
β_{0k}	${\sigma_k}^2$	inter − batch variability (technical)
β_{0l}	${\sigma_l}^2$	intra − batch variability (technical)
β_{0ijm}	${\sigma_{ijm}}^2$	variability in qPCR reaction (technical)
ε_{ijklmn}	${\sigma_e}^2$	residual variability

Y_{ijklm} represents the Cq value (on the natural scale) for the i^{th} individual at the j^{th} time point in the k^{th} extraction batch, the l^{th} within-batch replicate, and the m^{th} technical replicate. The intercept, β_0, is defined as the baseline Cq value averaged across all individuals, at all time points, for all extraction batches on all plates. The fixed effect coefficient, β_1, represents a change in miR Cq value per unit change in RNA concentration for each sample, which we used as precaution against any effects that were not accounted for by using the same input (µg) of RNA for each RT-PCR reaction. The random effects are defined as the change in expression (Cq value) from baseline levels for each index. β_{0i} is the unit change from the baseline average for the i^{th} (i= 1…13) individual. β_{0j} is the unit change from the baseline average at the j^{th} (j= 1, 2, 3) time point. B_{0k} is the unit change in the k^{th} (k=1, 2) extraction batch. β_{0l} is the unit change from baseline for the l^{th} (l= 1, 2) within-extraction replicate. β_{0ijm} is the unit change from baseline for the m^{th} (m =1, 2, 3) technical replicate for the i^{th} individual and the j^{th} timepoint. (This is the only nested random effect measure.) Lastly the term, $\boldsymbol{e_{ijklmn}}$, is defined as the 'residual variability,' which includes differences from plate-to-plate and other unaccounted for sources of variability. Nested F-test were run for each small RNA model to determine which random effects terms were significant (p<0.05).

Estimating minimum detectable fold-changes based on qPCR data

We observed several measurable sources of variability for two miRs from the qPCR experiment, miR-185 and miR-451. These two miRs were used to determine if repeated measures would improve detection of fold-changes in a theoretical study. The constraints for the theoretical study were a sample size of N= 75 vs. 75 subjects in two comparison groups (e.g., disease vs. healthy controls) under two experimental designs conditions. Study 1 had no repeated measures and Study 2 had four repeated measures for 50% of the subjects for each of the four modeled sources of variability based on our empirical qPCR data— seasonal, between-batch, within-batch and qPCR replicate. The estimates of the parameters of interest, which is

the minimum detectable fold-change in the mean level of the miRNAs, are computed using the variance values attributed to the four different sources of variability from the empirical study data of *N*=12 subjects.

Therefore, in order to obtain estimates of variability for the parameters of interest associated with repeating the empirical study, a clustered bootstrap procedure (using 100 bootstrap samples) was used. This procedure provides both a point estimate and confidence intervals for these fold-change estimates. Each nonparametric bootstrap sample consists of data associated with 12 subjects drawn with replacement from the *N*=12 subjects of the empirical study. The data associated with a bootstrap sample was used to estimate the variances attributed to the four sources of variability using linear mixed models (27). The distribution of the minimum detectable fold-changes over the 100 bootstrap samples was used to estimate the confidence intervals of this parameter of interest. (As the sample size used for the bootstrap bootstrap sample is relatively smaller than the sample size for the theoretical study, the inferences made from the simulation must be interpreted with caution.)

A simulation procedure was used to estimate the parameters of interest for Study 1 and Study 2 given the variance estimates from a bootstrap sample, *bs*. For each miR within a given study design, data for 100 studies were simulated (assuming a normal distribution for each random effect) using the *bs* variances estimates. The standard deviation of the mean expression across these 100 studies provides an estimate of the standard error ($SE_{i,bs}$) of the mean level of the miR in a *bs* sample for a given study design. This standard error was then used to estimate the minimum detectable fold-change, $FC_{i,bs}$, with 80% power (corresponding to a 5% family-wise error rate) by the following equation:

$$FC_{i,bs} = \sqrt{}2(Z_{\alpha} + Z_{\beta})SE_{i,bs} \quad (2)^{*}$$

Where Z_{α}= 1.64 (desired level of statistical significance)

Z_{β} = 0.84 (desired power)

Estimating minimum detectable fold-changes based on previously published data

To expand upon our findings of variability for individually-tested miRs by qPCR, we examined a publicly-available microarray dataset by Honda *et al* (25) that measured hundreds of blood miRs simultaneously at several time-points for each subject. The study looked at the effects of chronic academic stress on miR levels in whole blood of medical students by obtaining measurements two months before, two days before, and one month after an exam for medical practitioners. The GSE49677 series from the Gene Expression Omnibus (28) Agilent-021827 Human miR Microarray (V3) was used in this study. The low-expressed miRs with mean

---

* This equation is derived from: $$n = \frac{2\sigma^2(Z_{\beta} + Z_{\alpha})^2}{(\text{fold change difference})^2}$$

* An exception: miR-375 did not initially pass the Shapiro-Wilk test. Analyses were repeated after removal of three outliers and resulting effect size and p-value were similar to the reported values here.

intensity levels <20 were filtered out in the study (25) leaving 143 miRs. The levels of these miRs across the four subjects and three time points are normalized using Cyclic Loess (29).

Variability estimates from the academic-stress data were used to determine whether repeated measures would increase detection of differentially expressed blood miRs between two comparison groups (e.g., disease versus healthy controls). We calculated estimates of inter-individual variability from this study and assumed that the residual (unexplained) variability in blood miR levels was due to other sources. For ease of exposition, we assumed the residual to be time-point (e.g., seasonal) variability, although it is probably composed of multiple sources.

Based on the expression data of blood miRNA measured in the academic stress study (25) with four subjects and three repeated measures, we determined if repeated measures would improve detection of fold-changes in a theoretical study. For our theoretical replicate design simulation study, our sample size was 75 vs. 75 subjects in each of two comparison groups (e.g., disease vs. controls). A total of 2000 markers were evaluated for purposes of multiple testing under realistic omic-level conditions. We assumed that the sample collection for the subjects in the two groups occurred at two different time points. Therefore, seasonal-effects on miR levels were not blocked in these experimental designs. We varied the proportion of subjects with repeated measures and the number of repeated measures per subject for each of the seven proposed designs shown in Table 1. We used a non-parametric clustered bootstrap method, similar to the one described for the simulation of qPCR data, in order to predict minimum detectable fold-changes in the mean levels for the 143 miRNAs under the seven theoretical experimental design conditions. (Again, as the sample size used for the bootstrap sample is relatively smaller than the sample size for the theoretical study, the inferences made from the simulation must be interpreted with caution.) We provided confidence intervals for our parameters of interest based on estimates of inter-individual and residual variability (assumed to be partially attributed to biological variability) in the study on academic stress.

Results and Discussion

Comparing methods of miRNA extraction

We evaluated miR extraction procedures to find the most efficient and accurate method for our downstream applications. We presumed that lower Cq values for a given extraction method would be a proxy for both greater overall yield of all miRs and lower technical variability (i.e. between and within a given batch of extractions). We compared four methods; miRNeasy, miRVana, Trizol, and Zymo-*Duet* (which extracts both RNA and DNA). The miRNeasy kit had the lowest Cq value for all small RNAs tested (see Supplementary Results and Supplementary Figure S1A) which is supported by similar previous findings (14). The miRNeasy kit also slightly out-performed the AllPrep kit (Qiagen) (Supplementary Figure S1B) and was thus chosen as the extraction method for the downstream measurements of miR variability in the volunteer blood samples.

Measuring sources of variability in miR from qPCR data

We measured several sources of variability for nine chosen miRs from 12 volunteer subject PBMC samples obtained at three time points over the course of several months. We included extraction replicates to account for both between- and within-extraction batch

variability. Our results were analyzed using mixed-effects models (shown in Equation 1). The variance attributed to each source of variability for a given miR is shown in the stacked bar graph (Figure 1A and 1B), and the significance of each term in the model is reported in Table 2.

A nested F-test for each of the random effects in each miR model was used to identify the significant random effect terms ($p<0.05$). For all endogenous and exogenous small RNAs tested, a significant proportion of variability was due to replicate qPCR reactions for a given sample. For miR-16 and miR-451, inter-individual variability was also significant. Furthermore, for the qScript SYBR Green assays, miR-451 showed a significant "time point" effect while miR-16 showed a significant batch-replicate effect. This is the first evidence to suggest time-point and extraction variability in miR expression. Other transcriptomics studies have also found differences in RNA expression over the course of 1-day to several weeks (7,8) while longer time-points were inconclusive due to confounding of technical variability (10,11).

The residual variability present for all miRs may be due to several sources. Covariate information was not included in this model, such as age, race, gender and BMI, which may contribute to the residual variability. Technical variability in sample processing may also be another contributing factor. For example, Ficoll separation of PBMCs is not 100% efficient, so miRs such as miR-16 and miR-451, both known to be highly-expressed in red blood cells (15)), could contribute to variability of these miR expression levels. SNU6 and the exogenous miR, cel-39, demonstrated the highest proportions of residual variability (Figure 1B). As expected, biological and technical variability contributed minimally to total variability and overall variability for the spike-in cel-39, and it had the lowest total variability of all miRs tested. For SNU6, the source of residual variability remains unknown, but perhaps this small RNA is not ideal to use for normalization of target miRs in future studies if sources of variability are not representative of other miRs.

The remaining miRs did not show significant contributions of variability from the other measured sources. This may be because 1) the sample size was too small too assign statistical significance or 2) the scaled-down volumes for the TaqMan assay were not sufficient to measure these effects. Comparing miR-16 measurements in both assays, the variance term for the qPCR replicates in the model was much smaller for the SYBR Green assay than the TaqMan assay (0.119 vs. 0.499) and are plotted for each individual in Supplementary Figure S2A and B. Additionally, three significant random-effects terms were found with the SYBR Green assay for miR-16, while only two were significant using the TaqMan kit, providing further evidence that perhaps other sources of variability could be unveiled if qPCR replicate variance was reduced. A less rigorous comparison of the two assays has been made in a previous study, however, the opposite results were found (13). Our results on the performance of miR SYBR Green-based qPCR are supported by a very recent study that examined miR expression analysis of qScript and several other platforms in much greater detail (30). As many studies use TaqMan-based assays, our results could help explain the lack of reproducibility reported between studies examining miRs in the same tissue for the same disease in similar populations.

Estimating effects of repeated measures from qPCR data

We used the estimates of inter-individual, intra-individual, and technical variability for two of the miRs (miR-185 and miR-451) for further analysis in a theoretical study of N= 75 vs. 75

subjects. We calculated the minimum detectable fold-change with 80% power in a study with no repeated measures (Study 1) versus four repeated measures for each of the following; between-batch, within-batch, time point and qPCR replicates for 50% of the subjects in each group (Study 2). The minimum fold-change estimate for miR-451 decreased with repeated measures from 3.77 (95% CI [1.75, 5.16]) to 2.38 (95% CI [0.97, 3.36]) and the minimum fold-change estimate for miR-185 decreased from 4.18 (95% CI [2.67, 6.46]) to 2.4 (95% CI [1.71, 3.74]) (Table 3). For miR-185, a marginally significant (90% CI) decrease in fold-change was observed with repeated measures from Study 1 to Study 2. Our estimate of detectable differences in fold-change for miR-185 is similar to previous findings that showed a 1.82-fold difference between disease conditions measured in PBMC samples (26).

Estimating variability in miR from previously published data

We expanded our investigation of repeated measures to examine more miRs simultaneously, as is currently done in omics-level studies. We used a previously published miR microarray dataset on four medical students over three time points to estimate the variability in 143 miRs (25). The inter-individual variability of each miR from our empirical study (miR-342, miR-451, miR-16, miR-185, miR-30d, let-7d, miR-130a) was compared to results obtained from Honda *et al*. There was no significant correlation between the two estimates (data not shown). This lack of correlation may be explained by differences in expression variability in each of the sample types (PBMCs versus whole blood), as high expression for some of these miRs has been reported in red blood cells (15), or by the small sample sizes used to estimate variability for the qPCR data and the Honda *et al.* dataset.

Estimating effects of repeated measures using simulated miR microarray data

From the Honda *et al.* dataset, we simulated data for a theoretical study to demonstrate benefits of including repeated measures. We ran the analyses with 20%, 50% and 100% of the subjects randomly chosen for repeat sampling, and tested a total of seven different experimental designs summarized in Table 1. The cumulative distribution over the 143 miRs for the minimum detectable fold-change across the 100 bootstrap samples is plotted for each of the seven replicate designs (Figure 2A-C). Without repeated measures, a ≥2-fold change could be detected in ~24% of miRs. Inclusion of repeated measures for 20% of the samples improved the detection rate to 34% and 59% for Designs 1A and 1B, respectively (Figure 2A). When repeated measures were included for 50% of the samples, the detection rate for a ≥2-fold change improved further to 46% and 69% for Designs 2A and 2B, respectively (Figure 2B). Only a minimal increase in detection rate was gained beyond this when performing repeated measures on 100% of the samples in Designs 3A and 3B (Figure 2C).

The estimates of 95% confidence intervals (based on bootstrapping) for these cumulative distribution curves overlapped under each design scenario (Supplementary Figure S3A-C), providing inconclusive evidence of statistically significant improvements of one design over another. To investigate this further, we compared each of the three designs to each other within the 50% repeated-measure parameter (Design 0 vs. 2A vs. 2B) and looked at a range of confidence intervals (i.e., *p*-values) for the minimum detectable fold-change of miRs in one study versus another. The proportion of miRs with detectable fold-change differences in one

design versus another at a given *p*-value is plotted for each pair of designs (Figure 3). From this analysis, Design 2B (with four repeated measures) shows lower detectable fold-changes for ~20% of the miRs (at $p<0.10$) than Design 0 (with no repeated measures). Similar comparisons for Design 1B vs. Design 0 showed very few miRs with lower detectable fold-changes, whereas Design 0 vs. Design 3B showed lower detectable fold-changes in ~40% of miRs (at $p<0.10$) (see Supplementary Results and Supplementary Figure S4A and B, respectively).

While inferences can be made from the simulations of this miR microarray dataset, there are still several limitations of this study. First, we assumed that the remaining variability after accounting for inter-individual differences is due to intra-individual variability over time, however, technical variability from time-point-to-time-point are included too. Our estimates of intra-individual variability may be higher than expected due to the lack of technical variability measurements. Also, changes in study participants' stress levels during the collection time points (25) might also lead to over-estimates of intra-individual variability. Note that in order to mimic realistic omic-level conditions, the theoretical experimental design for this study was limited to $N= 75$ vs. 75 subjects, and we used 2000 measured endpoint markers (including the 143 miRs) for purposes of multiple-hypothesis testing. Altering these parameters by including a larger (or smaller) number of subjects and/or a greater (or reduced) proportion of repeated measures would shift all three curves to the left (or right) (see Figure 2). Thus, we consider the improvements of incorporating repeated measures observed herein to underestimate those expected for detecting smaller fold-changes in molecular epidemiological studies with larger sample sizes.

