The Second Genome

Microbiome-based diagnostic markers

Sharmila Mande

IN BRIEF

We carry more microbe cells than our own cells! We know that these microbes play a crucial role in our health and well-being. Although one can now obtain genome sequence data corresponding to these microbes using state-of-art sequencing machines, gaining meaningful insights from such complex and voluminous data is challenging.

TCS' Life Sciences Research team has not only developed efficient algorithms for every step of analysis and management of the data, but has also carried out cutting edge research to understand the link between microbial community (called "microbiome") and diseases. The team's innovative microbiome-based solutions can be utilized for not only monitoring health status of an individual, but also for predicting at an early stage his/her risk for a number of asymptomatic diseases and disorders.

> Our algorithms can be used for not only understanding microbial communities present in our body, but for obtaining key insights into role of such communities which are present in other environments (soil, water, etc.) too

Can you believe that microbes resident inside us outnumber not only our own cells, but also contribute to close to billion microbial genes as against about 20,000 human genes!! Thus, apart from our own genome, we have our second genome (also called metagenome) contributed by DNA of all the inhabiting microbial community, collectively called the microbiome. In order to understand whether and how these tiny microbes residing within us influence our health, we need to study their genomes. As most of these microbes cannot be cultured in the laboratory, they cannot be studied using traditional genomics approaches. Thanks to the progress in science and technology and engineering, a new generation of DNA sequencing machines can now sequence the entire DNA of all the microbes that reside in any environment. This field which bypasses the culturing step is called metagenomics.

The field of Metagenomics has advanced rapidly in the last decade. Enormous amounts of DNA data has been/ is being generated and analyzed. However, there are some challenges in analyzing DNA sequencing data obtained from a microbial community:

- The sequencing data is voluminous and noisy

- Analysis requires complex algorithms, given the fact that we have no prior idea about the

Fact File

TCS Research: Metagenomics and Microbiome

Outcomes: Several patented algorithms for analysis, visualization, and management of metagenomics data, First of its kind "Microbiome-based biomarkers" for risk assessment of diseases/disorders

Principal Investigator: Sharmila S. Mande

Academic Partners: National Institute of Cholera and Enteric Diseases (NICED, Kolkata), Translational Health Science and Technology Institute (THSTI, Faridabad), Dr. Mohan's Diabetes Research Foundation (MDRF, Chennai), Hinduja Hospital (Mumbai), Institute of Genomics and Integrative Biology (IGIB, Delhi), National Centre for Cell Sciences (NCCS, Pune), Institute of Advanced Study in Science and Technology (IASST, Guwahati), Tata Chemical (Pune), Novo-Nordisk Foundation Centre for Basic Metabolic Research (University of Copenhagen, Denmark)

Techniques Used: Next Generation DNA Sequencing, Genomics, Metagenomics, Functional Metagenomics, Data Mining, Machine Learning

Industries Benefited: Healthcare, Insurance, Pharma, Biotech

Patents: 84 filed, 23 granted

Papers: 92 in peer-reviewed international journals

types and amount of bacteria to look for

• At each step of the analysis, specialized algorithms, visualization tools, and statistical methods are needed to obtain biologically meaningful insights from metagenomic data.

TCS' Life Sciences R&D team has developed efficient algorithms for not only every step of the analysis, but also for management of metagenomic "BIG data". These algorithms can be used for analyzing metagenomic data as well as for comparing multiple metagenomic datasets sampled across space and/or time. This is specifically important if one wants to compare datasets corresponding to healthy state and diseases. Our algorithms can be used for not only understanding microbial communities present in our body, but also for obtaining key insights into role of such communities which are present in other environments (soil, water, etc.).

Screening for Risk of Asymptomatic Diseases

When we talk about human health, the first question which comes to our mind is whether we can diagnose diseases at a very early stage so that preventive measures can be taken. A large number of diseases and disorders, like cancers, diabetes, cardiovascular diseases do not show any symptoms in their early stages. In a healthy

> Unlike existing preterm birth diagnostic solutions, which are only applicable at later stages of pregnancy, our biomarker works in the first trimester with significantly high accuracy (>95%)

individual, there is a fine balance between various microbial groups residing in the body. This balance is lost in diseased individuals. Therefore, "variations" in the bacterial communities residing within us have the ability to serve as "diagnostic biomarkers" that can foretell the presence and/or stage of disease. Taking this as the clue, TCS' Life Sciences Research team has captured these patterns of imbalances, i.e. "variations" in order to develop microbiome-based diagnostic markers.

