I have been awarded the Fellowships of the "Indian Academy of Sciences" (https://www.ias.ac.in/) in 2023 and the "West-Bengal Academy of Science and Technology" (https://www.wast.in/) in 2016. Both of these Fellowships are in acknowledgement of my long-standing contribution to Science and Society.

After my Masters in Statistics, my research training during PhD had primarily been in the field of population and statistical genetics. We analyzed real data and also developed statistical models and computational algorithms, which were inspired by problems of population genetics as well as genetic epidemiology. Our study of analyzing the population diversity of India – encompassing mitochondrial DNA, Y-chromosomal DNA and autosomal DNA variations –was the first comprehensive reconstruction of the trails and dynamics of how anatomically modern humans(AMH) entered and populated the Indian sub-continent. Our work provided deep and fundamental insights on the population histories and affinities among Indian population groups. This paper in *Genome Research* is one of the most cited papers on Indian population genetics. In addition to providing support to aphorisms about Indian population diversity and structure, we showed that India has been populated by a small number of 'founding mothers'. We showed that the Austro-Asiatic speaking tribal populations are the autochthones of the India. Besides the widely studied population migrations through North-Western India, this study reinstated the importance of the North-East corridor for ancient migrations of people into mainland India. We also postulated that Dravidian speakers were widespread throughout India and were possibly pushed to the southern region by the entry of Indo-European speakers from South-Central Asia. The major findings of this study have stood the test of time and it were later substantiated by larger population genetic studies.

During the conduct of the large-scale genome diversity study, I also devised novel statistical methods for identifying genomic signatures in the DNA isolated from members of population groups [*Genome Research* 15: 67-77]; which were applicable in finding epistatic interactions in case-control study designs as well. We made a comparative evaluation of methods for estimating the antiquity of a population [*Journal of Genetics* 82: 7-12] and also inferences about coalescent times in populations.

During my post-doctoral tenure at Stanford University and University of California San Francisco, I developed statistical methods for admixture analysis and expanded my interest to genetic epidemiology. We devised a novel approach to detect signatures of natural selection in admixed populations [*Human Genetics* 124: 2008]. We also carried out multiple studies on mapping disease genes, including devising innovative techniques of gene-mapping using admixed populations. Our results are suggestive of why even after adjusting for socioeconomic conditions, does people with African ancestry generally have healthier lipid levels than people with European ancestry [*Human Molecular Genetics* 18: 2009]. We applied our methods on genotype data to find genes related to metabolic disorders [*Human Genetics* 124: 2008, *Obesity* 17: 2009, *Human Molecular Genetics* 18: 2009]. Our investigation was not confined to disease gene mapping. Using novel methods we

characterized the intricacies of admixture in Hispanics [*Genome Biology* 10(11): 2009] and African-Americans [*Genome Biology* 10(12): 2009]. Among African-Americans sampled from different parts of United States, we have shown that there is no major spatial variation of African ancestry. We have also shown that, because the trans-Atlantic slave trade largely transported western African populations into United States, the ancestry estimates of African-Americans in United States are almost independent of the ancestral population chosen, and thus there is very little population structure in their African ancestry to be worried about when conducting a case-control design [*Genome Biology* 10(12): 2009]. Our study among the Hispanics revealed that inferences about social behavior like preferences in marriage and the effect of the ascertainment on the genomes of populations can be confidently deduced from genomic data.

Besides studying admixed populations of the new world, I was also engaged in large-scale studies in genetic epidemiology, aiming to map genes for cardiovascular diseases [*Human Molecular Genetics* 17: 2008, *Human Genetics* 123: 2008, *Atherosclerosis* 198: 2008]. We showed that the genetic component of High-Density Lipoproteins (HDL) is strong and is more difficult to alter by changing lifestyle, like diet. We also showed that there are genomic regions (may as well be the same set of genes), which simultaneously impact different lipid traits, particularly HDL and Triglyceride(TG). Using linkage analysis we found a region on chromosome 8 which was strongly related to absolute pitch in humans [*American Journal of Human Genetics* 85: 2009].

