Aging Pathobiology and Therapeutics 2023; 5(4): 159-161 ***159***
DOI: 10.31491/APT.2023.12.130

On epigenetic clocks and cancer risk—an interview with Prof. Andrew E. Teschendorff

Shikha Sharma[a,*]

[a] *Institute for Stem Cell Science and Regenerative Medicine, Bangalore, India.*

Abstract

The article is an interview with Prof. Andrew E. Teschendorff of the CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Shanghai, China, conducted by Shikha Sharma of the Institute for Stem Cell Science and Regenerative Medicine, Bangalore, India, on behalf of *Aging Pathobiology and Therapeutics*.

Andrew E. Teschendorff, PhD

Andrew E. Teschendorff is a principal investigator at the Shanghai Institute of Nutrition and Health of the Chinese Academy of Sciences. He develops and applies advanced statistical/computational methodology to help analyze and interpret complex multi-omic data, with a particular focus on applications in aging, cancer risk prediction, and single-cell systems biology. He has a long-term interest in using computational and statistical approaches to improve our systems-biological understanding of aging and oncogenesis, and how to translate these insights into novel strategies for cancer risk prediction (https://aeteschendorff-lab.github.io/team/andrew/).

* Corresponding author: Shikha Sharma, PhD
Mailing address: Institute for Stem Cell Science and Regenerative Medicine, Bangalore 560065, India.
Email: shikha.ss19254@gmail.com
Received: 22 November 2023 / Accepted: 29 November 2023
Published: 26 December 2023

Shikha Sharma: You obtained your PhD in physics in 2000, where you did wonderful work on intersecting branes, calibrations, and supersymmetry, and you continued to work in the same area for some years, and then later you transitioned your expertise to analysis in the biological field with your breakthrough work in 2010 on DNA methylation profiling to predict cancer risk and its association with aging in a very short time. What made you transition from calibrating branes to calibrating methylation sites in stem cells and cancer models?

Andrew Teschendorff: Thanks for the question. Towards the end of my PhD in superstring theory, I realized that continuing in this area of theoretical physics would be extremely risky and potentially also very disappointing, since it is a very abstract subject marked by the complete absence of data. I could not see myself working all my life on "science" that could not be tested experimentally. So, after my PhD, I spent 1 year at the Complexity Research Lab of British Telecom modelling data-traffic in 3G networks, which sparked my general interest in applied statistics. However, I also longed to return to academia, and so in 2001 I took up a Research Assistant position in Mathematical Biology at the University of Warwick. This was the time when the first big microarray studies were being published, and I immediately realized that it would be extremely exciting and satisfying to work in human genomics, especially in trying to understand the molecular basis of a deadly disease like cancer. I felt extremely motivated if I could join the international effort to "beat cancer". I was fortunate enough to be offered a job by Prof Carlos Caldas at Cambridge University to work in his Breast Cancer Genomics Laboratory, and indeed this was the job that really helped kickstart my career. During this time in Cambridge, my research was very much focused on developing novel molecular classifications of breast cancer based on gene expression and copy-number variation. I soon felt though

that something important was missing in these efforts. The importance of epigenomics in cancer was becoming ever more apparent and so I moved to UCL in London to join various experts working in cancer epigenomics.

Shikha Sharma: Your work (Genome Research 2010) became a pioneer for the development of epigenetic aging clocks that are widely available to predict aging in humans, animals, different tissues and different disease conditions, which is a vast field now. When you were working on the prediction of DNA methylation sites in cancer, stem cells and normal tissues as well as associated with aging, have you thought at that time that your work would create a founding milestone in the development of the epigenetic clock field? How do you weigh the epigenetic clock field now? What do you think still needs to be done in this field? In 2016, your lab demonstrated epiTOC2, a novel mitotic clock that can measure the stem cell division for better mitotic age estimation and cancer risk prediction. Is your lab still working on developing more epigenetic clocks?