Preterm Deliveries

To give an example, take the case of preterm births (PTB). Every year, about 15 million babies are born pre-term[1] . Out of them, more than a million die due to medical complications. Premature birth results in more neonatal deaths than any other disease. Every 30
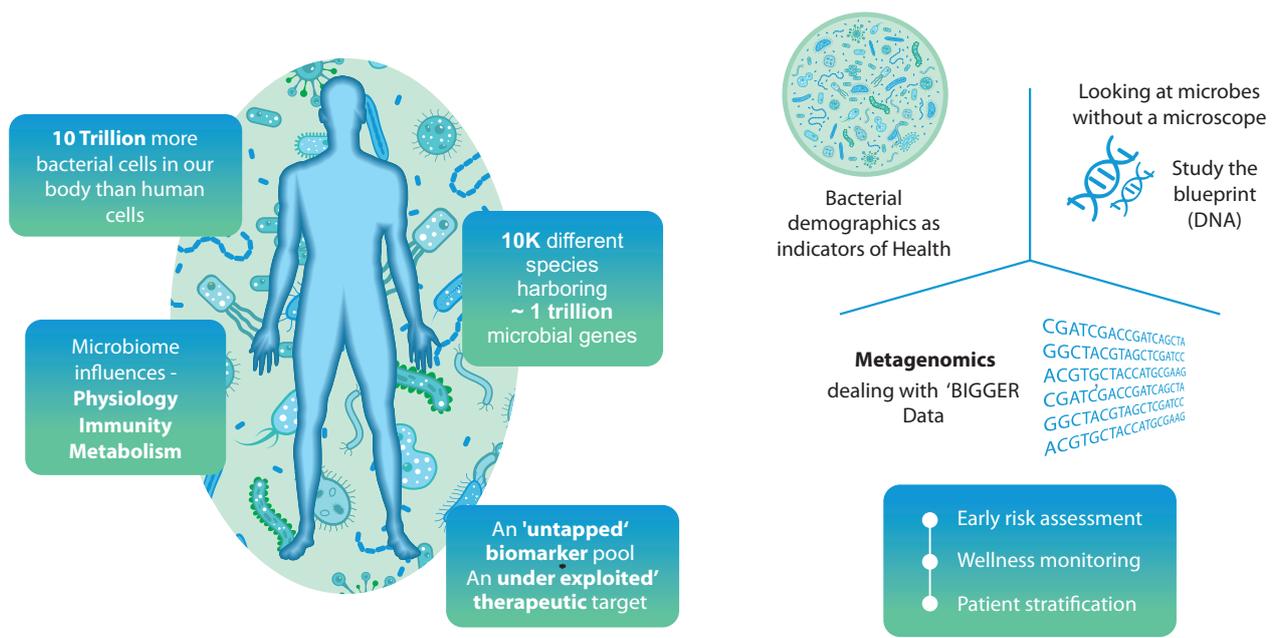
seconds, a new born child dies. A preterm delivery is difficult to predict and at the moment, there are no diagnostic methods that can accurately raise an early alarm. Therefore, TCS' microbiome research team looked beyond physical traits as well as biochemical tests and focused on studying the bacterial communities in pregnant women. We studied a lot of publicly available microbiome data from pregnant women and deciphered patterns that have the ability to act as biomarkers for preterm delivery. Unlike existing PTB diagnostic solutions, which are only applicable at later stages of pregnancy, our biomarker works in the first trimester with significantly high accuracy (>95%). Our biomarker relies on metagenomic sequencing and analysis of a single microbiome sample (either a vaginal swab or a saliva/ stool sample) collected from a pregnant woman. The

10 Trillion more bacterial cells in our body than human cells

Microbiome influences -
Physiology
Immunity
Metabolism

10K different species harboring ~ **1 trillion** microbial genes

An **'untapped'** **biomarker** pool An **under exploited'** **therapeutic** target

Bacterial demographics as indicators of Health

Looking at microbes without a microscope

Study the blueprint (DNA)

Metagenomics dealing with 'BIGGER Data

CGATCGACCGATCAGCTA
GGCTACGTAGCTCGATCC
ACGTGCTACCATGCGAAG
CGATCGACCGATCAGCTA
GGCTACGTAGCTCGATCC
ACGTGCTACCATGCGAAG

Early risk assessment

Wellness monitoring

Patient stratification

Figure 1: *Microbiome: new descriptors of health*

1 https://www.who.int/news-room/fact-sheets/detail/preterm-birth

The Gut Factor

The TCS "Gut Health Score" can be used to monitor an individual's gut health status with ease.

