

# Non-coding regulatory variations: the dark matter of pancreatic cancer genomics

Aldo Scarpa,<sup>1,2</sup> Andrea Mafficini<sup>1</sup>

Pancreatic ductal adenocarcinoma (PDAC) is the seventh cause of death for cancer worldwide and the third in the USA, where it is expected to become the second by year 2030. Unlike other cancers, little progress has been made when it comes to therapeutic options other than surgery, which is possible only for a small fraction (~20%) of patients presenting with localised disease.<sup>1,2</sup>

For the above reasons, large efforts have been undertaken to get a deeper understanding of the molecular alterations and their effects on cancer cells and tumour microenvironment, by exploiting both innovative disease models and high-throughput studies for genomic and transcriptomic profiling of PDAC.<sup>3,4</sup> In the meanwhile, genome-wide association studies, and the study of familial pancreatic cancer, have been looking for and found genetic variations associated to PDAC onset and outcome.<sup>5,6</sup>

At the genomic level, mutations in the coding region of several genes have been consistently identified, together with disruptive structural alterations whose effect has been linked to the altered functionality (eg, *KRAS*, *TP53*) or loss (eg, *CDNK2A*, *SMAD4*, *ARID1A*, *ROBO2*, *BRCA1/2*) of the respective gene products. Frequent and rare alterations identified converge in specific pathways, such as Wnt/Notch, Hedgehog, axon guidance, transforming growth factor beta, SWI/SNF (SWItch/sucrose non-fermentable) and DNA damage repair.<sup>3</sup> RNA expression profiles, on the other hand, allowed the definition of at least four different subtypes of PDAC that correlate with histopathological characteristics and are characterised by differential activation of selected gene programs and enrichment for specific gene mutations.<sup>4</sup>

Despite the amounts of data generated so far, RNA expression profiles and DNA sequence variations have a nexus still not explored: variations in non-coding regulatory regions of the DNA. These may be either somatic (mutations) or germ line, defined as expression quantitative trait loci (eQTLs) and fall in the sequence of binding sites for proteins involved in DNA transcription or RNA regulation, thus affecting gene expression.

The abundance of non-coding variations and their indirect effect makes it more difficult to determine the actual biological significance for each of them. Most of genetic variations indeed, both in normal and neoplastic cells, lie in the non-coding fraction of the genome, which largely exceeds the coding part in size. For a long time, the non-coding DNA has been the genetic equivalent to the dark matter for cosmology: it was known to affect gene expression, yet it was difficult to determine its full impact due to the computationally intensive calculations needed to simultaneously process genomic and RNA expression data.

Large cooperative efforts such as the ENCYClopedia Of DNA Elements consortium (ENCODE [www.encodeproject.org](http://www.encodeproject.org)) have allowed to jump-start and carry over the characterisation of DNA regulatory elements in the human genome, while parallel projects like the Genotype-Tissue Expression (GTEx, [www.gtexportal.org](http://www.gtexportal.org)) have investigated how eQTLs affect gene expression in normal human tissues.<sup>7</sup> These efforts have led to knowledge accumulation on regulatory elements and to faster, more powerful analysis algorithms to identify non-coding regulatory variants from parallel genomics and transcriptomics data.<sup>8</sup> Taking advantage of this knowledge, the analysis of non-coding regulatory variations is now being carried on to elucidate which specific subsets impacts PDAC biology.

While a recent work by Feigin, Garvin and colleagues at Cold Spring Harbor Laboratory has tackled the study of somatic non-coding regulatory mutations and their effect on gene expression in PDAC,<sup>9</sup> the manuscript published in the present issue by Zhang and colleagues

from the National Cancer Institute Laboratory of Translational Genomics takes on the role of eQTLs on the expression profiles of normal pancreatic tissue versus PDAC.<sup>10</sup> Given that still computational power poses limitations to a full analysis, these works have focused on *cis*-variations, affecting DNA regions in proximity of the regulated genes.

Feigin and colleagues have found that regulatory regions of PDAC cells, particularly those with stronger effect, are enriched in somatic mutations. Interestingly, these frequently mutated regulatory regions target the expression of genes converging in the same pathways that were previously identified by the analysis of coding mutations and DNA rearrangements. Moreover, regulatory mutations seem to affect only genes whose behaviour has not been already altered by coding mutations. Low expression of one gene in particular, *PTPRN2*, associated with repressive regulatory mutations and frequent hypermethylation, was found to be an independent predictor of poorer prognosis.

