



HAL
open science

Epigenomic technologies: an interview with Jorg Tost

Jorg Tost

► **To cite this version:**

Jorg Tost. Epigenomic technologies: an interview with Jorg Tost. *Epigenomics*, 2022, 14 (6), pp.359-363. 10.2217/epi-2022-0005 . cea-04352535

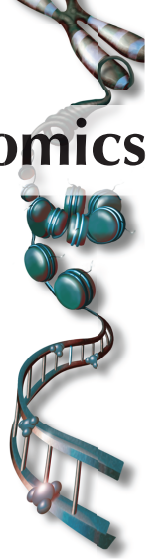
HAL Id: cea-04352535

<https://cea.hal.science/cea-04352535>

Submitted on 19 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Epigenomic technologies: an interview with Jörg Tost

Jörg Tost^{*},¹ 

¹Laboratory for Epigenetics & Environment, Centre National de Recherche en Génomique Humaine, CEA-Institut de Biologie François Jacob, Evry, 91000, France

^{*}Author for correspondence: tost@cng.fr

Jörg Tost received his PhD in genetics from the University of Saarbrücken (Germany) in 2004 for devising novel methods for the analysis of haplotypes and DNA methylation patterns. After a postdoctoral stay in the technology development department of the Centre National de Génotypage (Evry, France), he led the Epigenetics groups from 2006 to 2012, before becoming the Director of Laboratory for Epigenetics and Environment at the Centre National de Recherche en Génomique Humaine (CNRGH). The laboratory is involved in the development and application of technologies to analyze DNA methylation, miRNAs and other epigenetic modifications quantitatively at high resolution at target loci and genome-wide using state-of-the-art sequencing technologies as well as the development of bioinformatic tools for the processing of such data. The laboratory mainly focuses on the analysis of epigenetic changes in neurodegenerative, autoimmune and inflammatory diseases as well as the alteration of the epigenetic profiles in function of environmental exposure. A second research axis investigates novel technologies for the analysis of mutations of clinical relevance present at very low proportions in the analyzed samples and their impact on treatment management. Tost has an H-index of 50 and is the author or co-author of more than 195 publications.

First draft submitted: 24 November 2021; Accepted for publication: 5 January 2022; Published online: 28 January 2022

Keywords: DNA methylation • epigenetics • epigenome

Your expertise in developing & applying DNA methylation technologies is renowned within the community, and what has been the most exciting project that you have ever worked on?

I think these are clearly the different applications of the pyrosequencing technology, which we adapted for DNA methylation analysis about 20 years ago. It's now a very widely used technology and has shown in different benchmarks to be among the easiest to use, accurate, fastest and quantitative DNA methylation technologies. The repurposed pyrosequencing technology can be used for locus-specific DNA methylation analysis, biomarker analysis and so on.

We also extended the use of the technology to more specialized applications such as allele-specific DNA methylation and single molecule DNA methylation analysis as well as DNA methylation analysis in plants, which have additional cytosine modifications in other sequence context. Also more recently, we have used the pyrosequencing technology to enrich specifically disease-related methylated molecules in liquid biopsy using a modified version of Enhanced-ice-COLD PCR [1]. So, there really is a broad range of applications for this technology in the DNA methylation field.

I think, of course, DNA methylation analysis also comes with bioinformatic tools, and we published our BeadChip Array Analysis Pipeline by making use of the functional annotation of the CpG probes nearly 10 years ago, which was quite innovative at that time [2]. Unfortunately, our approach didn't scale up to the newer generation of the array, so we had to abandon this, but nonetheless I can say we are proud of this work from a historical point of view.

DNA methylation-based biomarkers have now been approved for screening colorectal cancer, and when can we expect the same for other common cancers & what challenges stand in the way?

First, it was a really critical step to have some of these DNA methylation-based markers pass regulatory approval, which for the first of its kind, always takes time. The novelty of the technology also deterred companies and research groups away from taking on this challenge. On the other hand, if you look at the literature, probably every gene in the genome has been associated with altered methylation and has been identified as a potential biomarker for a disease.

There are numerous data out there, but most of the studies we have seen so far are limited in size and are exploratory. I think there are currently only a limited number of studies, which are conducted with more rigor and robust validation steps. On the other hand, there are several applications where DNA methylation is actually already used, although perhaps not with all the regulatory approvals. For example, the open access brain cancer classifier at The German Cancer Research Centre, allows any pathologist or clinician to upload their BeadChip data and to classify or to subtype a brain tumor for improved diagnostic and patient management decisions. You do have, of course, the *MGMT* assays too, for predicting the response to temozolomide in glioblastoma.