In summary, miRs have great potential as reliable blood biomarkers of early effects of the disease state. In measuring miRNA expression, we concluded that variability due to the qPCR reaction replicates generally outweighs other measured sources of variability. In the future, it would be advantageous to either troubleshoot qPCR reaction conditions to reduce this variance (e.g., increase reaction volumes, modify reaction temperatures, etc.) or increase the number of replicates per sample to account for this source of variability. Previous publications on small-fold changes in blood-miR analyzed by qPCR should be viewed with skepticism in light of this finding. Additionally, methods of extraction and miR quantification must be rigorously tested to maximize yield and interpretable results, as performance can vary by kit and assay. For unavoidable sources of variability, block experimental designs or repeated measures should be implemented. Identifying sources of variability in future omics-level experimental designs and estimating power *a priori* using these described methods can save precious resources, funding, and time for molecular epidemiological studies.

Authorship Contributions

S.I.D. and R.T. conceived and designed the study and developed methodology; S.I.D., F.S., A.G. and B.Y. performed experiments and acquired data; S.I.D. and R.T. analyzed and interpreted data; S.I.D. and R.T. wrote the manuscript; F.S., L.Z. and M.T.S reviewed and edited the manuscript; R.T., L.Z., and M.T.S. supervised the study; All authors contributed to the final manuscript and approved its content.

Tables and Figures

Table 1. Summary table of experimental designs used for simulations of miR microarray data

Design	N1	N2	Number of subjects with repeated measures (%)	Number of repeated measures
0	75	75	0 (0)	0
1A	75	75	30 (20)	1
1B	75	75	30 (20)	4
2A	75	75	75 (50)	1
2B	75	75	75 (50)	4
3A	75	75	150 (100)	1
3B	75	75	150 (100)	4

Table 2. Variance terms and *p*-values for sources of variability
The variability for each random effect term in the model is reported, as well as *p*-values based on ANOVA tests for each term of each modeled miR.

	qPCR Replicate		inter-individual		Seasonal		between batches		within batch		residual	Total Var
random effect	Var	*p*-value	Var,	*p*-value	Var,	*p*-value	Var	*p*-value	Var	*p*-value	Var	
miR-130a	0.83	<0.001	0.15	0.387	0.00	1.00	0.00	1.00	0.00	1.00	0.09	1.07
miR-30d	0.54	<0.001	0.05	0.560	0.00	1.00	0.00	1.00	0.00	1.00	0.10	0.69
miR-185	0.82	<0.001	0.22	0.153	0.03	0.91	0.00	1.00	0.15	1.00	0.10	1.32
let-7d	0.44	<0.001	0.23	0.103	0.00	1.00	0.04	0.77	0.27	0.65	0.04	1.02
miR-16 (TaqMan)	0.50	<0.001	0.24	<0.05	0.05	0.71	0.04	0.89	0.00	1.00	0.11	0.94
c.eleg-39	0.14	<0.001	0.00	1.000	0.00	1.00	0.01	0.82	0.00	1.00	0.08	0.23
RNU48	0.39	<0.001	0.10	0.264	0.00	1.00	0.00	0.21	0.00	1.00	0.09	0.58
SNU6	0.38	<0.001	0.00	1.000	0.00	1.00	0.00	1.00	0.00	1.00	0.46	0.84
miR-16 (qScript)	0.12	<0.001	0.13	<0.001	0.00	1.00	0.00	1.00	0.35	<0.05	0.01	0.61
miR-451	0.31	<0.001	0.39	<0.001	0.40	<0.05	0.03	0.69	0.48	0.07	0.03	1.64

Table 3. Estimations of minimum detectable fold-changes for 2 study designs

The estimates of variability obtained from the empirical qPCR study of miR-451 and miR-185 were used to determine the minimum detectable fold-change with 80% statistical power for a theoretical study (N=75 vs 75) of two miRs. The mean fold-change, standard error (SE) and 95% CI and 90% CI are reported given no replicates (Study 1) versus a study given five extraction batches, five within-batch replicates, five time point replicates, and five qPCR replicates for 50% of the subjects (Study 2).

	miR-451		miR-185	
	Study 1	Study 2	Study 1	Study 2
mean	3.77	2.38	4.18	2.4
SE	0.95	0.53	1.2	0.67
95% CI	(1.75, 5.16)	(0.97, 3.36)	(2.67, 6.46)	(1.71, 3.74)
90% CI	(1.88, 5.11)	(1.07, 3.20)	(3.69, 6.37)	(2.11, 3.68)

Figure 1. qPCR measurements of sources of blood miRNA variability

A. Proportions of inter-individual, intra-individual, and technical variability were estimated for *N*=12 subjects using a mixed-effects model of qPCR data from seven target miRs (miR-16, miR-342-3p, miR-30d, miR-185, let7d, miR-130a, miR-451), two endogenous control small RNAs (RNU48 and snRNA U6) and one exogenous spike-in (cel-39). Technical variability includes variability within- and between-extraction batches as well as plate-to-plate variability. **B.** Interclass correlation (ICC) for each source of variability was calculated as the proportion of total variance for each miR.

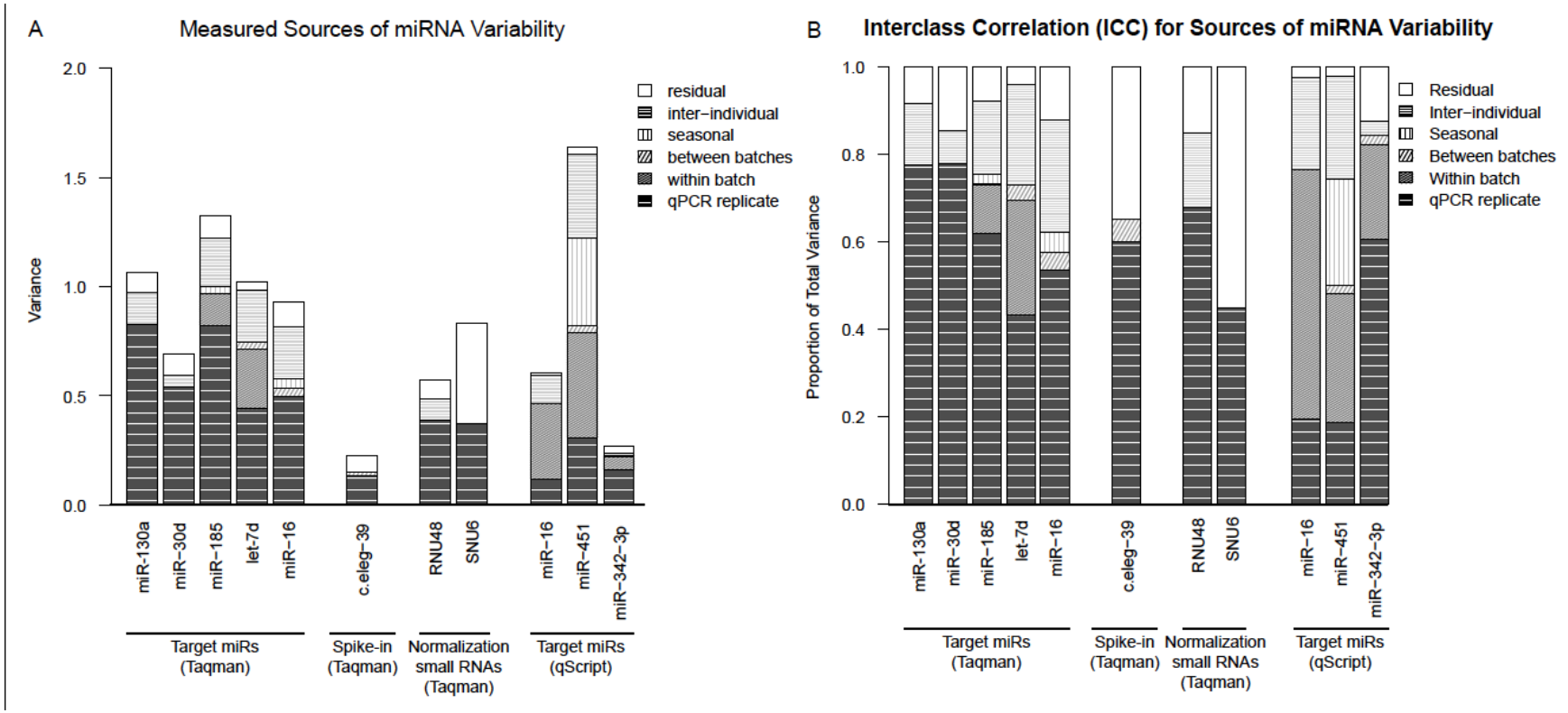


Figure 2. Cumulative distributions of minimum detectable fold-changes in miRs using repeated measures

Smallest fold-changes detected for the 143 miRNAs are plotted under seven experimental design conditions, which vary in proportion of repeated measures, A) 20%, B) 50%, C) 100%, and number of repeated measures per subject (*n*1=0, *n*1=1 or *n*1=4). Fold-changes are reported with 80% power for 100 bootstrap simulations using previously published data (25). The vertical line in each figure compares distributions at a 2-fold change

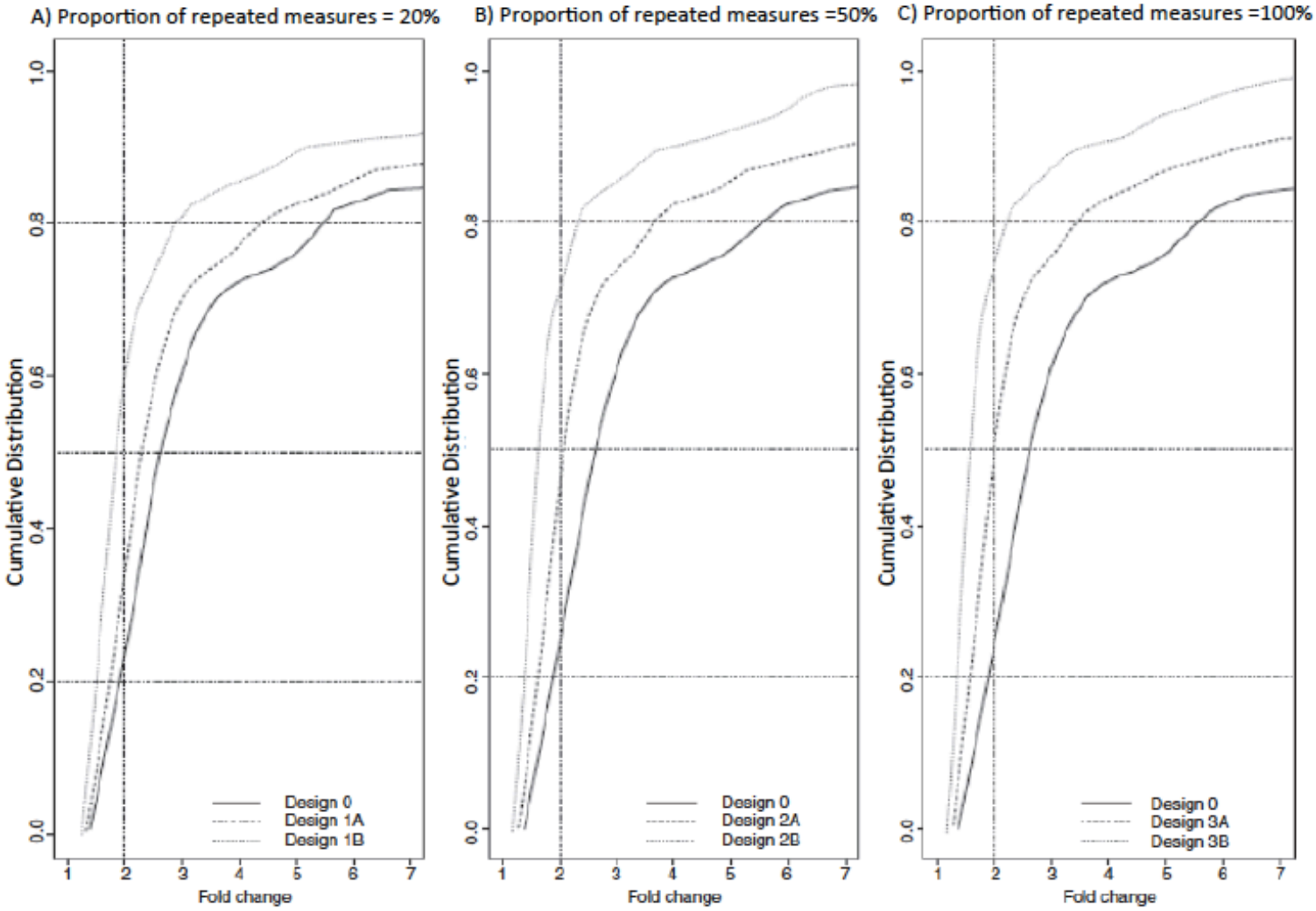
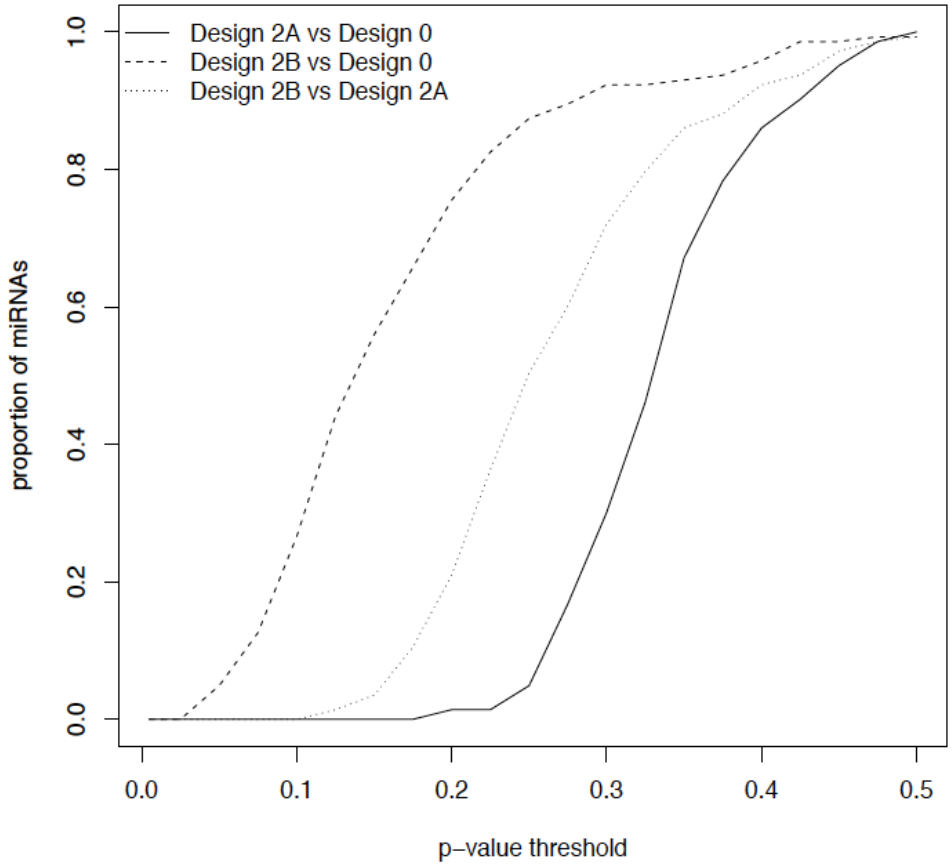


Figure 3. Comparison of 50% repeated measure designs for detection of significant fold-changes in miRs
Designs 0, 2A and 2B were compared to each other to calculate the proportion of the 143 miRNAs for which two designs' confidence intervals do not overlap at a given *p*-value. (Designs differ by number of repeated measures for each subject.)