While somatic regulatory mutations are in the DNA of cancer cells only, eQTLs modulate gene expression in all the cells of an individual, including cells of the normal pancreatic tissue and the non-neoplastic cells of the desmoplastic reaction often constituting the largest part of PDAC tissue. Zhang and colleagues analysed eQTLs in 95 normal pancreas, 76 of which adjacent to cancer, and in 115 PDAC. The identified eQTLs were enriched in regulatory elements important for exocrine pancreas specification. Despite a partial overlap of eQTLs, 23% of them resulted to be specific to PDAC tissue and 42% to the normal pancreas, while eQTLs identified in the latter showed a large overlap (82%) with available GTEx data from pancreas of healthy donors, indicating little if no effect of the adjacent tumour on the normal pancreas. This parallel study of eQTLs in both normal pancreas and PDAC gives a first idea of how the inherited modulation of gene expression changes when a limited portion of cancer cells (31% on average) in the analysed pancreatic tissue interacts with and perhaps drives the reorganisation of the surrounding non-neoplastic cells. An interesting open question is how much of this difference can be ascribed to the presence of neoplastic cells with their dysregulated gene expression, to the fact that non-neoplastic cell populations in the normal pancreas and in the desmoplasia are radically different and to the effect that cancer cells elicit on the surrounding tissue via molecular crosstalk.

<sup>1</sup>ARC-Net Research Centre, University and Hospital Trust of Verona, Verona, Italy

<sup>2</sup>Department of Diagnostics and Public Health, Section of Anatomical Pathology, University and Hospital Trust of Verona, Verona, Italy

**Correspondence** to Professor Aldo Scarpa, ARC-Net Research Centre, Policlinico GB Rossi, Piazzale L.A. Scuro, 10, Piastra Odontoiatrica (II floor), University and Hospital Trust of Verona, Verona I-37134, Italy; [aldo.scarpa@univr.it](mailto:aldo.scarpa@univr.it)

These analyses are thus only the beginning of a path that calls for further integration of genomics, epigenomics and transcriptomics data. Considering the organisation of PDAC tissue, one expected development would be the parallel analysis of sorted neoplastic and non-neoplastic cells to better understand how genetic variation influences the response of non-neoplastic cells to the stimuli of malignant ones.

**Acknowledgements** Supported by AIRC (12182).

**Contributors** AS and AM wrote the commentary.

**Competing interests** None declared.

**Provenance and peer review** Commissioned; internally peer reviewed.



**OPEN ACCESS**

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build

upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.



**To cite** Scarpa A, Mafficini A. *Gut* 2018;**67**:399–400.

Received 21 May 2017

Accepted 23 May 2017

Published Online First 28 June 2017



► <http://dx.doi.org/10.1136/gutjnl-2016-313146>

*Gut* 2018;**67**:399–400.

doi:10.1136/gutjnl-2017-314310

**REFERENCES**

1 Siegel RL, Miller KD, Jemal A, *et al.* Cancer statistics, 2016. *CA Cancer J Clin* 2016;**66**:7–30.

2 Rahib L, Smith BD, Aizenberg R, *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014;**74**:2913–21.

3 Waddell N, Pajic M, Patch AM, *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 2015;**518**:495–501.

4 Bailey P, Chang DK, Nones K, *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* 2016;**531**:47–52.

5 Roberts NJ, Klein AP. Genome-wide sequencing to identify the cause of hereditary cancer syndromes: with examples from familial pancreatic cancer. *Cancer Lett* 2013;**340**:227–33.

6 Amundadottir LT. Pancreatic Cancer Genetics. *Int J Biol Sci* 2016;**12**:314–25.

7 GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;**348**:648–60.

8 Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;**28**:1353–8.

9 Feigin ME, Garvin T, Bailey P, *et al.* Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma. *Nat Genet* 2017;**49**:825–33.

10 Zhang M, Lykke-Andersen S, Zhu B, *et al.* Characterising cis-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. *Gut* 2018;**67**:521–533.