Your stool sample can point to the presence of several diseases, some of them serious and often asymptomatic. No bullshitting, this! Analysis of metagenomes is growing in importance as a diagnostic tool; and a stool sample is an indicator of not just gut health, but of other organs as well.

The gut microbiome is a fascinating one; here are some reasons why:

- A majority of microbes in our body reside in our gastrointestinal track (gut).

- While moving down the alimentary canal, food is stopped by trillions of microbes in the gut microbiome. (The total weight of all the microbial community in the gut is estimated to be around 2 kg – which incidentally is similar to the weight of our brain!)

- Two-thirds of the gut microbiome is unique to each individual, the uniqueness being dependent on the food we eat, ethnicity, health condition, and other environmental factors.

- Some food we eat cannot be digested by our own enzymes. When the stomach and small intestine are unable to digest such foods, certain gut bacteria help ensuring that we get the nutrients that can be absorbed by our intestine. Certain gut bacteria modulate secretion of specific hormones (e.g. serotonin) and vitamins (vitamin B and K). The gut microbiome also plays a role in bioavailability of drugs as well as resistance to antibiotics.

- Aberrations in gut microbial communities have been shown to be connected with several diseases and disorders. Interestingly, gut microbial community also controls functioning of distal organs like heart, brain, lung, and liver.

TCS' Life Sciences Research team has been actively working on the gut microbiome and has demonstrated expertise in not only developing metagenomics algorithms, but also in translational microbiome research.

method has been validated using publicly available microbiome data pertaining to pregnant women (on over 1100 clinically collected samples from diverse geographies and ethnicities). Since our microbiome-based diagnostic biomarker can accurately predict the risk of preterm birth as early as in the first trimester of pregnancy, clinicians will have enough time to act and prevent the preterm delivery.

Apart from predicting the risk of preterm delivery, TCS team has successfully developed microbiome-based biomarkers that can be used for accurate screening of colorectal cancer and

breast cancer using microbiome sequence data obtained from an individual's stool sample. The biggest advantage of using our biomarkers is that these are non-invasive (unlike traditional clinical diagnostics/screening methods) as well as low cost. They can also predict early the risk of a number of asymptomatic diseases/disorders. This gives doctors a chance to start timely therapeutic intervention and manage the patient and the disease in a much more effective manner.

Future of Personalized Healthcare

While linking imbalance in gut microbiome composition to diseases isn't new, TCS' Life Sciences Research team has pioneered in (1) microbial function based assessment of gut health and (2) microbiome-based biomarker discovery for early risk assessment of diseases/disorders. With the technological advances and rapidly declining costs of DNA sequencing, it will be possible in the near future to routinely sequence gut microbial communities just from stool samples of individuals in a cost-effective manner. A simple analysis of this data using TCS' innovative methods will be able to provide clues on the health condition of the individuals. This will help in not only defining the best diet and nutrition plan for them, but also will indicate whether they need to undergo further medical investigations or interventions.

What makes for a healthy gut microbiome? We carried out collaborative work with National Institute of Cholerae and Enteric Diseases (NICED, Kolkata) and Translational Health Science and Technology Institute (THSTI, Faridabad) to understand the

Minimally-Invasive
>95% accuracy

Vaginal Swab

~ 80% accuracy

Saliva Sample

~ 75% accuracy

Stool Sample

Pregnant Woman (1st Trimester)

Potential Benefits

- Early risk assessment
- High accuracy
- Affordable
- Non-invasive
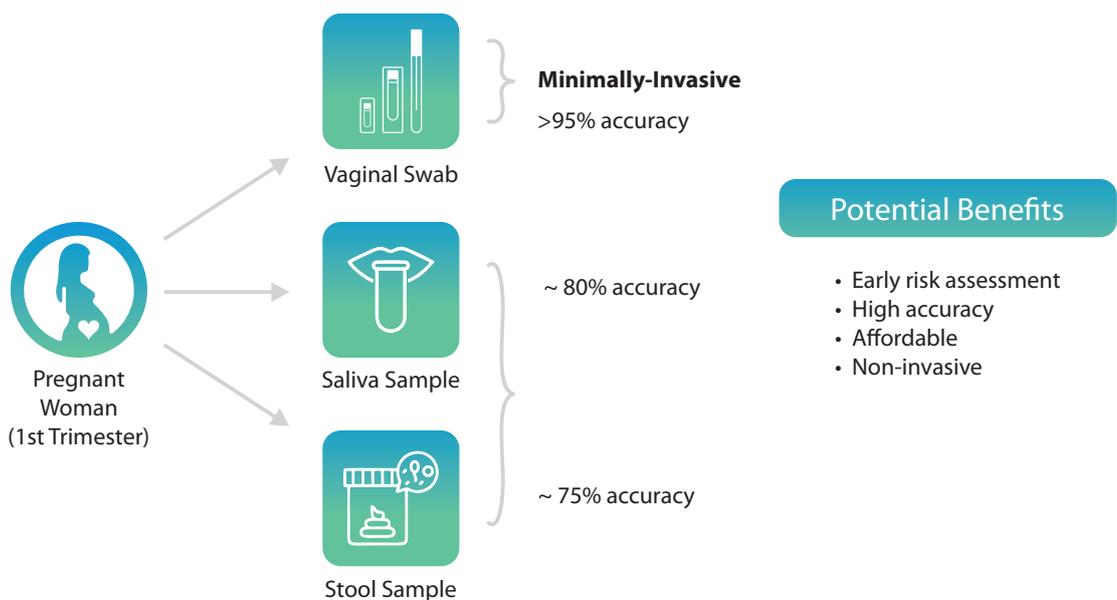
Figure 2: *Preterm birth diagnostics*