References

1. McHale CM, Zhang L, Lan Q, Vermeulen R, Li G, Hubbard AE, et al. Global Gene Expression Profiling of a Population Exposed to a Range of Benzene Levels. Environ Health Perspect. 2010 Dec 13;119:628–34.
2. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009 Jan;19:92–105.
3. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005 Jan 14;120:15–20.
4. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, et al. An Analysis of Human MicroRNA and Disease Associations. PLoS ONE. 2008 Oct 15;3:e3420.
5. Yang Q, Qiu C, Yang J, Wu Q, Cui Q. miREnvironment Database: providing a bridge for microRNAs, environmental factors and phenotypes. Bioinformatics. 2011 Dec 1;27:3329–30.
6. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011 Jun;3:267–77.
7. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci. 2003 Feb 18;100:1896–901.
8. Radich JP, Mao M, Stepaniants S, Biery M, Castle J, Ward T, et al. Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics. 2004 Jun;83:980–8.
9. Eady JJ. Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. Physiol Genomics. 2005 May 24;22:402–11.
10. McLoughlin K, Turteltaub K, Bankaitis-Davis D, Gerren R, Siconolfi L, Storm K, et al. Limited dynamic range of immune response gene expression observed in healthy blood donors using RT-PCR. Mol Med. 2006;12:185.
11. Karlovich C, Duchateau-Nguyen G, Johnson A, McLoughlin P, Navarro M, Fleurbaey C, et al. A longitudinal study of gene expression in healthy individuals. BMC Med Genomics. 2009 Jun 7;2:33.
12. Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, et al. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA. 2010 May;16:991–1006.
13. Redshaw N, Wilkes T, Whale A, Cowen S, Huggett J, Foy CA. A comparison of miRNA isolation and RT-qPCR technologies and their effects on quantification accuracy and repeatability. BioTechniques. 2013 Mar;54:155–64.
14. Ach RA, Wang H, Curry B. Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. BMC Biotechnol. 2008;8:69.

15. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, et al. Blood Cell Origin of Circulating MicroRNAs: A Cautionary Note for Cancer Biomarker Studies. Cancer Prev Res (Phila Pa). 2012 Mar 1;5:492–7.
16. Cheng HH, Yi HS, Kim Y, Kroh EM, Chien JW, Eaton KD, et al. Plasma Processing Conditions Substantially Influence Circulating microRNA Biomarker Levels. PLoS ONE. 2013 Jun 7;8:e64795.
17. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, et al. Haemolysis during Sample Preparation Alters microRNA Content of Plasma. PLoS ONE. 2011 Sep 1;6:e24145.
18. Kirschner MB, Edelman JJB, Kao SC-H, Vallely MP, van Zandwijk N, Reid G. The Impact of Hemolysis on Cell-Free microRNA Biomarkers. Front Genet [Internet]. 2013 May 24 [cited 2014 May 12];4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3663194/
19. Mraz M, Malinova K, Mayer J, Pospisilova S. MicroRNA isolation and stability in stored RNA samples. Biochem Biophys Res Commun. 2009 Dec 4;390:1–4.
20. Masotti A, Caputo V, Da Sacco L, Pizzuti A, Dallapiccola B, Bottazzo GF. Quantification of Small Non-Coding RNAs Allows an Accurate Comparison of miRNA Expression Profiles. J Biomed Biotechnol [Internet]. 2009 [cited 2014 May 12];2009. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2735750/
21. Eikmans M, Rekers NV, Anholts JDH, Heidt S, Claas FHJ. Blood cell mRNAs and microRNAs: optimized protocols for extraction and preservation. Blood. 2013 Mar 14;121:e81–89.
22. Eldh M, Lötvall J, Malmhäll C, Ekström K. Importance of RNA isolation methods for analysis of exosomal RNA: evaluation of different methods. Mol Immunol. 2012 Apr;50:278–86.
23. Remáková M, Škoda M, Faustová M, Vencovský J, Novota P. Validation of RNA extraction procedures focused on micro RNA expression analysis. Folia Biol (Praha). 2013;59:47–50.
24. Kanof ME, Smith PD, Zola H. Isolation of Whole Mononuclear Cells from Peripheral Blood and Cord Blood. Current Protocols in Immunology [Internet]. John Wiley & Sons, Inc.; 2001 [cited 2013 Dec 5]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/0471142735.im0701s19/abstract
25. Honda M, Kuwano Y, Katsuura-Kamano S, Kamezaki Y, Fujita K, Akaike Y, et al. Chronic Academic Stress Increases a Group of microRNAs in Peripheral Blood. PloS One. 2013;8:e75960.
26. Karolina DS, Armugam A, Tavintharan S, Wong MTK, Lim SC, Sum CF, et al. MicroRNA 144 Impairs Insulin Signaling by Inhibiting the Expression of Insulin Receptor Substrate 1 in Type 2 Diabetes Mellitus. PLoS ONE [Internet]. 2011 Aug 1 [cited 2012 Nov 6];6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3148231/
27. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7:177–88.
28. Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol. 2006;411:352–69.
29. Ballman KV, Grill DE, Oberg AL, Therneau TM. Faster cyclic loess: normalizing RNA arrays via linear models. Bioinforma Oxf Engl. 2004 Nov 1;20:2778–86.

30. Mestdagh P, Hartmann N, Baeriswyl L, Andreasen D, Bernard N, Chen C, et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. Nat Methods. 2014 Aug;11:809–15.

Supplementary Materials

Figure S1. Comparison of miRNA Extraction kits by qPCR

A) Four protocols, miRNeasy, miRVana (MirV), Trizol (Tri) and Zymogen, were compared to determine the method with the best miRNA extraction yield from PBMCs stored in RNAProtect. The miRNA yield for the miRNeasy kit is used as the baseline comparison. We measured five target miRNAs (miR-16, miR-142-3p, miR-9, let-7a*, miR-370), two 'housekeeping' small RNAs (RNU48 and snRNA U6), and two synthetic miRNA spike-ins (cel-39 and cel-54). Extraction methods were performed 1-2 times per set of samples. The Zymogen kit (Zymo) was not tested for miR-9 and let-7a*. B) The miRNA obtained from the PBMC samples using the AllPrep Kit was compared to miRNA from the miRNeasy kit also using qPCR analysis. The miRNA yield for the miRNeasy kit is used as the baseline for comparison. Experiments were performed three times on three aliquots of two PBMC samples. We measured two control small RNAs (RNU48 and snRNA U6), one high expressed miR (miR-16), and two spike-in oligos (cel-39 and cel-54).

A) B)

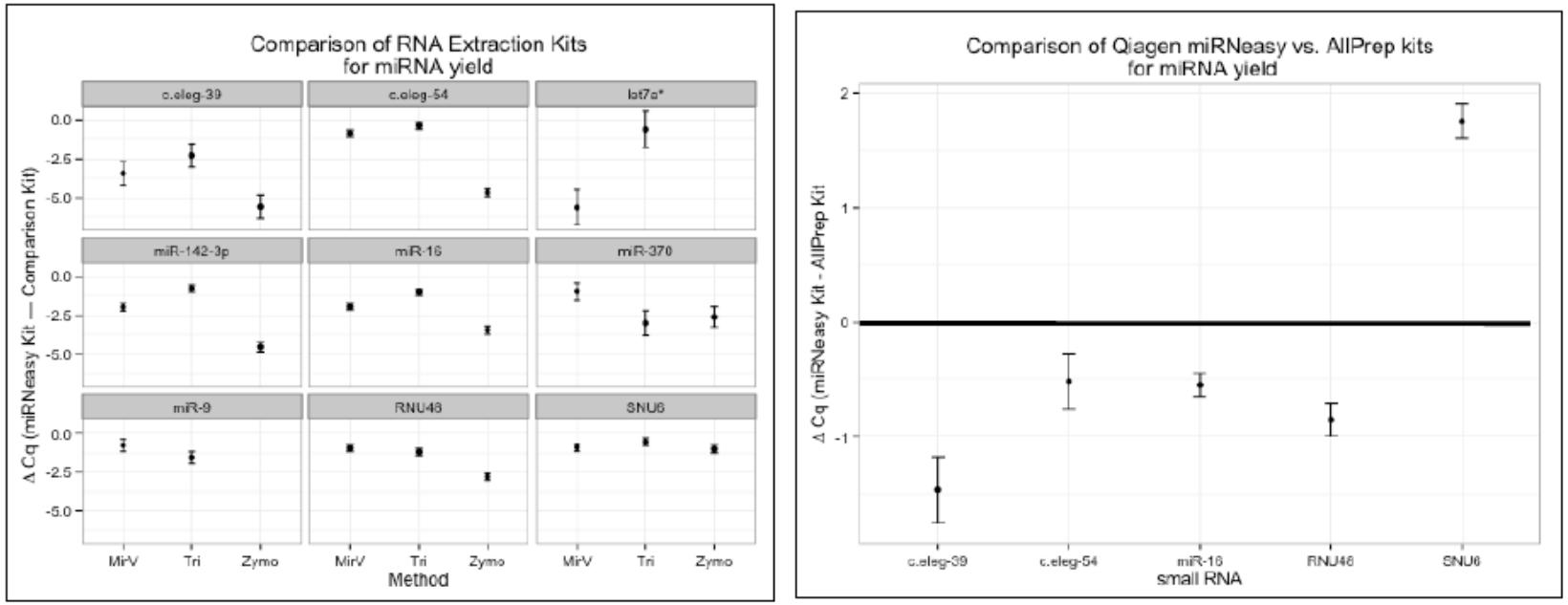
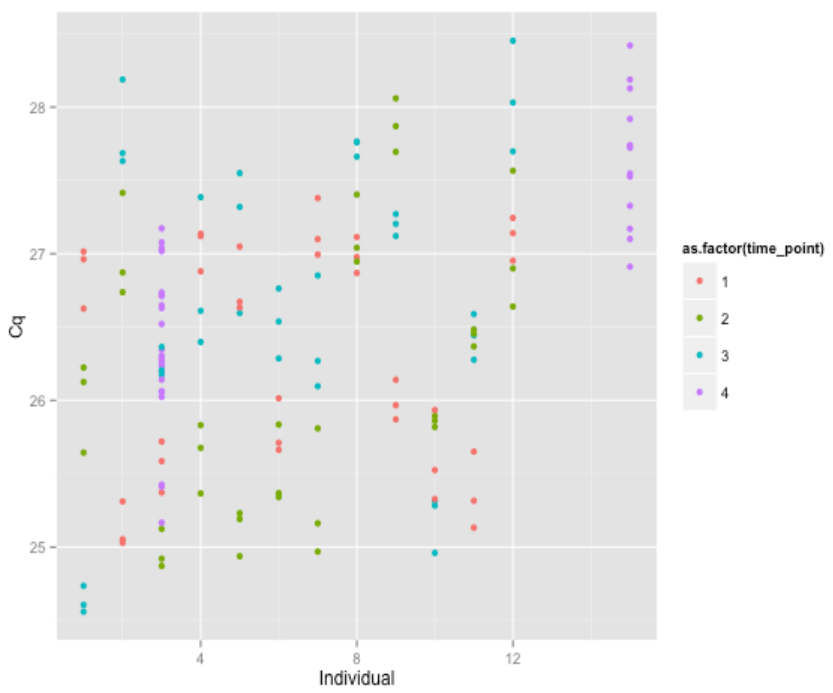


Figure S2. Comparison of miR qPCR assays for miR-16: TaqMan vs. SYBR Green

The Cq values for the miR-16 are plotted for each of the 12 individuals at three time points for A) the TaqMan assay and B) the qScript (SYBR Green) assay. The qPCR reactions were completed in triplicate for each of the samples tested. Individuals with extractions at a fourth time point represent extraction replicates.

A) miR-16 TaqMan



B) miR-16 Syber Green

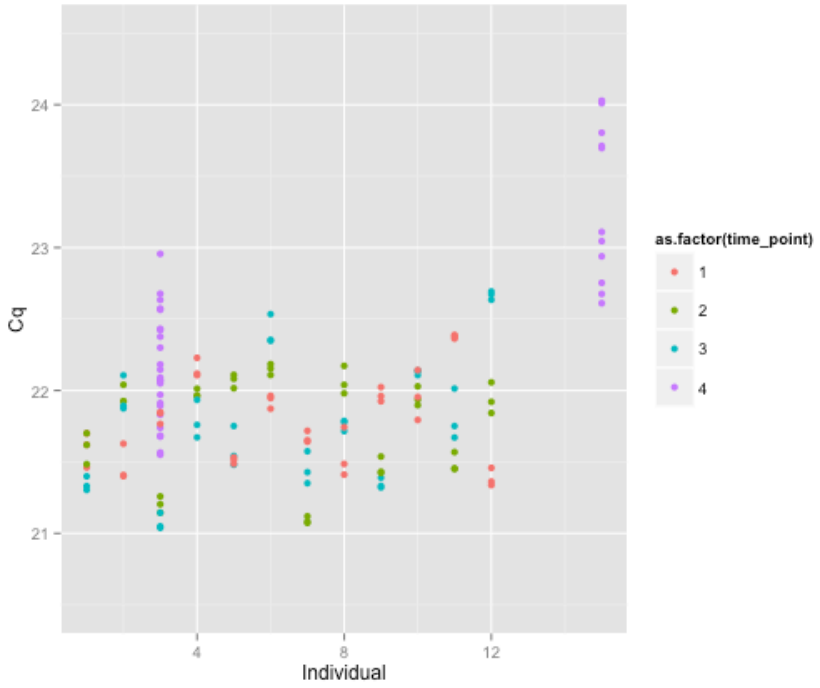


Figure S3. Confidence intervals for minimum detectable fold-changes in miRs

The 95% confidence intervals for minimum detectable mean fold-changes in miRs from the bootstrap simulation of miR microarray data are plotted. These designs vary in proportion of repeated measures A) 20%, B) 50%, C) 100%, and number of repeated measures per subject (n1=0, n1=1 or n1=4).

