THE PREMATURITY RESEARCH CENTERS DATA REPOSITORY IS ONE OF THE MOST IMPORTANT INNOVATIVE NEW TOOLS FOR COLLABORATIVE RESEARCH. NOW IT'S READY TO BECOME AN ENGINE OF DISCOVERY.

From the time she was in high school, Marina Sirota was interested in the intersection of math and biology. Her timing was perfect. The field she would eventually excel in, bioinformatics, was just being born from the conjunction of the great surge of computing power being invented and advances that made it possible to examine deeper and deeper layers of information previously hidden in molecular biology.

As an undergraduate at Stanford, she had an independently designed major: math and computer science curricula with an emphasis on machine learning and artificial intelligence, and a biology curriculum focused on genetics and genomics. As a graduate student, she again found herself in exactly the right place at the right time—working in the lab of a bioinformatics pioneer, Atul Butte, to develop the Prematurity Research Centers Data Repositor.

The Data Repository was launched in 2014, at the March of Dimes Prematurity Research Center (PRC) at Stanford University School of Medicine, integrating the work of researchers and clinicians from a broad range of disparate disciplines. More than 200 scientists, engineers, statisticians, sociologists and medical professionals work together to foster innovation that will accelerate discoveries of the unknown causes of preterm birth and new ways to prevent it. The March of Dimes Data Repository for Preterm Birth Research is truly the backbone of the research effort.

The Data Repository was designed to be both a catalog and catalyst. It provides the capability to ask new and important questions about preterm birth. Dedicated to enhancing research collaboration and coordination, the Data Repository currently contains information from 13 studies, with individual molecular profiles on 365 patients, aggregated genetic data from more than 30,000 patients, and molecular measurements on more than 8,000 samples. More data is being added every day, and the Data Repository's contents were recently opened to all the other researchers studying preterm birth around the world, in an effort to encourage them to add their own findings to the Data Repository.

The advent and continued development of genotyping and next generation gene sequencing technologies has enabled researchers to generate a vast amount of molecular data. The March of Dimes Data Repository is highly diverse and includes genomic, transcriptomic, immunological and microbiomic data. In addition, relevant datasets are being augmented with data and results from participating laboratories with public and government databases and private data sources

Like most of the work it houses, the Data Repository is itself a synergistic collaboration between the Stanford PRC and several other entities and organizations. These include Northrop Grumman Health Solutions (NG), a partner of the National Institute of Allergy and Infectious Diseases (NIAID) Division of Allergy, Immunology, and Transplantation at the National Institutes of Health (NIH). Since 2004, NG has been working on the ImmPort database, a data sharing portal to ensure that NIH-funded discoveries serve as the foundation for future research. Bioinformatics, scientific data management, and analysis are rapidly becoming essential components of any large-scale research effort and the March of Dimes Data Repository will eventually join that category.

"The Prematurity Research Centers Data Repository is truly the backbone of the preterm birth research effort."
Marina Sirota, Ph.D., Assistant Professor Institute for Computational Health Sciences University of California, San Francisco

"Essentially, we're creating a platform where researchers from all over the world can start applying machine learning and artificial intelligence tools to the data that's coming out of all six March of Dimes PRCs," says Dr. Sirota. "In order to do that, we're in the process of integrating and organizing all of this data and making it accessible and available not just to the PRCs, but also to the research community at large. And secondly, we want to also start using it for research, mining this data and performing meta-analyses in order to begin asking questions we simply weren't able to ask before."

This second goal is of particular significance to Dr. Sirota, who is not only part of the Stanford PRC but also a faculty member at the University of California, San Francisco, where she leads a team of computational biologists. Her group is already using the Data Repository for two research projects. Her group has already had great success in mining publicly available datasets to enable computational drug discovery in the areas of autoimmunity and cancer (see Sirota et al, Science Translational Medicine 2011). Her group's current work will be added to the curated trove of knowledge.

Moms and babies in the U.S. are facing an urgent health crisis:

• In this country 1 in 10 babies is born prematurely each year

• Worldwide 15 million babies are born prematurely each year.

• Premature birth and its complications are the largest contributors to infant death in the United States and globally.

• More than 380,000 babies are born prematurely in the U.S. each year.

• In addition to the human toll, the societal cost of premature birth is more than $26 billion in the U.S. per year.

• Women of color are up to 50 percent more likely to give birth prematurely and their children can face a 130 percent higher infant death rate.

• In this country black women have maternal death rates over three times higher than women of other ethnicities.

• More than 20 percent of premature babies are born to black women– that's 1 in 5 babies.

• Employers pay 12 times as much in health care costs for premature/ low birthweight babies compared to babies born without these complications.

Because premature birth has many possible causes, each PRC is charged with exploring a different transdisciplinary research target that is likely to be crucial to the prevention of premature birth. Stanford University research themes have unique strengths in the study of the microbiome, the transcriptome, and the immunome. In addition, they also house the Data Repository for all the Prematurity Research Centers enhancing collaboration and data coordination.

"A large part of the work we did on the data for these two projects was to aggregate and normalize the data to integrate it as a cohesive whole," says Dr. Sirota. "That's an important step because not all of the data came to us on the same computational or technical footing. But once that's completed, we will release our final dataset so that others don't have to do that work again."

This is far from the traditional approach of a single researcher or small team working alone in an isolated laboratory and publishing their results without much communication or collaboration. This is a new world and the public availability of data, coming together with emerging technologies to process and analyze it, represents a significant opportunity to advance the current state of understanding about disease processes and what we might do to prevent or cure them.

The overarching goal of this effort is to enable new scientific discoveries from the rich molecular resources that have been funded by March of Dimes, and others, to advance the research in preterm birth. The resulting Data Repository, led by Marina Sirota and her team, is rapidly becoming a powerful tool to do just that.

"Precision medicine is an emerging integrative approach for disease prevention," Dr. Sirota said. "Individual variability in genetics, epigenetics, environment, and lifestyle can now be taken into account for early detection and treatment. The public availability of genomics data and our efforts to share it, combined with powerful bioinformatics tools, are inspiring groundbreaking new work in the search to end preterm birth."

THEME 4 LEADER

Marina Sirota, Ph.D., *Assistant Professor*
Institute for Computational Health Sciences
University of California, San Francisco

DONATE

For more information on how you can be a part of this effort please contact: 914.997.4492

MARCHOFDIMES.ORG

MARCH OF DIMES
PREMATURITY RESEARCH CENTER
Stanford University