

## CANCER GENOMICS

## A fluid route to gene expression

Liquid biopsies — such as the sequencing-based analysis of cell-free DNA (cfDNA) circulating in blood — have garnered increasing interest in oncology as a minimally invasive method to obtain information on cancer-associated genetic and epigenetic aberrations. A new study suggests that sequencing of plasma cfDNA can also yield functional information, pinpointing which genes are expressed and which are silent.

Previous reports have shown that nucleosome occupancy differs at actively transcribed versus inactive promoters. Ulz *et al.* hypothesized that this is reflected in differences in read-depth coverage and thus set out to determine whether whole-genome sequencing of plasma DNA could predict gene expression.

The team sequenced plasma DNA from 104 healthy donors and compared read-depth patterns at the transcription start sites (TSSs) of 3,804 housekeeping genes with those of 670 unexpressed genes. Indeed, the coverage pattern was found to reflect nucleosome organization. Actively transcribed TSSs exhibited depleted coverage (representing a reduction in nucleosome occupancy), flanked by regions

of wave-like patterns of peaks (indicative of well-positioned nucleosomes). These patterns mirrored those previously established with micrococcal nuclease (MNase) assays. By contrast, read-depth coverage at TSSs of inactive promoters was increased, reflecting a region with tightly packed nucleosomes.

The authors then focused on a lymphoblastoid cell line and compared existing MNase assay data with plasma read-depth data for the 1,000 (representing 1,334 TSSs) most highly expressed and the 1,000 (representing 1,109 TSSs) least expressed genes in blood, as determined on the basis of published plasma RNA-sequencing (RNA-seq) data. The maps were similar for both approaches, and read-depth patterns differed depending on the level of gene expression.

Analyses of plasma DNA read depth identified two discrete genomic regions that differed between expressed and silenced genes: one based on the reduction in nucleosome occupancy in the 2,000 bp region centred on the TSS (2K-TSS), and a second region mapping to the most frequent position of the nucleosome-depleted region (from –150 bp to +50 bp with

respect to the TSS). Statistical analysis of the coverage patterns at these two regions combined with machine-learning approaches enabled the authors to predict the expression status of the 100 most highly and least expressed genes with a sensitivity and accuracy of 0.91.

Next, Ulz *et al.* analysed primary tumour samples and plasma DNA samples from two patients with metastasized breast cancer, comparing copy number alteration profiles as well as expression patterns (obtained by RNA-seq in the primary tumour and correlated with plasma DNA promoter coverage). These analyses revealed that the read-depth patterns of sequenced cfDNA were able to correctly classify expressed cancer driver genes and determine the expressed isoform of genes with several TSSs.

Finally, the authors analysed 426 additional plasma samples from patients with cancer, showing that more than half of these samples would be suitable for nucleosome promoter analysis, thus highlighting the broad applicability of their approach.

Linda Koch

**ORIGINAL ARTICLE** Ulz, P. *et al.* Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* <http://dx.doi.org/10.1038/ng.3648> (2016)

**FURTHER READING** Waldron, D. A nucleosome footprint reveals the source of cfDNA. *Nat. Rev. Genet.* **17**, 125 (2016)