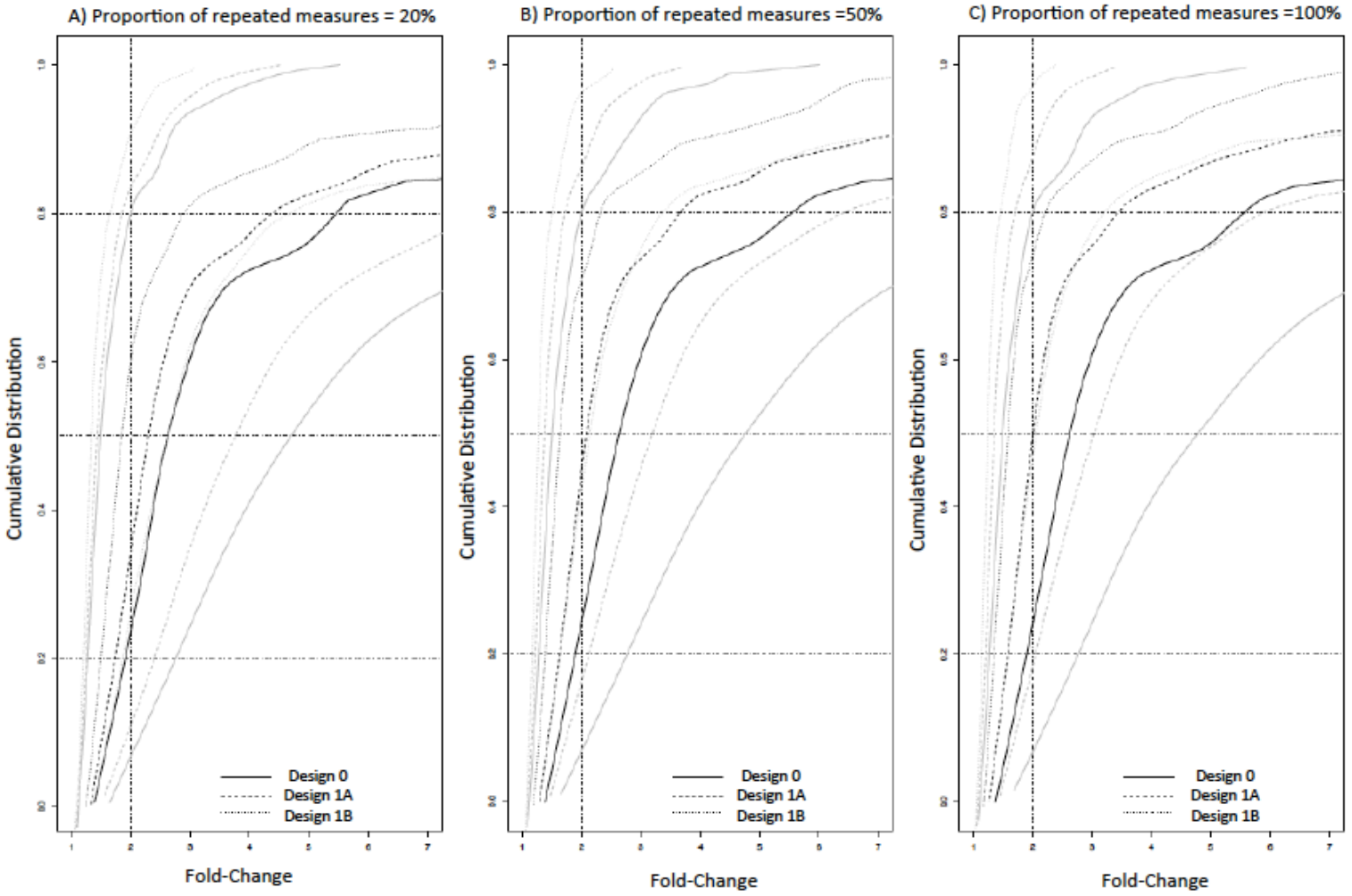
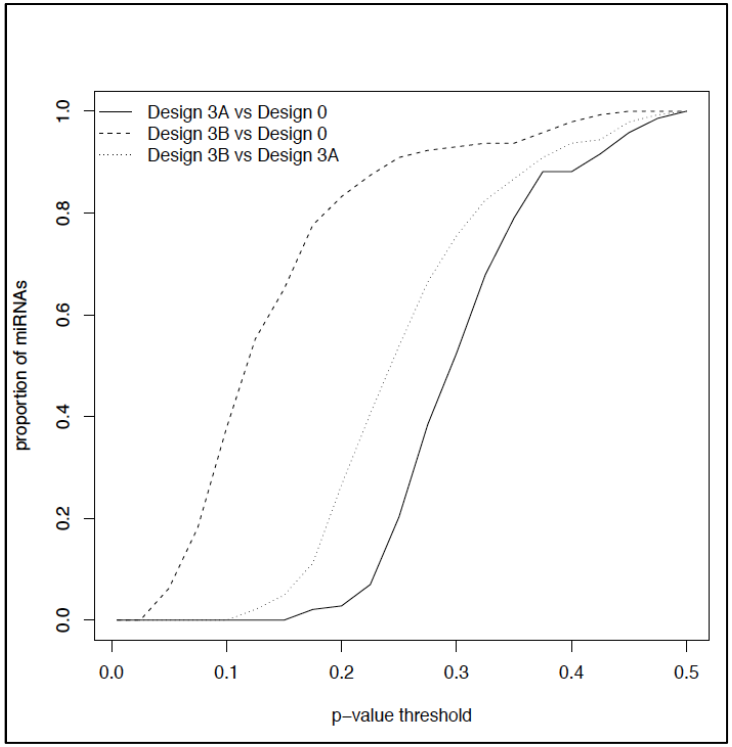


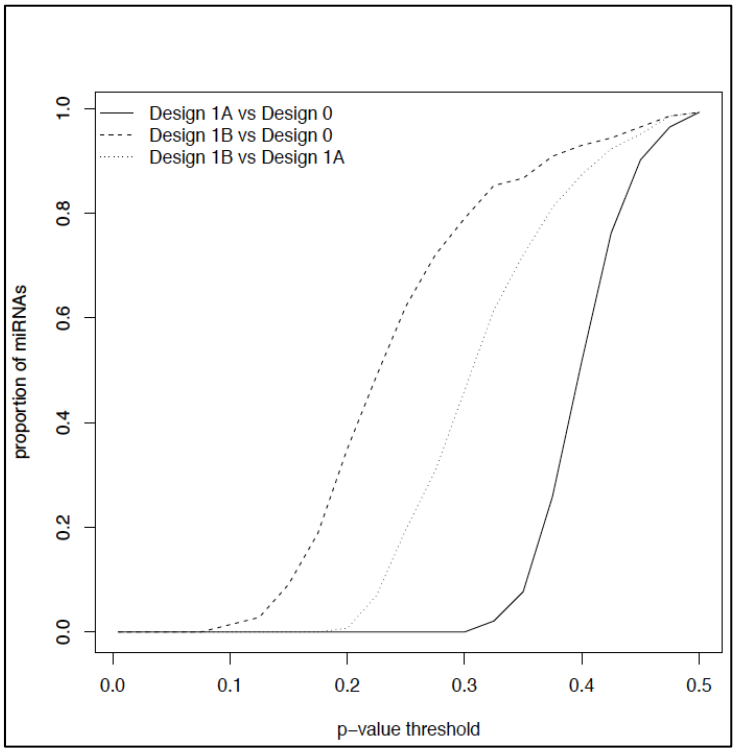
Figure S4. Comparison of repeated measure designs for detection of significant fold-changes in miRs

A) Designs 0 vs. 1A vs. 1B for 20% repeated measures and B) Design 0 vs. 3A vs. 3B for 100% repeated measures were compared to each other to calculate the proportion of the 143 miRNAs for which two designs' confidence intervals do not overlap at given p-values.

A)



B)



Supplementary Methods

miRNA Extraction Procedures

Several miRs were measured by real-time PCR to determine which kit gave the highest quality and quantity of miR yield following protocols for each given kit. This experiment was completed with four human PBMC samples preserved in RNAprotect. Samples were thawed, distributed into 4 equal aliquots and spun at 5000 x *g*. The RNA preserving-reagent was aspirated off and RNA was extracted from the pellet using one of the four kits. For the volunteer study of intra-, inter- and technical variability, 36 experimental samples were extracted in two batches on two separate days. Two extra PBMC 'reference' samples were included in the extraction of both batches to account for between-batch variability and extracted in duplicate on each day to account for within-extraction batch variability. The concentration of the RNA eluate for all samples was measured by Nanodrop analysis (Nanodrop / Thermo Scientific, Wilmington, DE) and confirmed by Quant-iT RiboGreen RNA Assay (LifeTechnologies, Grand Island, NY).

The miRNeasy kit was compared to a slightly modified version of the AllPrep DNA/RNA kit, to determine miR yield-lost during the collection of both DNA and RNA with on-column kits. Between 25-250nmoles of *c.eleg*-39 and 25nmoles of *c.eleg*-54 were added to each sample upon addition of the cell lysis reagent. The AllPrep kit protocol was slightly altered for elution of both small and total RNA. (Upon obtaining eluate from the DNA spin column, 1.5 volumes of 96%-100% ethanol was added before transfer to the RNeasy Mini spin column, 500µL of RPE was used to wash the spin column twice, and then an additional 500µL of 96%-100% ethanol was added to the spin column before the 2 minutes spin at full speed.)

miRNA Quantification by qPCR

TaqMan Assays were completed using proportionally smaller volumes (5uL reactions volumes) for both the RT and PCR step as described previously (1). Total RNA input for each reaction ranged from 2-10µg. miR assays were tested for efficiency of >70% in the qPCR step using cDNA dilutions before using the assays for this study. To evaluate very low-expressed miRs, miR-370 and let-7a*, were used in spite of efficiency <70%. Two of the six TaqMan Assays showed high levels of background signal with a "no reverse transcriptase" control in the reverse-transcription step. Thus, a competing miR quantification assay, qScript-PerfeCTa microRNA Assays (Quanta Biosciences), was used to measure miR-451 and miR-342-3p. (5 µL reactions for the poly A tail synthesis reaction, 10 µL for the RT reactions, and 15µL volume for the PCR reactions). For all qPCR experiments, three technical replicates were used for each sample on a 96-well plate, and two separate reference samples were run on each plate to calibrate for plate-to-plate differences. Upon obtaining qPCR results, Ct value thresholds for each plate were determined by setting the 25^{th}-75^{th} quartile distributions of the signal from each plate equivalent to each other before continuing with the downstream analysis.

Statistical Analysis of qPCR results for extraction kit comparison

A linear mixed model was used to compare the differences between the 4 miRNA extraction kits, or the two Qiagen miRNA extraction kit methods.

$$Y_{ijk}^{m} = \beta_0{}^{m} + \beta_{0i}{}^{m} + \sum_{j=1}^{N_{extraction}-1} \beta_j{}^{m} I\left(j == k\right) + \varepsilon_{ijk}{}^{m}$$

$$I\left(j == k\right) = 1 \quad \text{if} \quad j = k \qquad (3)$$

$$= 0 \quad \text{otherwise}$$

Here Y_{ijk}^{m} represents the k^{th} (k=1,2) replicate *Cq* value for the m^{th} (m=1,2,...9) miR in the i^{th} (i=1,2,3,4) individual using the j^{th} (j=0, 1,..., $N_{extraction}-1$) extraction kit. $\beta_0{}^{m}$ represents the mean *Cq* value for the m^{th} miR using the miRNeasy kit. $\beta_j{}^{m}$ represents the change in the mean *Cq* value for the m^{th} miR using the j^{th} extraction method versus the miRNeasy kit. $\beta_{0i}{}^{m}$ represents the random effect terms that capture the change in the mean *Cq* level of the m^{th} miR of the i^{th} individual compared to the mean *Cq* value of this miR across all individuals. $\varepsilon_{ijk}{}^{m}$ represents the residual for this model.

References

1. Kroh EM, Parkin RK, Mitchell PS, Tewari M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). Methods. 2010 Apr;50:298–301.

Chapter 4: Validation of Circulating miRNA Profiling for Type II Diabetes in Asian Indians

Sarah I. Daniels[1], Brenda Yee[1], Kevin Zhang[1], Ellen F. Key[1], Sylvia Sanchez[1], John Chambers[2], Jaspal Kooner[2], Paul Elliott[2], Luoping Zhang[1], Martyn T. Smith[1]

(Authorship order to be determined)
Associated Institutions:
[1]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California
[2]Department of Epidemiology and Biostatistics, Imperial College London, UK; MRC-PHE Centre for Environment and Health, Imperial College London, London, UK; Ealing Hospital NHS Trust, Middlesex, UK; Imperial College Healthcare NHS Trust, London, UK

ABSTRACT

Circulating blood miRNAs hold potential as new biomarkers of type II diabetes (T2D), as specific miRNAs in the blood milieu may be secreted from tissues directly affected by T2D. However, in the cancer field, it has been shown that some of the proposed tumor-related circulating miRNAs were actually originating from the patients' peripheral blood cells. In order to determine whether peripheral blood cells confounded the association between miRNA expression and T2D, we attempted to validate 20 candidate blood miRNAs previously reported as differentially expressed in T2D cases versus healthy controls. After adjusting for technical sources of variability, we initially observed higher expression of 17/20 miRNAs in cases versus control subjects. However, lymphocyte counts were also significantly higher in the diseased subjects and positively associated with most miRNA expression levels. After adjusting for lymphocyte counts in the regression models, only three miRNAs remained differentially expressed and were not associated with lymphocyte counts. While a majority of the T2D circulating miRNAs previously reported may be due to immunological changes related to the disease process, at least 3 miRNAs, miR-24, miR-423, and miR-375, are potentially useful biomarkers and deserve further evaluation as signaling molecules in the development of T2D.