Andrew Teschendorff: Yes, the work we published in Genome Research in 2010 was important for at least two reasons. It was one of the first studies to use the Illumina Infinium DNA methylation bead array technology, and by profiling many hundreds of blood samples from ovarian cancer cases and controls, we accidentally discovered an age-associated DNA methylation signature that also displayed associations with chronological age in many other normal tissue types, including even mesenchymal stem cells. In effect, it laid the groundwork for Steve Horvath to develop his pan-tissue epigenetic clock. Our study was also important in demonstrating that this age-associated DNA methylation signature is enriched for sites marked by repressive marks in stem cells, many of which encode transcription factors that are important for differentiation. It had previously been hypothesized that DNA methylation changes at these differentiation factors could causally contribute to cancer development, so demonstrating that these changes happen in normal cells as a function of the most prominent cancer risk factor (*i.e.*, age) was extremely exciting. Moreover, we went on to show that these DNAm changes were also seen in precancerous lesions. At the time, my colleagues and I were therefore entirely focused on what this DNAm signature could mean for cancer development, and hence we did not realize the significance of our findings for developing a DNA methylation-based predictor of chronological age. That is why we never worked on developing a clock, although we knew from our data that it would be possible to predict age fairly accurately. It took the genius of Steve Horvath to realize that predicted deviations from chronological age may be informative biological age. Looking back at it, I am very happy that the 2010 Genome Research paper played such an important role in the development of the epigenetic clock field, which has now blossomed into one of the most promising research areas in epigenomics and aging. Indeed, some of the most exciting recent works hint at how DNA methylation changes may play a critical role in rejuvenating cells. Building on our 2010 work, we have continued to explore the significance of age-associated DNAm changes in the context of cancer-risk, developing a number of epigenetic mitotic clocks that yield proxies for mitotic age. Here mitotic age refers to the cumulative number of stem cell divisions in a tissue, which is believed to be a determinant of cancer-risk. Our work has shown that DNAm is an ideal molecular substrate for measuring mitotic age in normal tissues before they turn cancerous, thus offering new avenues for cancer risk prediction. We are currently trying to improve upon the existing epigenetic mitotic clocks, as many challenges, including statistical/computational ones, remain.

Shikha Sharma: In 2013, you took another transformative turn and invented the BMIQ normalization algorithm, and in 2017, you co-led the development of the ChAMP R Bioconductor package, and in the same year came up with the proposal to use diffusion network entropy for data analysis. Meanwhile, you kept working on methylation profiling, and between 2012 and 2016, you came up with the idea of using differential DNA methylation variance to identify cancer risk from precancerous lesions. From 2012 to now, you have developed various novel algorithms, including EVORA, SEPIRA, CELLDMC, EPISCORE, and DICE for methylation profiling. What drives you to do this innovative and transformative work?

Andrew Teschendorff: Well, the BMIQ normalization algorithm was developed out of necessity, since there were a number of biases in the DNA methylation data generated by the newer Infinium bead array technologies. Normalization is always a critical part of analyzing any type of omic data, so we have spent a substantial amount of time over the years developing tools and software like ChAMP to help streamline such normalization analyses. To be honest, developing normalization methods is not the most exciting area of statistical bioinformatics, and yet, ironically, such work is often the most cited. In parallel to the above work, we have remained very active in developing other algorithms to tackle other statistical challenges. For instance, the CellDMC and EpiSCORE algorithms tackle the challenge of cell type heterogeneity, as most DNA methylation data is still generated in bulk tissue. Given that in the foreseeable future single-cell DNAm technologies will remain unscalable and too costly, cell type heterogeneity remains one of the biggest challenges facing the Epigenomics field, hence why we continue to be active in this field. Identifying cancer-risk markers from DNAm data is also challenging because evidence points towards the earliest DNAm changes being highly stochastic: this is why we developed an algorithm like EVORA, that performs feature selection using an entirely different paradigm based on the concept of differential variance, and that is tailored to capture this underlying stochasticity. I personally consider the EVORA work to be one of my most significant ones, both from a biological/clinical as well as statistical standpoint. Indeed, EVORA represents a radical departure from the traditional paradigm of selecting features based on differences in average DNAm. Alongside our work on DNAm data, it has been equally