Because DNA methylation is tissue specific, this property can be used for clinical applications. Several studies have shown that you can determine where a metastatic tumor comes from based on its DNA methylation profile even if you do not see the primary tumor. Therefore, this technology has the potential to change clinical routine and I think in the liquid biopsy field there are also many exciting developments. The recent acquisition of Grail by Illumina is one example. Grail is looking at circulating DNA and methylation profiles to define if there is cancer-related methylation somewhere in the genome and provides leads where the tumor might be. Strategies are now moving from individual cancer-specific markers to a more universal screen, which in theory could be implemented in clinical routine in the next five years or so.

There are also other companies like Volition, which don't look at DNA methylation, but at specific modifications of nucleosomes and histone modifications, which have a very similar approach, always making use of the tissue specificity. So, I think this field will expand in the next few years with a number of technologies coming out. The developmental phase and implementation of biomarkers is always lengthy, it may easily take up to 10 years. But I see for example, methylation of *PITX2* is coming up in breast cancer; marketed by a major life science supplier in Germany.

DNA methylation assays are starting to be used in clinical studies and showing robustness and clinical benefit, but their application is still in the research-oriented clinics, rather than in a routine clinical laboratory. To completely integrate DNA methylation into routine testing, we need to provide formats for biomarker analysis, which are completely compatible with the workflows based on quantitative PCR or amplicon sequencing, which are already implemented in the clinical diagnostic laboratories to make them acceptable.

I would say one of the challenges of DNA methylation-based biomarkers is the specificity. For example, *RASSF1A* is methylated in almost every cancer, during pregnancy and probably in several other diseases. But there are numerous papers that were claiming *RASSF1A* to be a good biomarker, but it's just universal. The same is true for microRNA expression, such as miRNA-21. There are probably 1000s of papers claiming this is a potential biomarker, but it's in fact a biomarker for hundreds of conditions which renders it less useful.

We have to accept that DNA methylation changes and microRNA expression are probably not going to yield a singular unique marker for disease. Perhaps, we need to be less dogmatic and stop thinking of diagnostic technologies solely as one test, but more of a combination of transcriptomic, proteomic and epigenetic data, including perhaps also the microbiome. This might be a way to get to the required specificity.

We want biomarkers to deliver results with a performance significantly above the predictive power of routine biochemical tests and the clinicians' diagnostics. So, these new technologies have to give us something more, and of course, this is not an easy task. To overcome this challenge, we need to include epigenetics, at least at an exploratory level, in clinical trials.

You have extensively researched the epigenetics of immune-related disease. What role does 'immunoepigenetics' play in inflammatory & autoimmune diseases & could this crosstalk also shed light on cancer mechanisms?

We have been looking at immune-related diseases from a pragmatic point of view because immune cells are easily accessible, which is much more complicated in, for example, neurodegenerative diseases.

Something, which is really fascinating me, is how the epigenome works as a memory to all the environmental exposure, and this has been really a central theme of my laboratory where we research many different exposures such as smoking or chemicals and organic pollutants. But also, let's say on a more positive level, for example what does a traditional farm environment do to your epigenome? What does it do to your immune system? We know that children born to a mother working on a farm have a reduced chance of developing allergies or asthma.

I think what is starting to emerge as a common theme in immune-related diseases is the plasticity of specific immune cell populations in these diseases. So you have some cell populations in an epigenetically altered state, there is evidence for this, for example in eosinophils in asthma. The innate immune system, upon vaccination, will leave an epigenetic mark on some of the cells of the innate immune system. I think this epigenetic altered state we see in immune-related diseases is something very fascinating. In one of our studies where we looked at food allergy, we could show that *Gata3*, the key transcription factor which defines the cellular identity is methylated in type II T-helper cells, which are pro-allergenic.

For cancer, the immune system, of course, plays a major role. We do see immunotherapy bringing much progress to cancer patient management. Recently, there was a study showing that cytotoxic CD8⁺ T-cells expressing granzyme B and K were predictive of the response to immunotherapy, and I am very much convinced that the epigenetic regulatory landscape in these immune cells might be different between immunotherapy responders and non-responders. This is something we will investigate in the next few years in more detail. But we must have highly detailed data on the immune populations, because just CD4 cells are still a heterogeneous mixture, in order to detect probable changes in subpopulations which might be important for such responses. Bioinformatic strategies and multi-omic single cell technologies will need to be utilized in this study. There are more and more single cell technologies becoming accessible for epigenetic modifications, although more on the chromatin level, not yet on the DNA methylation, but that is something to look forward to.