Introduction

Circulating blood microRNAs (miRNAs) are of recent interest in the field of type II diabetes (T2D). MiRNAs are small (19-23 nucleotides) non-coding RNAs, that bind to the 3'-untranslated region of mRNAs and target mRNAs for degradation or inhibition of translation, thus regulating 30-60% of mRNA transcription[1,2]. In human subject samples, miRNA are extremely stable in comparison to mRNA and protein[3–5], and can be measured using various techniques. Altered expression of miRNA profiles in whole blood, white-blood cell subtypes, serum, and plasma have been linked to type II diabetes (T2D) onset, progression and other metabolic disease state(s) [6–13]. Circulating miRNAs in the extracellular blood mileu are of particular interest as they may act as signaling molecules for cell-to-cell communication between T2D target tissues such as the liver, pancreas, and adipose.

Studies in separate study populations have reported similar findings of specific circulating blood miRNA related to T2D (summarized in Table 1). For example, three separate studies demonstrated that miR-29a, associated with hyperlipidemia, is more highly expressed in blood samples of diabetics versus control subjects. Yet, the directionality of the association can also differ between studies. One large study (N>800) found miR-126 significantly down-regulated 5-years prior to T2D and negatively associated with fasting-glucose status [6]; One large study (N>800) found miR-126 significantly down-regulated 5-years prior to T2D and negatively associated with fasting-glucose status [6]. miR-126 was observed to be down-regulated in two additional human studies[6,7], and up-regulated in a third[14]. Expression of miR-320a, involved with insulin signaling pathways was shown to be increased in two studies[7,9] and decreased in a third study[6]. Differences in sampling media (plasma vs. serum vs. whole blood), normalization methods (to account for technical variability), and baseline expression levels among different populations [12,15] may also contribute to the lack of consistency between these studies.

Changes in blood miRNA profiles related to chronic disease outcomes may also be due to the underlying source of the circulating miRNA. In the field of cancer, circulating miRNA that were deemed as potential biomarkers from the tumor tissue of interest were later found to be derived from the cancer patients' peripheral blood cells[16]. Moreover, explicit warnings have been made in the cancer field, to take heed in making generalizations about the origin of circulating miRNAs. Other contributing factors, such as hemolysis during the blood collection process [17–19], may also inadvertently bias the circulating miRNA profiles. Therefore, it is crucial to take these considerations into account when analyzing measurements of circulating blood miRNAs studies.

Here, we use a T2D case-control study on Asian Indians to exemplify how to control for biological and technical variability when assessing circulating blood miRNAs. Asian Indians have 2-3-fold greater risk of type II diabetes (T2D) than whites [20–22], thus determining biomarkers before disease manifestation could curb T2D progression in this high-risk population. We attempt to validate some of the circulating blood miRNAs associated with T2D that were previously reported in the literature. We consider peripheral blood cell counts as important sources of confounding and provide analytical methods for comparing relative expression levels that account for variability during RNA extraction, improving upon the delta-delta-Ct convention.

Methods

Ethical Statement

A. Population and Biospecimen collection

The London Life Sciences Prospective Population Study (abbreviated as "LOLIPOP") was established in 2002 and is a prospective cohort comprised of South Asian Indians and European whites living in West London. Subjects that enter into the cohort are followed-up annually, which includes a detailed health assessment and collection of blood and urine samples from subjects at each visit. Adult (>21 years) volunteers, specifically of Telugu or Sri Lankan Tamil descent, were newly recruited for the LOLIPOP study in 2012. As is protocol for the LOLIPOP study, all participants were subject to a full health physical that included physiological, clinical and blood measurements. Survey data and demographic information, including family history, current occupation, time of residence in England, and smoking, drinking and exercise habits, was collected by a registered nurse. T2D cases were defined as fasting-blood glucose ≥7mmol/L; a more stringent cut-off for abnormal blood-glucose levels was defined as HbA1c >5.6%.

B. miRNA extraction

RNA was extracted from 200uLs of plasma from 48 subject samples. Plasma samples were thawed on ice and spun at 5000x*g* for 5 minutes at 4°C prior to obtaining 200uL volumes for extraction using the miRNeasy kit (Qiagen), a chloroform-phenol and filter-based extraction procedure. The following modifications were made for plasma RNA extraction; 1) Two spike-in miRs (*cel*-39 and *cel*-54) were used as exogenous controls for technical variability in extraction methods 2) larger volumes of Qiazol lysis reagent (1:5, plasma:Qiazol), 3) proportionately more chloroform was added to obtain phase separation, and 4) proportionately more ethanol added to the aqueous phase before loading onto the filter. To account for batch effects, two reference plasma samples were extracted side-by-side with each batch of subject samples. The RNA quantity eluted in PCR-grade water was quantified by Nanodrop Spectrometer.

C. miRNA measurement (qPCR)

The specific candidate miRNAs were chosen based on previous results summarized in Table 1, along with negative control miRNAs that were not expected to differ between cases and controls. The qScript-PerfeCTa microRNA Assays (Quanta Biosciences) was used to measure all candidate miRNAs with the company's protocol (10 µL reactions for the poly A tail synthesis reaction, 20 µL for the RT reactions, and 50µL volume for the PCR reactions). Most qScript assays were tested with dilution series of RNA extracted from plasma to ensure >80% efficiency for reactions of this sample type. For each qPCR experiment for each target miRNA, all samples were loaded onto a 96-well plates and measured by BioRad CFX96 RT-qPCR instrument. Plasma samples were tested for hemolysis contamination as done previously [12], and the difference in Cq values between miR-451 and miR-23a was below the cut-off for hemolysis (<7-8) in the chosen samples [17].

D. Analytical Methods

Raw Cq threshold values obtained from RT-PCR were used for analysis of all assays with a cycle number (Cq value) <35 for all subject samples. No stable miRNA could be determined for purposes of normalization, as occurred before in a previous study in a similar population [13]. In addition, normalization methods that only use synthetic spike-in miRNAs have been found to be unrepresentative of differences also seen in extraction variability of endogenous miRNA (reviewed in [23]). Therefore, in order to effectively examine associations between miRNA expression and case-control status, linear regression was used to model the relationship between diabetes status (predictor variables) and miRNA expression (dependent variable), as the miRNA expression signal was normally distributed (Shapiro-Wilk test $p>0.05$)[*]. Regression models were adjusted for technical variability (to account for extraction effects measured by batch and spike in *c.elegans* miRNA levels) and some sources of biological variability (the subject's age and lymphocyte count). The more rigorous false discovery rate (FDR) was applied to correct for multiple-hypothesis testing (20 miRNAs) in these experiments ($q<0.05$). Data analysis was conducted in R statistical software.

Results

In this study we measured 20 circulating miRNAs from 31 cases versus 17 healthy controls, with cases determined by HbA1c>5.6%. The HbA1c cut-off was ultimately used for the main analysis, because some subjects had high HbA1c values but low fasting-glucose levels. Without definitive diagnosis from a physician, we presumed that HbA1c is a long-term marker of diabetes that may capture a greater proportion of total diabetic and pre-diabetic cases. The biometric and demographic characteristics were fairly similar between the two comparison groups except for LDL concentrations, lymphocyte counts, and basophil counts (Table 2). Similar comparisons were observed when using fasting blood glucose concentration cut-off (≥7 mmol/L), except differences were also observed in age between cases versus controls (Table S1) (and was included as a covariate in the analyses).

Correlation of miRNAs with each other and with differential white blood cell counts

Paired correlations of Cq values (adjusted for technical variability) were calculated for all miRNA targets to determine similarities in expression patterns across plasma miRNAs. The color-coded matrix (Figure 1) shows the high-degree ($r>0.8$) of correlation between most of the diabetes-related miRNA markers. Differential white blood cell counts were also included in the correlation matrix to determine if any miRNAs correlated with these counts. Many of the miRNAs showed a direct linear relationship with lymphocyte counts, including miR-29a, miR-423, miR-23a, miR-146a, miR-425, miR-130b, miR-191, miR-16, miR-27a, miR-320a, let-7d, miR-374, miR-155, miR-223, miR-24. Several miRNAs that are not specific to white blood cells, including miR-451 (red blood cells), miR-122 (liver), and miR-375 (pancreas), did not correlate well with either the clustered miRNAs nor the differential white blood cell counts.

[*] An exception: miR-375 did not initially pass the Shapiro-Wilk test. Analyses were repeated after removal of three outliers and resulting effect size and p-value were similar to the reported values here.

Regression models of miRNA expression in cases versus controls

Linear regression models were used to identify miRNAs that were differentially expressed between cases and controls. All results here are interpreted with caution, as the total sample size, *N,* was relatively small. In the first set of models, they were only adjusted for technical variability and age of the subjects. The fold-change differences in these models showed a significant increase in a majority of the miRNAs (Table 3). Then, an additional covariate for lymphocyte count was added to the model, as these particular white blood cell subsets were associated with both the miRNA expression levels and the T2D disease state in our subjects. In this lymphocyte- adjusted model, there was only seven miRNAs that remained significant. Of these, only miR-375, miR-423 and miR-24 had no significant association with lymphocyte count ($p>0.05$). While miR-375 had the greatest measure of effect (>2-fold), none of the miRNAs withstood correction for multiple hypothesis testing (FDR<0.05). The relative expression values adjusting for technical and biological variability are plotted for cases versus controls for miR-375, miR-423, and miR-24 (Figure 2). Similar patterns in effect measures are seen when using the fasting-glucose cut-off of ≥7mmol/L for cases versus controls (comparing N=24 vs. N=24), however miR-375 and miR-122 are the only significant miRNAs after adjustment for lymphocyte count and only miR-375 withstood correction for multiple hypothesis testing (FDR<0.05). (Table S2 and corresponding Supplemental Figure S1).

Discussion

While studies on circulating miRNAs have increased in popularity in recent years, little attention has been given to the true source of these extracellular miRNA expression levels. Several reports have suggested that proposed disease-related blood miRNAs may simply be acting as proxys for a third unmeasured variable in the blood, including blood cells[16], platelets[24], and lipids[25]. In the field of cancer, white blood cell counts were correlated with circulating miRNAs that were considered to be candidate cancer biomarkers in cancer patients[16]. Peripheral immune cell counts increase not only in cancer cases, but in other chronic illnesses, such as T2D, as well. Positive associations have routinely been found between white blood cell counts and diabetes status, even prior to T2D onset (in prospective studies)[26]. We were able to confirm that, while a majority of target circulating miRNAs were differentially expressed between cases and controls, many of these miRNAs are highly correlated with lymphocyte counts. This influenced the measure of association between circulating miRNA expression and T2D status, as many of the target miRNAs in this study were no longer significant after adjusting for lymphocyte counts. As circulating miRNA will continue to be considered viable blood biomarkers, differential white blood cell counts must be accounted for in the experimental design for studies of chronic illnesses.

In this case-control study of Asian Indians, we confirmed differential expression of various circulating plasma miRNAs that were previously reported in other T2D-related population studies. In Asian Indians specifically, increased expression of miR-128, miR-374a, miR-130b was found in subjects with diabetes [13] and decreased expression of miR-423, miR-122, miR-15a, miR-197, miR-320a, miR-486 was found in subjects progressing to pre-diabetes [27]. We observe that, after adjusting for technical variability, 17/20 of the miRNAs in our study (including miR-320a, miR-128, miR-423) were initially found to be associated with T2D status. However, 14/20

of these T2D-associated miRNAs were likely confounded by the correlation with white blood cell counts, namely lymphocytes. This also helps to explain why many of the 20 miRNAs tested were so highly correlated with each other (r>0.8). While some of these miRNAs may still be involved in disease etiology, currently they appear to be representative of changes in lymphocyte counts and, as such, are relatively non-specific and insensitive biomarkers.

Three miRNAs, miR-375, miR-423, and miR-24 were not associated with lymphocyte counts but were associated with glycemic impairment. Interestingly, miR-375 also had the largest measure of effect, a 2-fold difference in expression between cases and controls. Pancreatic beta-islet cells highly express miR-375, and knock-out mouse models have shown severe hyperglycemia and reduced pancreatic cell proliferation in the absence of miR-375 [28]. Conversely, increased miR-375 expression is demonstrative of beta-islet cell death in a mouse model [29]. T2D cases in human epidemiological studies have shown increased miR-375 expression levels in Chinese populations [8,30]. Previous studies on Asian Indians have not identified miR-375 as a differentially expressed circulating miRNA, citing expression levels below the detection limit in one study [27]. This miR was low expressed in our study as well, and may imply the source of secretion is pancreatic tissue specific and not due to changes in lymphocyte counts. Circulating miR-423, previously-reported as a cardiac-specific miRNA in several distinct populations [31–34], was more highly expressed in T2D cases of our sample population, however, previous studies on Asian Indians with glycemic impairment reported decreases miR-423 levels [13,27]. Similarly, miR-24 is also expressed by heart cells (cardiomyocytes) yet is involved with angiogenesis and inhibition of apoptosis [35–37]. miR-24 was previously shown to be differentially expressed in two other human studies on T2D as well [6,12]. These three miRNA are more robust biomarkers of T2D status, and more functional studies are warranted to determine their role in T2D onset and progression.

Acknowledgements

We appreciate the productive discussions with Davide Risso, Kelsi Perttula, and Fenna Sille about our experimentation, statistical analysis, and interpretation of our results. We thank Audrey Goldbaum and Justin Kim for their exploratory work on plasma miRNAs that supported this paper.