link between gut microbiome and malnourishment. Since the gut microbiome plays a crucial role in nutrient pre-processing, assimilation, and energy harvest from food, one would expect dysbiosis of the gut microbial community in malnourished children. We analyzed microbial DNA extracted from faecal samples (which represent gut microbiome) obtained from children from rural areas in West Bengal. We selected children with varying nutritional status, i.e. ranging from severe acute malnourished children at one end of the spectrum to apparently healthy children at the other end. An interesting question is "Regardless of living under similar conditions of hygiene and sanitation, why are some children malnourished while the others are apparently healthy?" Our study findings suggest that the indicative characteristics of the gut environment rest not only on the presence of specific microbial groups, but also on the inherent inter-microbial interactions present in it. The results showed that with decrease in the nutritional status, there is a possible increase in functional interdependence among the various microbes residing in the gut. As per the analysis it is possible to modulate the gut microbiome by disrupting certain key players (microbes) so as to achieve conditions that could result in the regression of the malnourishment phenotype. In other words, appropriate nutritional regimes as well as new probiotic and prebiotic supplements can be developed for tackling malnourishment problem.

Monitoring Gut Health: Towards Personalized Diet and Therapy

Realizing that a healthy and balanced gut microbiota is one of the keys for ensuring good health, it would be ideal to be able to monitor and understand health of the gut (and its microbiome) periodically. It is to be noted that microbial communities may be very different in two healthy individuals from two different geographies, but they may perform similar biological functions. Thus, in order to assess "gut health" one needs to look at not just the microbial groups residing therein, but also their functional potential. TCS' Life Sciences Research group has explored specific functional signals from the microbiome for developing a method for assessing and quantifying gut health status. TCS "gut health score" can be used to monitor an individual's gut health status just from collected stool samples!

How robust is the TCS gut health score in differentiating between a dysbiotic and healthy gut? We have validated our score using around 600 publicly available microbiome datasets corresponding to healthy and diseased samples. The diseases included Crohn's disease, inflammatory bowel disease (IBD), diabetes, asthma, Colo-rectal cancer (CRC), nonalcoholic fatty liver disease, Parkinson's disease, and cystic fibrosis. Our gut health score could indeed differentiate the disease samples from the healthy ones. The findings also revealed that this score could indicate disease conditions which affect not only the gut (e.g. in the case of IBD, Crohn's, CRC), but also distal

The TCS team has successfully developed microbiome-based biomarkers that can be used for accurate screening of colorectal cancer and breast cancer

organs (like brain, lung, liver). This suggests that one can utilize a simple stool-based test to assess gut health of an individual and identify risk of various diseases and disorders. Going beyond this, identifying functions determinant of gut health condition of an individual can provide insights into designing personalized therapeutics which may be in the form of prebiotics, probiotics or nutritional supplements.

Sharmila S Mande

Sharmila S. Mande, a distinguished chief scientist at TCS Research, has 35 years of research experience in Life Sciences domain. She received her Ph.D. degree in the year 1991 in Physics from Indian Institute of Science (IISc), Bangalore. She had her postdoctoral training at University of Groningen, The Netherlands and University of Washington, Seattle, USA. After returning to India, she continued her research at Institute of Microbial Technology (IMTech) and Post Graduate Institute of Medical Education and Research (PGIMER) in Chandigarh, before joining TCS in 2001 to start TCS' Life Sciences R&D activities. She has published several papers in international journals and holds a number of patented algorithms and software solutions that address challenges faced by researchers in analyzing large-scale biological data.