To understand the crosstalk between the immune system and the cancer, we also have to consider the microbiome. We understand that several epigenetically active metabolites are actually created by the microbiome, but we don't really understand how this translates into changes of the immune system. On the other hand, we know that if the microbiome is dysregulated that this might lead to autoimmune, inflammatory and allergic diseases. For the moment this crosstalk is mainly correlative; we see changes in specific populations of the microbiome producing for example short chain fatty acids or polyunsaturated fatty acids, and we know that these are epigenetically active. But we are not yet able to make a direct functional link between these changes and the occurrence of the disease. This gives an interesting therapeutic handle because there is a lot of evidence now in different models, that application for example of histone deacetylase inhibitors will slow down or even prevent the atopic march (the successive development of different allergic diseases) and perhaps even topical intervention of these epigenetically acting therapeutics might have an impact in the future. Again, this is all on cellular and animal models for the moment.

In your 2020 editorial '10 years of Epigenomics: a journey with the epigenetic community through exciting times', you expressed the potential of spatial transcriptomics in epigenetic modification analysis [3], can you please expand on this?

In this editorial, I referred to the potential of spatial analysis as we know that epigenetics is something which is stable because it defines the cell type, but it's also dynamic, because it responds to the environment. We always think of the environment as something external, but it's also very likely that it's also reliant on the position of cells. So cells have a different epigenome depending on their position where they are in a tissue and where they are in the body. At the moment, we can investigate highly pure cell populations obtained through multiparametric flow cytometry or magnetic sorting. We can drill down even more into these cell populations using different single cell technologies that we have just alluded to.

Let's take the example – a T cell, which is very close to a cancer cell – might have some changes in its metabolism and in its epigenome compared to a T cell, which is a few centimeters away and thus is not exposed to the same signaling molecules and same cytokine environment. We need to focus not only on specific cell populations, but also on where exactly these cells are. I am not yet aware that epigenetic modifications can be easily analyzed using spatial analysis. We do see a lot of development of course in the RNA and protein expression field where these have become rapidly adapted technologies with many different suppliers on the market.

I think this is crucial at least for DNA methylation; for histone modifications, this will a little bit more complicated and it might be difficult to get the granularity we need. DNA methylation analysis might respond better to our

needs to really understand the immune cancer crosstalk compared to transcriptomics and proteomics, but at the moment with the tools we have, we cannot really disentangle those.

Your team's methodology paper 'Analysis & annotation of DNA methylation in two nonhuman primate species using the Infinium Human Methylation 450K & EPIC BeadChips' has received significant attention from the community [4], and what do these annotations mean for future research & progression in the field?

I think the BeadChip technology is really becoming the workhorse of the epigenetic community because they are relatively cost effective. You can perform the analysis on a large number of samples, and you can give them to most technological platforms able to use genotyping chips. A point not to neglect is that there are a lot of very good bioinformatic pipelines that do exist for this. There are many compilations and different online repositories where you find significant information on different cell types, and you can compare those to your samples if you are interested in specific CpGs or specific diseases. There are probably hundreds of 1000s of samples which have been processed on the 27K, 450K and now EPIC BeadChip, but it's all only humans.

Several studies have tried to use this for mouse, but in the end only about 7–8% of the probes are reliable. I doubt that this is a very cost-effective tool, but it might still be an opportunity for people who don't have access to sequencing or the resources to do so. On the other hand, the pandemic has reinforced the use of other preclinical models, notably non-human primates, which are a tool of choice for neurological applications and also infectious diseases where the mouse is of relatively little value.

So, in this study [4], we show that in two widely used non-human primate species, the rhesus macaque and the African green monkey, we can actually use 30–40% of the array to give reliable data. This may also be a useful technique in COVID-19-related research to study DNA methylation and epigenetic memory. Creating the tools is really the prerequisite to advance research on disease and developmental questions.

Bioinformatics & *in silico* analysis of the epigenome are rapidly gaining traction, and how can we best manage & utilize these big datasets?

I think one of the major challenges is that although we do make a large number of datasets publicly available, it's really hard to reproduce the results. This is not something related to the conclusions of the authors, it's rather that the documentation on how the results were obtained is not sufficient. We need better documentation on which version with which parameter and what tool was used. For example, where we can have access to the raw data, for example the IDAT files from the BeadChips, the samples can often be re-used for meta-analysis, while for those where only pre-processed methylation values are available you can still use them as an independent validation of your data, but it's actually very challenging to use them in a combined analysis.