Tables and Figures

Table 1. Review of literature on circulating miRNA expression level differences between T2D cases and controls

miR	Study	Year	Fold Change (±SD)	Compartment	Adj OR (95%CI)	Significant
130a	Karolina et al	2012	-1.37 ±0.01	blood		Yes
	Karolina et al	2011	1.94	blood		Yes
130b	Karolina et al	2011	2.00	blood		Yes
	Prabu et al	2015	1.25	plasma		Yes
197	Karolina et al	2012	-1.35±0.14	blood		Yes
	Wang et al	2014		plasma	1.11(.66-1.86)	No
	Zampetaki et al	2010	-1.87	plasma	0.65 (0.50-0.80)	Yes
150	Karolina et al	2012	1.57 ±0.02	blood		Yes
	Karolina et al	2011	2.85±0.09	blood		Yes (internal validation)
	Wang et al	2014		plasma	1.41(.82-2.45)	No
	Zampetaki et al	2010	-1.43	plasma	0.18 (0.12-0.30)	Yes for Odds Ratio
192	Karolina et al	2012	1.86±0.13	blood		Yes
	Karolina et al	2011	2.34±0.08	blood		Yes (internal validation)
320a	Karolina et al	2012	2.33±0.19	blood		Yes
	Karolina et al	2011	3.61±0.16	blood		Yes (internal validation)
	Wang et al	2014		plasma	1.53 (0.87-2.70)	No
	Zampetaki et al	2010	-1.33	plasma	0.22 (0.10-0.44)	Yes
320b	Wang et al	2016	~3.0	plasma		Yes
21	Wang et al	2014		plasma	1.61 (0.93-2.80)	No
	Zampetaki et al	2010	-3.33	plasma	0.76 (0.65-0.86)	Yes
24	Wang et al	2014	1.18	plasma	2.39 (1.26-4.54)	Yes
	Zampetaki	2010	-2.5	plasma	0.58 (0.43-0.78)	Yes
15a	Wang et al	2014	1.26	plasma	2.39 (1.00-5.70)	Yes for Swedes
	Zampetaki et al	2010	-6.66	plasma	0.53 (0.41-0.66)	Yes
	Zhang et al	2013	~-4.0	plasma		No
15b	Pescador et al	2013	~2.0	serum		No

miR	Study	Year	Fold Change (±SD)	Compartment	Adj OR (95%CI)	Significant
126	Karolina et al	2011	1.51	blood		Yes
	Wang et al	2014		plasma	1.40 (0.79-2.48)	No
	Zampetaki et al	2010	-1.67	plasma	0.39(0.22-0.51)	Yes
	Zhang et al	2013	~-6.0	plasma		Yes
191	Wang et al	2014		plasma	0.99 (.57-1.72)	No
	Zampetaki	2010	-2.0	plasma	0.58(0.42-0.77)	Yes
223	Wang et al	2014		plasma	1.19(0.66-2.14)	No
	Zampetaki et al	2010	-2.85	plasma	0.42 (0.3-0.6)	Yes for Odds Ratio
	Zhang et al	2013	~-1.25	plasma		No
486	Wang et al	2014		plasma	1.29 (.75-2.23)	No
	Zampetaki	2010	-1.43	plasma	0.2 (.16-.32)	Yes for Odds Ratio
28	Wang et al	2014		plasma	1.19 (0.70-2.02)	No
	Zampetaki	2010	1.25	plasma	1.25 (1.01-1.60)	Yes for Odds Ratio
	Zhang et al	2013	N/A	plasma		No
146a	Karolina et al	2011	-3.38±0.13	blood		Yes (internal validation)
	Kong et al	2011	~4.0	serum		Yes
	Zampetaki et al	2010		plasma		No
30d	Karolina et al	2011	-1.38±0.10	blood		Yes (internal validation)
	Kong et al	2011	~4.0	serum		No
144	Karolina et al	2011	3.07±0.13	blood		Yes (internal validation)
	Wang et al	2014	1.58	plasma	2.43(1.07-5.55)	Yes for Swedes only

miR	Study	Year	Fold Change (±SD)	Compartment	Adj OR (95%CI)	Significant
29a	Karolina et al	2011	2.09±0.14	blood		Yes (internal validation)
	Kong et al	2011	~8.0	serum		No
	Zhao (GDM)		3.90	serum		
29b	Karolina et al	2011	2.38	blood		Yes
	Wang et al	2014	0.86	plasma	1.93(1.11-3.36)	Yes
	Zampetaki et al	2010	1.54	plasma	0.95 (0.85-1.05)	No
	Zhang et al	2013	N/A	plasma		No
375	Karolina et al	2011	2.00	blood		Yes
	Kong et al	2011	~8.00	serum		Yes
27a	Karolina et al	2012	2.53 ±0.17	blood		Yes
	Karolina et al	2011	2.46	blood		Yes
23a	Karolina et al	2012	-1.17 ±0.07	blood		Yes
	Karolina et al	2011	1.86	blood		Yes

Table 2. Subject Characteristics for Asian Indian cases of T2D or pre-diabetes (defined by HbA1c>5.6%) and Controls

	Controls N=17	Cases N=31	p- value
HbA1c (SD)	5.34 (0.23)	7.67(1.50)	<0.001
Glucose (SD)	4.67 (0.18)	8.03 (1.63)	<0.001
Sex Male (%)	11 (64%)	22 (71%)	0.70
Age (SD)	49.57(8.13)	54.16 (9.76)	0.10
Current Smokers (%)	2 (12%)	4 (13%)	0.99
BMI (SD)	27.14 (3.56)	27.22 (3.66)	0.94
WHR (SD)	0.94 (0.06)	0.96 (0.07)	0.29
SBP (SD)	126.49 (15.29)	129.92 (15.72)	0.48
DBP (SD)	79.26 (12.12)	80.39 (7.90)	0.75
HDL (SD)	1.29 (0.28)	1.23 (0.36)	0.49
TG (SD)	1.58(0.70)	1.82 (0.94)	0.25
Chol (SD)	5.14 (1.05)	4.55 (1.32)	0.09
LDL (SD)	**3.13 (0.86)**	**2.49(1.07)**	**0.03**
WBC (SD)	6.51 (2.15)	7.20 (2.12)	0.29
Lymphocyte cnt (SD)	**1.80 (0.50)**	**2.30 (0.69)**	**0.007**
Monocyte cnt (SD)	0.43 (0.18)	0.51 (0.17)	0.16
Neutrophil cnt (SD)	4.00 (1.59)	4.14(1.68)	0.781
Basophil cnt (SD)	**0.019 (0.01)**	**0.032 (0.02)**	**<0.001**
Eosinophil cnt (SD)	0.25 (0.24)	0.23(0.15)	0.72

Table 3. Fold-Change Differences in miR Target expression levels for cases of T2D or pre-diabetes (defined by HbA1c>5.6%) versus controls, adjusted for fold-change differences in miR target expression levels associated with 1-unit increase in lymphocyte counts

	Unadjusted			Adjusted for Lymphocyte				
Target	Fold-change (T2D Case)	p-value	q-value (FDR)	Fold Change (T2D Case)	p-value	q-value (FDR)	Lymphocyte count 10^6 cell/uL	p-value
miR-375	**1.89**	**0.010**	**0.013**	**2.06**	**0.008**	**0.06**	**0.872**	**0.42**
miR-423	**1.72**	**0.000**	**0.002**	**1.51**	**0.006**	**0.06**	**1.212**	**0.05**
miR-223	1.61	0.001	0.003	1.4	0.016	0.0862	1.238	0.02
miR-23a	1.57	0.000	0.002	1.38	0.008	0.06	1.219	0.01
miR-128	1.55	0.002	0.005	1.37	0.030	0.0936	1.201	0.04
miR-24	**1.55**	**0.001**	**0.005**	**1.39**	**0.022**	**0.0936**	**1.176**	**0.08**
miR-425	1.54	0.000	0.002	1.34	0.009	0.06	1.241	<0.01
miR-130b	1.52	0.006	0.012	1.33	0.074	0.1533	1.224	0.05
miR-27a	1.48	0.003	0.007	1.32	0.045	0.1378	1.198	0.04
miR-320a	1.48	0.006	0.012	1.3	0.079	0.1533	1.228	0.03
miR-29a	1.46	0.009	0.013	1.28	0.106	0.1721	1.23	0.04
miR-191	1.46	0.003	0.007	1.29	0.047	0.1378	1.207	0.02
miR-146a	1.45	0.008	0.013	1.26	0.106	0.1721	1.23	0.03
miR-122	1.44	0.157	0.166	1.66	0.074	0.1533	0.801	0.23
miR-155-5p	1.43	0.010	0.013	1.23	0.136	0.1911	1.255	0.02
miR-374	1.42	0.008	0.013	1.26	0.093	0.1685	1.197	0.04
miR-let7d	1.4	0.016	0.019	1.23	0.145	0.1911	1.205	0.05
miR-16	1.31	0.010	0.013	1.21	0.081	0.1533	1.122	0.11
miR-29b	1.27	0.103	0.114	1.24	0.181	0.1911	1.033	0.76
miR-451	1.17	0.432	0.432	1.14	0.556	0.6342	1.041	0.79

Figure 1. Spearman's Correlation for miRNA expression levels and cell count

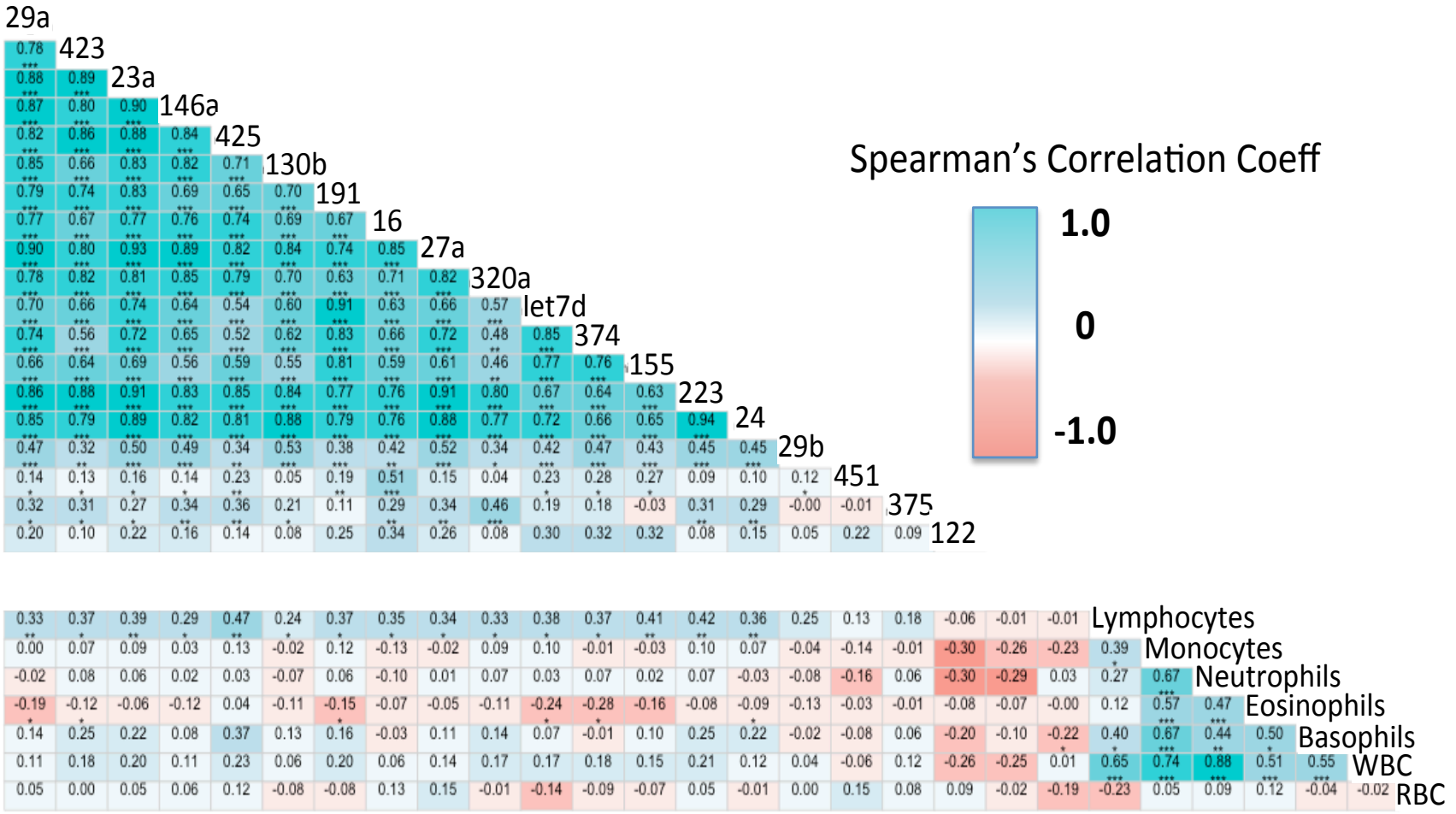
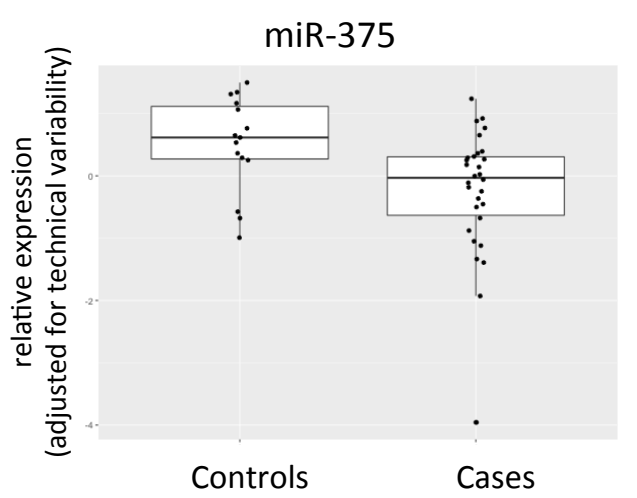
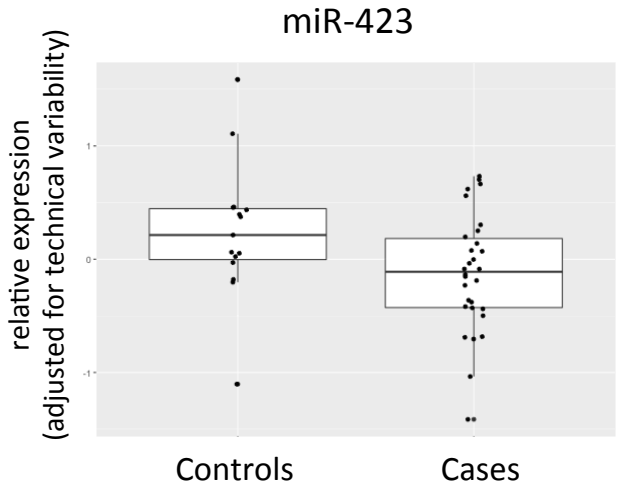
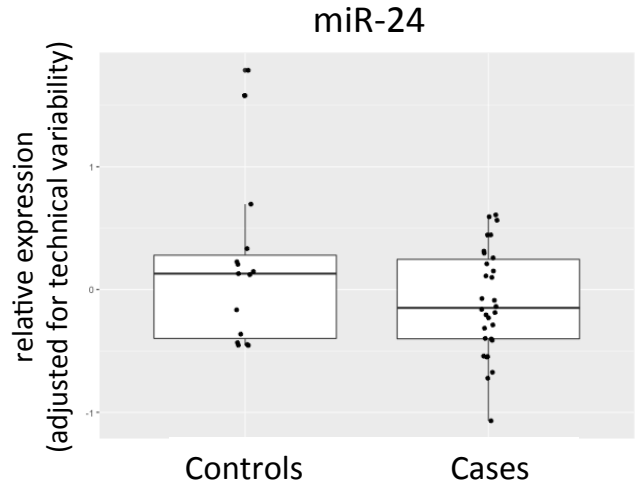


Figure 2. Relative expression differences in cases (defined by HbA1c>6.5%) versus controls for miRNAs that are not influenced by lymphocyte counts.