important to embrace the single-cell data tsunami, as the high granularity of such data offers unprecedented opportunities for the cancer-risk prediction field. Indeed, one main focus has been on developing a network theoretical framework called diffusion entropy and practical tools for estimating stemness, including cancer stemness, in preneoplastic cell populations. Preliminary data indicates that it may be possible to stratify single preneoplastic cells by their cancer-risk and so we are actively pursuing this research direction. One key lesson learned from developing all these algorithms is that such innovation often relies on getting your hands on novel and unique data.

Shikha Sharma: You have worked on various cancer models and their methylation profiles. Have you observed any common characteristics between these cancers based on epigenetic signature data? Is it possible to use these epigenetic signatures to detect the risk of occurrence of cancers? What do you think epigenetic alteration is a slow or fast process in cancer? Is it possible to stop cancer growth by correcting the methylation signature predicted by the epiTOC2 mitotic clock?

Andrew Teschendorff: An important insight we have gained from studying DNAm patterns across many cancer types is that a significant proportion of these DNAm changes are associated with cell division and that they are shared between all cancer types. This is not entirely surprising since an increased rate of cell proliferation is a common cancer hallmark. Thus, many of the DNAm changes observed in cancer are just a consequence of this increased proliferation rate. This is why it becomes paramount to measure DNAm changes in precancerous lesions or in normal tissues exposed to cancer risk factors. In 2012 and later in 2016, we published two proof-of-principle studies demonstrating that DNAm has the potential to predict cancer three years before diagnosis. In 2012, this was done in the context of a prospective study nested within the ARTISTIC trial, which profiled DNAm in histologically normal cervix from healthy women, with half of whom developed high-grade cervical intraepithelial neoplasia 3 years later. This demonstration relied critically on the aforementioned EVORA algorithm. Epigenetic mitotic clocks could also be valuable in this context. The hope is that such cancer-risk prediction algorithms could serve as a means to prevent cancer development or to triage patients for more frequent follow-up examinations so as to ensure that cancers are detected as early as possible. As to whether DNAm changes could be edited to stop or prevent cancer, this is still a very long way off, but we are working on trying to understand if early DNAm changes could be causally implicated. We have a number of hypotheses supporting the view that a small fraction of acquired DNAm changes could contribute causally by blocking differentiation and increasing aberrant cellular plasticity. It would be a dream come true if in the future it were possible to reverse the DNAm changes at these genes in vivo, for instance, via some fancy and safe future gene editing tool, and that such editing could stop, prevent or at least reduce the risk of cancer.

Shikha Sharma: You have accomplished remarkable achievements in your career. How do you assess your journey so far? What are your future goals and what are your other passions?

Andrew Teschendorff: Working as a computational biologist, it is not easy to drive an entirely in-silico research program that aims to address specific biological or clinical questions. I often find myself walking a tightrope, trying to keep a balance between developing novel statistical and bioinformatic methods on the one hand, whilst also trying to tackle the biological and clinical challenges. However, I feel that keeping this balance is critically important in order to overcome the biggest challenges we are facing. In this regard, my passion and research goal is certainly to help elucidate the systems-biological "unifying" principles of cancer development, and in doing so to also help develop improved tools for cancer-risk prediction, given that improved cancer-risk prediction and early detection have been shown to be the most efficient means of reducing the huge cancer-associated mortality burden. My feeling is that as molecular genetics becomes ever more quantitative, that there is going to be an ever increasing need to apply concepts and methods from more quantitative disciplines, such as engineering, materials science, AI, and physics, in order to crack the biggest challenges of complexity in molecular biology research.

Shikha Sharma: Thank you for your time.