Then, there is a need also to make the datasets available. If this is covered by the informed consent of the patients that we have analyzed, then that's something one should keep in mind, especially for large-scale data. It is possible to obtain genetic information from epigenetic data especially sequencing-based data, so it should be considered as sensitive data and has to be treated accordingly, but still whenever its possible data should be made available.

We should raise the bar on the quality of re-analysis, notably International Cancer Genome Consortium/The Cancer Genome Atlas data, to really question the novelty of the study. I think just an analysis of DNA methylation or microRNAs will not be sufficient to answer biomarker questions, so we should be able to integrate, miRNA, genetic, epigenetic and metabolomic data. This is challenging because what do you take as a minimal nomenclature? Correlating with gene expression on an individual level is a very simplistic view and we know for example miRNAs can potentially target hundreds of genes. Likewise, reporting methylation at a specific site and reducing the association to a gene name might lead to the loss of important information as the CpG site may also be a regulatory region or regulate a transcription factor, which will have an impact on multiple genes. So, there is a complexity that I don't think we yet are able to fully cope with.

Because of the tissue specificity, most analyses are still conducted in limited numbers of samples, except perhaps cancer. The datasets are just too small to be used in advanced algorithms such as machine learning. Although I see quite some hype about deep learning and neural networks, if you look at the data, in many cases a simple logistic regression or random forest analysis might give you the same results with less computational effort. The datasets need to be much bigger, but we also need to have all the phenotypic information.

I think, in general, the idea of including epigenetics in clinical trials, which are gathering data on multiple molecular levels to evaluate the efficiency of a treatment including epigenetic modifications on different levels is

crucial. Another challenge I think is that at least for whole genome bisulfite sequencing or even oxidative bisulfite sequencing which is even worse, the cost of these analysis is still very high, and we can't easily work with algorithms for imputation as we do in genetics. There have been a number of algorithms devised for imputation which do work, but I think there is still room for improvement.

Data integration to get a grip on the big picture will be the challenge for the next decade, I think. We will need samples, which have been specifically collected for epigenetic analysis, which if possible, are well-characterized in their composition to account for cellular heterogeneity. And of course, we need an increased willingness of people to devote their precious clinical samples to specific projects. Of course, it's always a question which is the project and which should be prioritized. There is a lot of discussion in the field of ethics, who actually owns the samples that are collected? Will it be the patient or who should be able to decide for which use the samples should be reserved? I think the field has made significant progress on the collaborative approach, but this could still be improved. On the other hand, one also has to acknowledge that the ethical requirements do make collaboration on the international level in many cases challenging.

Is there anything the readers should keep an eye-out for?

For the next few years, I think we will continue our work in the allergy field, and we will reinforce the work in the immuno-oncology field. The current data make it seem unlikely that the tumor itself will contain the important clinical information on how people will respond to therapies. So, my lab will rather look at the immune-cell populations and try to understand if there are epigenetic changes that we can exploit for moving on to a more personalized approach for cancer management. Similarly, in the field of inflammatory and allergic diseases, the analysis of epigenetic modifications might be a good option to predict allergen immunotherapy. In the field of immune-related diseases, we are still way behind the cancer field where epigenetics research has been much more intensive and longer ongoing than in other complex diseases. One should not forget that the immune-related diseases do not have in most cases a strong a genetic driver as cancer has, so epigenetics might be even more important in these diseases.

Financial & competing interests disclosure

J Tost is a member of the *Epigenomics* editorial board and receives an annual honorarium for his contribution. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Interview disclosure

The opinions expressed in this interview are those of J Tost and do not necessarily reflect the views of Future Medicine Ltd.

References

1. Mauger F, Kernalguen M, Lallemand C, Kristensen VN, Deleuze J-F, Tost J. Enrichment of methylated molecules using enhanced-ice-co-amplification at lower denaturation temperature-PCR (E-ice-COLD-PCR) for the sensitive detection of disease-related hypermethylation. *Epigenomics* 10(5), 525–537 (2018).
2. Touleimat N & Tost J. Complete pipeline for Infinium[®] Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 4(3), 325–341 (2012).
3. Tost J. 10 years of *Epigenomics*: a journey with the epigenetic community through exciting times. *Epigenomics* 12(2), 81–85 (2020).
4. Pichon F, Shen Y, Busato F *et al.* Analysis and annotation of DNA methylation in two nonhuman primate species using the Infinium Human Methylation 450K and EPIC BeadChips. *Epigenomics* 13(3), 169–186 (2021).