References

1. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005 Jan 14;120(1):15–20. PMID: 15652477
2. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009 Jan;19(1):92–105. PMID: 18955434
3. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen BS, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci. 2008 Jul 29;105(30):10513–10518. PMID: 18663219
4. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang W, Zhang Y, Wang J, Jiang X, Xiang Y, Xu C, Zheng P, Zhang J, Li R, Zhang H, Shang X, Gong T, Ning G, Wang J, Zen K, Zhang J, Zhang C-Y. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008 Oct;18(10):997–1006.
5. Eikmans M, Rekers NV, Anholts JDH, Heidt S, Claas FHJ. Blood cell mRNAs and microRNAs: optimized protocols for extraction and preservation. Blood. 2013 Mar 14;121(11):e81-89. PMID: 23327925
6. Zampetaki A, Kiechl S, Drozdov I, Willeit P, Mayr U, Prokopi M, Mayr A, Weger S, Oberhollenzer F, Bonora E, Shah A, Willeit J, Mayr M. Plasma MicroRNA Profiling Reveals Loss of Endothelial MiR-126 and Other MicroRNAs in Type 2 DiabetesNovelty and Significance. Circ Res. 2010 Sep 17;107(6):810–817.
7. Karolina DS, Armugam A, Tavintharan S, Wong MTK, Lim SC, Sum CF, Jeyaseelan K. MicroRNA 144 Impairs Insulin Signaling by Inhibiting the Expression of Insulin Receptor Substrate 1 in Type 2 Diabetes Mellitus. PLoS ONE [Internet]. 2011 Aug 1 [cited 2012 Nov 6];6(8). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3148231/ PMCID: PMC3148231
8. Kong L, Zhu J, Han W, Jiang X, Xu M, Zhao Y, Dong Q, Pang Z, Guan Q, Gao L, Zhao J, Zhao L. Significance of serum microRNAs in pre-diabetes and newly diagnosed type 2 diabetes: a clinical study. Acta Diabetol. 2011;48(1):61–69.
9. Karolina DS, Tavintharan S, Armugam A, Sepramaniam S, Pek SLT, Wong MTK, Lim SC, Sum CF, Jeyaseelan K. Circulating miRNA Profiles in Patients with Metabolic Syndrome. J Clin Endocrinol Metab. 2012 Oct 2;97(12):E2271–E2276.
10. Pescador N, Perez-Barba M, Ibarra JM, Corbaton A, Martinez-Larrad MT, Serrano-Rios M. Serum Circulating microRNA Profiling for Identification of Potential Type 2 Diabetes and Obesity Biomarkers. PLoS ONE [Internet]. 2013 Oct 15 [cited 2014 Jan 21];8(10). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3817315/ PMCID: PMC3817315
11. Ortega FJ, Mercader JM, Catalán V, Moreno-Navarrete JM, Pueyo N, Sabater M, Gómez-Ambrosi J, Anglada R, Fernández-Formoso JA, Ricart W, Frühbeck G, Fernández-Real JM.

Targeting the circulating microRNA signature of obesity. Clin Chem. 2013 May;59(5):781–792. PMID: 23396142
12. Wang X, Sundquist J, Zöller B, Memon AA, Palmér K, Sundquist K, Bennet L. Determination of 14 Circulating microRNAs in Swedes and Iraqis with and without Diabetes Mellitus Type 2. PLoS ONE. 2014 Jan 30;9(1):e86792.
13. Prabu P, Rome S, Sathishkumar C, Aravind S, Mahalingam B, Shanthirani CS, Gastebois C, Villard A, Mohan V, Balasubramanyam M. Circulating MiRNAs of "Asian Indian Phenotype" Identified in Subjects with Impaired Glucose Tolerance and Patients with Type 2 Diabetes. PLOS ONE. 2015 May 28;10(5):e0128372.
14. Zhang T, Lv C, Li L, Chen S, Liu S, Wang C, Su B. Plasma miR-126 Is a Potential Biomarker for Early Prediction of Type 2 Diabetes Mellitus in Susceptible Individuals. BioMed Res Int. 2013 Dec 25;2013:e761617.
15. Barry SE, Chan B, Ellis M, Yang Y, Plit ML, Guan G, Wang X, Britton WJ, Saunders BM. Identification of miR-93 as a suitable miR for normalizing miRNA in plasma of tuberculosis patients. J Cell Mol Med. 2015 Jul;19(7):1606–1613. PMCID: PMC4511358
16. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, Tait JF, Tewari M. Blood Cell Origin of Circulating MicroRNAs: A Cautionary Note for Cancer Biomarker Studies. Cancer Prev Res (Phila Pa). 2012 Mar 1;5(3):492–497.
17. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, Dahlsveen IK. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. 2013 Jan;59(1):S1–S6.
18. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, Reid G. Haemolysis during Sample Preparation Alters microRNA Content of Plasma. PLoS ONE. 2011 Sep 1;6(9):e24145.
19. Kirschner MB, Edelman JJB, Kao SC-H, Vallely MP, van Zandwijk N, Reid G. The Impact of Hemolysis on Cell-Free microRNA Biomarkers. Front Genet [Internet]. 2013 May 24 [cited 2014 May 12];4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3663194/ PMCID: PMC3663194
20. Tillin T, Hughes AD, Godsland IF, Whincup P, Forouhi NG, Welsh P, Sattar N, McKeigue PM, Chaturvedi N. Insulin resistance and truncal obesity as important determinants of the greater incidence of diabetes in Indian Asians and African Caribbeans compared with Europeans: the Southall And Brent REvisited (SABRE) cohort. Diabetes Care. 2013 Feb;36(2):383–393. PMCID: PMC3554271
21. Gujral UP, Pradeepa R, Weber MB, Narayan KV, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. Ann N Y Acad Sci. 2013 Apr;1281(1):51–63. PMCID: PMC3715105
22. Pandit K, Goswami S, Ghosh S, Mukhopadhyay P, Chowdhury S. Metabolic syndrome in South Asians. Indian J Endocrinol Metab. 2012;16(1):44–55. PMCID: PMC3263197
23. Schwarzenbach H, da Silva AM, Calin G, Pantel K. Data Normalization Strategies for MicroRNA Quantification. Clin Chem. 2015 Nov;61(11):1333–1342. PMCID: PMC4890630

24. Mitchell AJ, Gray WD, Hayek SS, Ko Y-A, Thomas S, Rooney K, Awad M, Roback JD, Quyyumi A, Searles CD. Platelets confound the measurement of extracellular miRNA in archived plasma. Sci Rep. 2016;6:32651. PMCID: PMC5020735
25. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are Transported in Plasma and Delivered to Recipient Cells by High-Density Lipoproteins. Nat Cell Biol. 2011 Apr;13(4):423–433. PMCID: PMC3074610
26. Gkrania-Klotsas E, Ye Z, Cooper AJ, Sharp SJ, Luben R, Biggs ML, Chen L-K, Gokulakrishnan K, Hanefeld M, Ingelsson E, Lai W-A, Lin S-Y, Lind L, Lohsoonthorn V, Mohan V, Muscari A, Nilsson G, Ohrvik J, Chao Qiang J, Jenny NS, Tamakoshi K, Temelkova-Kurktschiev T, Wang Y-Y, Yajnik CS, Zoli M, Khaw K-T, Forouhi NG, Wareham NJ, Langenberg C. Differential white blood cell count and type 2 diabetes: systematic review and meta-analysis of cross-sectional and prospective studies. PloS One. 2010;5(10):e13405. PMCID: PMC2956635
27. Flowers E, Gadgil M, Aouizerat BE, Kanaya AM. Circulating micrornas associated with glycemic impairment and progression in Asian Indians. Biomark Res. 2015;3:22. PMCID: PMC4785747
28. Poy MN, Hausser J, Trajkovski M, Braun M, Collins S, Rorsman P, Zavolan M, Stoffel M. miR-375 maintains normal pancreatic alpha- and beta-cell mass. Proc Natl Acad Sci U S A. 2009 Apr 7;106(14):5813–5818. PMCID: PMC2656556
29. Erener S, Mojibian M, Fox JK, Denroche HC, Kieffer TJ. Circulating miR-375 as a biomarker of β-cell death and diabetes in mice. Endocrinology. 2013 Feb;154(2):603–608. PMID: 23321698
30. Chang X, Li S, Li J, Yin L, Zhou T, Zhang C, Chen X, Sun K. Ethnic differences in microRNA-375 expression level and DNA methylation status in type 2 diabetes of Han and Kazak populations. J Diabetes Res. 2014;2014:761938. PMCID: PMC3972833
31. Nabiałek E, Wańha W, Kula D, Jadczyk T, Krajewska M, Kowalówka A, Dworowy S, Hrycek E, Włudarczyk W, Parma Z, Michalewska-Włudarczyk A, Pawłowski T, Ochała B, Jarząb B, Tendera M, Wojakowski W. Circulating microRNAs (miR-423-5p, miR-208a and miR-1) in acute myocardial infarction and stable coronary heart disease. Minerva Cardioangiol. 2013 Dec;61(6):627–637. PMID: 24253456
32. Fan K-L, Zhang H-F, Shen J, Zhang Q, Li X-L. Circulating microRNAs levels in Chinese heart failure patients caused by dilated cardiomyopathy. Indian Heart J. 2013 Feb;65(1):12–16. PMCID: PMC3860780
33. Goldraich LA, Martinelli NC, Matte U, Cohen C, Andrades M, Pimentel M, Biolo A, Clausell N, Rohde LE. Transcoronary gradient of plasma microRNA 423-5p in heart failure: evidence of altered myocardial expression. Biomark Biochem Indic Expo Response Susceptibility Chem. 2014 Mar;19(2):135–141. PMID: 24506564
34. Thomé JG, Mendoza MR, Cheuiche AV, La Porta VL, Silvello D, Dos Santos KG, Andrades ME, Clausell N, Rohde LE, Biolo A. Circulating microRNAs in obese and lean heart failure patients: A case-control study with computational target prediction analysis. Gene. 2015 Dec 10;574(1):1–10. PMID: 26211628

35. Qian L, Van Laake LW, Huang Y, Liu S, Wendland MF, Srivastava D. miR-24 inhibits apoptosis and represses Bim in mouse cardiomyocytes. J Exp Med. 2011 Mar 14;208(3):549–560. PMCID: PMC3058576
36. Meloni M, Marchetti M, Garner K, Littlejohns B, Sala-Newby G, Xenophontos N, Floris I, Suleiman M-S, Madeddu P, Caporali A, Emanueli C. Local inhibition of microRNA-24 improves reparative angiogenesis and left ventricle remodeling and function in mice with myocardial infarction. Mol Ther J Am Soc Gene Ther. 2013 Jul;21(7):1390–1402. PMCID: PMC3702112
37. Wang L, Qian L. miR-24 regulates intrinsic apoptosis pathway in mouse cardiomyocytes. PloS One. 2014;9(1):e85389. PMCID: PMC3893205

Supplementary Materials

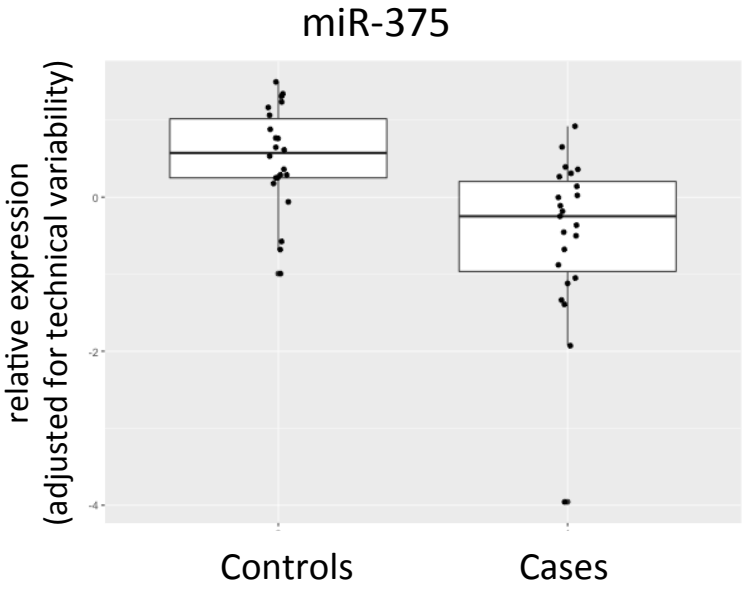
Table S1. Subject Characteristics for Asian Indian T2D Cases (defined by Glucose>= 7mmol/L) and Controls

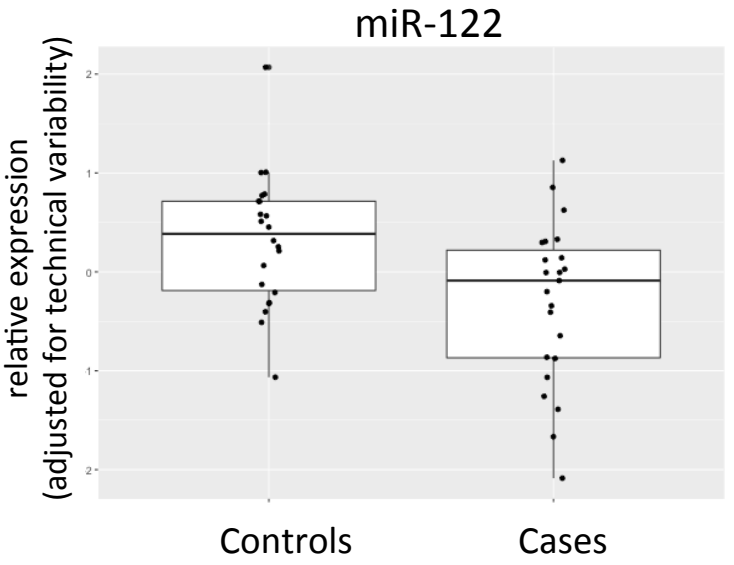
	Controls N=24	Cases N=24	p-value
Glucose	4.67 (0.18)	9.02 (1.63)	**<0.01**
HbA1c	5.5 (0.33)	8.2 (1.29)	**<0.01**
Sex	16 (0.67)	17 (0.71)	0.99
Age	49.49 (8.13)	55.58 (9.76)	**0.02**
Current Smokers (%)	3 (12.5)	3 (12.5)	0.99
BMI	28.15 (3.56)	26.23 (3.66)	0.07
WHR	0.94 (0.06)	0.97 (0.07)	0.18
SBP	125.72 (15.29)	131.7 (15.72)	0.19
DBP	78.7 (12.12)	81.28 (7.9)	0.39
HDL	1.27 (0.28)	1.23 (0.36)	0.66
Trig	1.58 (0.7)	1.9 (0.94)	0.20
Chol	5.23 (1.05)	4.3 (1.32)	**0.01**
LDL	3.24 (0.86)	2.16 (1.08)	**0.00**
WBC	6.63 (2.13)	7.29 (2.13)	0.29
Lymphocyte	1.94 (0.58)	2.3 (0.71)	0.06
Monocytes	0.46 (0.16)	0.5 (0.19)	0.47
Neutrophils	3.98 (1.65)	4.21 (1.65)	0.63
Basophils	0.02 (0.01)	0.03 (0.02)	**0.01**
Eosinophils	0.23 (0.21)	0.24 (0.16)	0.75