After I returned to India and joined the National Institute of BioMedical Genomics (NIBMG). I continued to carry out studies on both genetic epidemiology and population genetics. Genomewide Association Studies were gaining popularity in India and I was engaged will multiple groups in investigating the association of single-nucleotide-polymorphisms with different genetic diseases like Primary Open Angle Glaucoma [*PLoS One* 5: 2013, *BMC Medical Genomics 9:* 2016], the angiogenic diabetic retinopathy [*Retina* 32: 2011, *Molecular Vision* 18: 2012] Cardiovascular Diseases and Diabetes [*Diabetes* 62: 2013, *Indian Journal of Medical Research 2020, Biomolecules, 9*(8), 2020; *Journal of Human Genetics* 64(6), 2020, *Journal of Genetics* 98(1), 2020, *Molecular Genetics and Genomics.* 292(3): 2020]. We did statistical inferences from the largest genome-wide data set on type 2 diabetes mellitus collected in India that has identified a new susceptibility locus and has been published in the renowned journal *Diabetes* [*Diabetes* 62: 2013]. Subsequently, we have utilized these large-scale genomic data to initiate molecular pharmacogenomics in the population level in India [*Pharmacogenomics* 15(10): 2014, *Pharmacogenomics* 15(5): 2014]. With Dr Mitali Mukherjee's group, we estimated that Indian Siddi populations are admixed with Africans who traveled east ward and they have approximately 60% African ancestry. We also showed that, owing to their recent admixture, high proportion of both the North-Indian and the African ancestral component and relatively small population size, the Siddi population is a wonderful candidate to initiate studies of admixture mapping. This work attracted major global attention. Not only has it been published in the *American Journal of Human Genetics* [*American Journal of Human Genetics* 89: 2011], but it was also selected for platform presentation in three of the most high-profile international meetings in Human Genetics -- the American Societies of Human Genetics (ASHG), European Society of Human Genetics (ESHG) and the Human Genome Organisation(HuGO). More recently, aligned with

the primary focus of NIBMG, some of my recent research contributions has been in the field of cancer research [*Scientific Reports* 5:2015, *PLoS One* 10: 2015, *PLoS One* 9: 2013, *Nature Communications* 4: 2013, *Mitochondrion*. 2015, *Mitochondrion*. 2019].

One arm of my research has always been in statistical modeling and algorithm development. We designed Bayesian semi-parametric model that we have devised outperforms the globally popular program, STRUCTURE in terms of computational time, without any compromise in accuracy [*Biometrics* 16: *2012]*. With Dr Anil Ghosh of the Indian Statistical Institute, we developed methods to identify hidden subpopulations within clusters [*Statistical Methods and Applications* 24: 2015].

In order to understand in greater detail the population diversity and evolution in the Indian-subcontinent, we systematically explored genome-wide DNA variation in about 400 unrelated Indians belonging to 20 ethnic groups. This is the largest DNA variation study conducted in India, both the number of ethnic groups (20 groups) and the number of DNA variants (over 1 million variants) examined on each individual.

We have corrected the conclusion of an earlier major study – published in 2009 (Reich et al *Nature* 461, 489-494) – that modeled the population history of India and concluded that the present-day Indians are derived from two ancestral groups of people, one of whom is ancestral primarily to all north Indians and the other ancestral south Indians. We have been able to provide robust evidence that four – not two – ancestral groups contributed to the genetic diversity of present-day mainland Indians [*Proceedings of the National Academy of Sciences* 113: 2016]. These ancestral groups are roughly identifiable with the four language families in India – Indo-European (north India), Dravidian (south India), Tibeto-Burman (north-east India) and Austro-Asiatic (fragmented in east and central India; spoken exclusively by the tribals). We also identified a fifth ancestral lineage that is dominant among the hunter-gatherer tribals of Andaman and Nicobar Islands (Jarawa and Onge). We also found evidence that this lineage is also ancestral to the present-day Pacific Islanders. We identified the ancestry predominant among the Indo-European speakers to be co-ancestral to South-Central Asia and the ancestry predominant among the Tibeto-Burman speakers to be co-ancestral to East and South-East Asia; but the ancestor groups or origin of the Austro-Asiatics and the Dravidian tribals remain unidentified. We have also shown that inter-marriage without major restriction was replaced by the formulation and declaration of social norms, leading to the formation of endogamous groups about 70 generations ago; probably during the Gupta Empire.

We have compared the genomic spectrum of populations of India and showed its uniqueness, under-representation and importance in understanding global diversity [*Genome Biol Evol.* 8 (11): 2016, *Indian Journal of History of Science: 2016, Journal of Biosciences: 2019*]. This discourse culminated in efforts to understand and catalogue the genetic variation in the (a) Indian-subcontinent, or broadly in the (b) Asian continent. In order to achieve (a), the Department of Biotechnology (DBT), has initiated a multi-institutional, multi-centric effort, known as “Genome India”; while to achieve (b), there is an international public-private effort: “GenomeAsia 100K”, where investigators from NIBMG have played lead roles [*Nature* 576: 2020].

Recently, my laboratory has also expanded in studying the impact of microbiome in disease

and development [Microbiology Spectrum: 2023; NPJ Biofilms and Microbiomes, 2022]. We have conducted a randomized-control-trial to access the impact of probiotics on Diabetes. On the theoretical method development, we have successfully implemented machine learning methods to identify regions in the genome which has undergone natural selection as well as extracting intricate details from images in Diabetic Retinopathy. We believe, all of these have immense translational potential.