Characterizing intergenic transcription at RNA polymerase II binding sites in normal and cancer tissues

Pierre de Langen 1, Fayrouz Hammal 1, Elise Guéret 1, Jean-Christophe Mouren 1, Lionel Spinelli 1, Benoit Ballester 1

PMID: 37868033 PMCID: PMC10589727 DOI: 10.1016/j.xgen.2023.100411

Abstract

Intergenic transcription in normal and cancerous tissues is pervasive but incompletely understood. To investigate this, we constructed an atlas of over 180,000 consensus RNA polymerase II (RNAPII)-bound intergenic regions from 900 RNAPII chromatin immunoprecipitation sequencing (ChIP-seq) experiments in normal and cancer samples. Through unsupervised analysis, we identified 51 RNAPII consensus clusters, many of which mapped to specific biotypes and revealed tissue-specific regulatory signatures. We developed a meta-clustering methodology to integrate our RNAPII atlas with active transcription across 28,797 RNA sequencing (RNA-seq) samples from The Cancer Genome Atlas (TCGA), Genotype-Tissue Expression (GTEx), and Encyclopedia of DNA Elements (ENCODE). This analysis revealed strong tissue- and disease-specific interconnections between RNAPII occupancy and transcriptional activity. We demonstrate that intergenic transcription at RNAPII-bound regions is a novel per-cancer and pan-cancer biomarker. This biomarker displays genomic and clinically relevant characteristics, distinguishing cancer subtypes and linking to overall survival. Our results demonstrate the effectiveness of coherent data integration to uncover intergenic transcriptional activity in normal and cancer tissues.