Table S2. Fold-Change Differences in miR Target expression levels for T2D cases (defined by glucose>=7mmol/L) versus controls, un/adjusted for fold-change differences in miR target expression levels associated with 1-unit increase in lymphocyte counts

	Unadjusted			Adjusted for lymphocytes				
Target	Fold-change (T2D Case)	p-value	q-value (FDR)	Fold Change (T2D Case)	p-value	q-value (FDR)	Lymphocyte count 10^6 cell/uL	p-value
miR-375	**2.31**	**<0.001**	**0.004**	**2.42**	**<0.001**	**0.004**	**0.9**	**0.48**
miR-122	**1.64**	**0.041**	**0.158**	**1.77**	**0.024**	**0.236**	**0.827**	**0.26**
miR-423	1.42	0.016	0.110	1.28	0.081	0.405	1.298	0.01
miR-425	1.35	0.010	0.096	1.21	0.064	0.405	1.298	0.00
miR-128	1.31	0.047	0.158	1.18	0.192	0.728	1.266	0.01
miR-23a	1.28	0.044	0.158	1.15	0.218	0.728	1.297	0.00
miR-320a	1.26	0.103	0.247	1.13	0.358	0.868	1.289	0.01
miR-155-5p	1.26	0.090	0.247	1.13	0.344	0.868	1.3	0.00
miR-223	1.25	0.111	0.247	1.11	0.428	0.868	1.333	0.00
miR-29a	1.23	0.145	0.290	1.11	0.456	0.868	1.29	0.01
miR-24	1.21	0.162	0.294	1.10	0.477	0.868	1.266	0.01
miR-27a	1.19	0.189	0.315	1.07	0.570	0.942	1.275	0.01
miR-146a	1.16	0.284	0.379	1.04	0.773	0.942	1.303	0.00
miR-130b	1.16	0.329	0.412	1.03	0.818	0.942	1.316	0.01
miR-191	1.16	0.224	0.345	1.05	0.685	0.942	1.283	0.00
miR-16	1.12	0.261	0.372	1.05	0.628	0.942	1.173	0.02
miR-let7d	1.12	0.411	0.483	1.01	0.935	0.942	1.275	0.01
miR-374	1.1	0.481	0.534	0.99	0.932	0.942	1.279	0.01
miR-451	1.09	0.652	0.686	1.06	0.767	0.942	1.068	0.64
miR-29b	1.05	0.733	0.733	1.01	0.942	0.942	1.095	0.37

Figure S1. Relative expression differences T2D cases (defined by fasting blood glucose>=7mmol/L) versus controls for miRs that are not influenced by lymphocyte counts





Chapter 5: Summary and Conclusion

The purpose of this dissertation is to identify viable blood biomarkers of exposure and disease related to type II diabetes (T2D) using improved methods and experimental design. Specific populations are at differential risk of the T2D, which provides an opportunity to make comparisons between groups. Incidence of T2D is particularly high in India, and prevalence is rising in diaspora Asian Indian populations in other nations, including the US and UK (1–3). By gaining greater understanding of this specific population's blood profile, we can generate new hypotheses to help explain the mechanisms of disease. A semi-targeted exposomics approach is used here to measure candidate blood biomarkers of exposure and disease in Asian Indians that have been previously reported to be associated with T2D in other populations.

This dissertation takes advantage of new methods used to examine biomarkers in relatively small volumes of blood. A recent development within Agilent Technologies requires only 200 µL of plasma to measure the blood concentrations of 66 environmental pollutants (as compared to previous methods which required milliliters of blood). For biomarkers such as miRNA, only 200 µL of plasma are needed to extract the RNA, and a small fraction of the eluted RNA is used for qPCR plasma expression. Pilot experiments for both the pollutant and miRNA measurements helped determine the extent of biological and technical variability *a priori* before expanding these studies to test precious samples.

Several valuable findings were obtained from these studies. After screening for many environmental pollutants in just 49 Asian Indians, the pilot study showed that DDE levels tended to be higher in diabetics compared to controls. In Chapter 2, this study was expanded to a larger sample population of N=200 individuals. Asian Indians showed 3-30x higher concentrations of organochlorine pesticides compared to European whites, including DDT, DDE, and β-HCH. For Asian Indians, there was >5-fold higher odds of (relatively) high exposure to DDE and β-HCH for subjects with T2D compared to healthy individuals. While these pesticides have previously been associated with T2D in populations from the US, Sweden, and elsewhere (4), this is the first study to demonstrate this in an Asian Indian population, even years after emigration from the main source of exposure. These observations add to the growing body of literature that early-life pesticide exposure can increase the risk of T2D onset later in life.

Interestingly, the metabolic phenotype of Asian Indians makes them more susceptible to T2D. T2D is more prevalent in Asian Indians regardless of traditional metrics of obesity and have increased insulin resistance and dyslipidemia at lower BMIs than whites and other ethnic groups (5,7). This may be due in part to the fact that Asian Indians also have disproportionately more visceral fat around their midsection and internal organs than other ethnic groups (5,7). Asian Indians have a greater prevalence of insulin resistance at baseline, and higher levels of adiponectin and C-reactive proteins (5)—all risk factors for T2D. These unique traits may be due to greater body fat and differences in fat deposition in Asian Indians compared to whites, even in early-life (6,7). The differences in mechanisms of glucose use and fat storage in Asian Indians versus whites are only recently being explored (8), and greater effort is needed to determine the underlying causes leading to T2D. As many POPs are lipophilic compounds predicted to be at 100-1000x higher concentrations in adipose tissue than in the blood (9), this may elicit an exacerbated effect on key endocrine organs, such as the pancreas and liver, for this population

(as described in animal models (10)). Thus, POPs exposures may lead to prolonged downstream effects on the endocrine system that augment the risk of T2D in Asians Indians.

Gaining understanding of environmental factors that are causally related to T2D requires well-designed longitudinal studies, which is essential to the next stage of pursuing exposomics. As progression to T2D can occur 5-10 years prior to clinical manifestation of the disease, future studies will need years of follow-up and meticulous measurement of blood exposure levels. As seen in this study, environmental pollutants in blood samples can correlate with each other and with T2D status, including DDT, DDE and β-HCH. Therefore it is important to determine whether these exposures (namely DDE and β-HCH) interact synergistically or independently to affect metabolic activity. While this study shows that DDE levels are higher in T2D cases versus controls, other studies have shown the parent compound, DDT, levels to be higher (4). Furthermore, animal and tissue culture models provide evidence that DDT maybe the causal agent of disease. It is plausible that DDT is, in fact, the more active compound and the mechanism of action that causes glucose dysregulation, occurring prior to degradation of the metabolite, DDE. Similarly, β-HCH, found in technical grade HCH and as a by-product in Lindane, may simply be a proxy for exposure to Lindane (ϒ-HCH), which has a much shorter half-life but perhaps more detrimental effects on glucose regulation *in vivo* that occur immediately upon onset.

An exposome approach within cohort studies will shed more light on which environmental pollutants contribute to T2D onset later in life. It is imperative that these T2D cohort studies include samples from early-life, as windows of greatest susceptibility are likely to occur before birth, in the first years of life, and during puberty. Multiple collection time-points will allow further investigation of the relationship between these analytes and their rates of metabolism over time with respect to disease onset. Ultimately, an individual's cumulative measure of exposures can be constructed along with risk of their disease outcomes. This process will be facilitated by the new high-throughput MS/MS platform described in this study that can measure tens to hundreds of persistent and non-persistent organic pollutants. Given technological advancements in recent years, screening of environmental exposures via exposomics will become high-throughput for more efficient and effective analysis of subject samples from cohort studies

The field of miRNAs as blood biomarkers is still in its infancy as well, thus, optimization of current methods is crucial before measuring differences in expression levels between comparison groups. For miRNA in particular, very small (<2-fold change) differences are generally observed in human population studies, therefore, fine-tuning the method and accounting for variability is essential to capture these small fluctuations in expression levels. Chapter 3 describes how targeted miRNA methods were used to measure the biological and technical sources of variability. A substantial amount of technical variability was initially observed when using a probe-based method of qPCR measurement. This technical variability was reduced when changing to a SYBR green-based qPCR method of detection. In addition, simulations were used, based on empirical data, to show that statistical power to detect small (2-fold) changes is increased by including technical and biological replicates in a hypothetical study. While greatly overlooked, identifying sources of variability for new biomarkers of interest is imperative in order to maximize our ability to discern differences between comparison

groups.

The objective in Chapter 4 was to verify 20 miRNAs that were previously reported as differentially expressed in separate T2D sample populations. A case-control study on Asian Indians was designed and executed based on the methods optimized in Chapter 3. Most miRNAs were directly correlated with lymphocyte count, which was higher in cases versus controls. Only 3 miRNAs, miR-375, miR-24, and miR-423, were associated with T2D and not confounded by lymphocyte counts.

This finding should be strongly considered with respect to the field of circulating plasma miRNA and blood biomarkers, in general. Other miRNA studies on T2D have shown a universal increase or decrease in circulating miRNA expression levels with respect to disease status, and these findings may also be highly influenced by the circulating white blood cell counts and should be evaluated with skepticism. In light of this finding, others published in the field showing associations between blood cell count and circulating miRNA expression levels (11). In Chapter 4 as well a direct relationship was found between lymphocyte (white blood cell) count and miRNA expression for a majority of miRNAs tested. New conventions are needed to adjust for white blood cell counts in future human and animal studies. Moreover, novel free-floating nucleotide, amino acid, or lipid material in the blood milieu should be assessed for correlation with major blood contents before assuming that these markers are likely secreted directly from tissues of interest related to the disease.

Pathogenesis of T2D is still poorly understood as well, and new biomarkers, such as miRNA, discovered through prospective studies are imperative for early detection, diagnosis, and treatment. While several other studies have identified these miRNAs as biomarkers of T2D status, cohort studies can determine if these are also predictive markers of the disease. There is potential for circulating miRNAs to be used as an early-marker of T2D, as has previously been shown for miRNA-126 in a large cohort (12). Within the context of the exposome, these miRNA targets of interest can also be correlated to individuals' blood exposure levels as well, as miRNAs have previously been associated with both environmental exposures and disease status, acting as an intermediate marker (13).

Identifying preventable risk factors for T2D and related metabolic outcomes that may be specific to exposures in India is crucial for the health and welfare of the Asian Indian diaspora. Semi-targeted exposomics approaches, exemplified here, show promise in identifying biomarkers of T2D. Cross-sectional studies such as this are an efficient and economical first step to determine baseline differences between populations and generate new hypothesis for further examination in longitudinal studies.

References

1. Barnett AH, Dixon AN, Bellary S, Hanif MW, O'hare JP, Raymond NT, et al. Type 2 diabetes and cardiovascular risk in the UK south Asian community. Diabetologia. 2006 Oct;49(10):2234–46.
2. Kanaya AM, Wassel CL, Mathur D, Stewart A, Herrington D, Budoff MJ, et al. Prevalence and correlates of diabetes in South asian indians in the United States: findings from the metabolic syndrome and atherosclerosis in South asians living in america study and the multi-ethnic study of atherosclerosis. Metab Syndr Relat Disord. 2010 Apr;8(2):157–64.
3. Kanaya AM, Herrington D, Vittinghoff E, Ewing SK, Liu K, Blaha MJ, et al. Understanding the high prevalence of diabetes in U.S. south Asians compared with four racial/ethnic groups: the MASALA and MESA studies. Diabetes Care. 2014 Jun;37(6):1621–8.
4. Thayer KA, Heindel JJ, Bucher JR, Gallo MA. Role of Environmental Chemicals in Diabetes and Obesity: A National Toxicology Program Workshop Review. Environ Health Perspect. 2012 Jun;120(6):779–89.
5. Unnikrishnan R, Anjana RM, Mohan V. Diabetes in South Asians: Is the Phenotype Different? Diabetes. 2014 Jan 1;63(1):53–5.
6. Bakker LEH, Sleddering MA, Schoones JW, Meinders AE, Jazet IM. Pathogenesis of type 2 diabetes in South Asians. Eur J Endocrinol Eur Fed Endocr Soc. 2013 Nov;169(5):R99–114.
7. Misra A, Bhardwaj S. Obesity and the metabolic syndrome in developing countries: focus on South Asians. Nestlé Nutr Inst Workshop Ser. 2014;78:133–40.
8. Bakker LEH, Guigas B, van Schinkel LD, van der Zon GCM, Streefland TCM, van Klinken JB, et al. Middle-aged overweight South Asian men exhibit a different metabolic adaptation to short-term energy restriction compared with Europeans. Diabetologia. 2015 Jan;58(1):165–77.
9. Toppari J, Larsen JC, Christiansen P, Giwercman A, Grandjean P, Guillette LJ, et al. Male reproductive health and environmental xenoestrogens. Environ Health Perspect. 1996 Aug;104 Suppl 4:741–803.
10. Bigsby RM, Caperell-Grant A, Madhukar BV. Xenobiotics released from fat during fasting produce estrogenic effects in ovariectomized mice. Cancer Res. 1997 Mar 1;57(5):865–9.
11. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, et al. Blood Cell Origin of Circulating MicroRNAs: A Cautionary Note for Cancer Biomarker Studies. Cancer Prev Res (Phila Pa). 2012 Mar 1;5(3):492–7.
12. Zampetaki A, Kiechl S, Drozdov I, Willeit P, Mayr U, Prokopi M, et al. Plasma MicroRNA Profiling Reveals Loss of Endothelial MiR-126 and Other MicroRNAs in Type 2 Diabetes. Circ Res. 2010 Sep 17;107(6):810–7.
13. Vrijens K, Bollati V, Nawrot TS. MicroRNAs as Potential Signatures of Environmental Exposure or Effect: A Systematic Review. Environ Health Perspect. 2015 May;123(5):399–411.