Antisense long non-coding RNAs in breast cancer: A transcriptome-wide disruption

Stephane Wenric, Sonia El Guendi, Jean-Hubert Caberg, Warda Bezzaou, Corinne Fasquelle, Benoit Charloteaux, Latifa Karim, Benoit Hennuy, Pierre Frères, Joëlle Collignon, Meriem Boukerroucha, Hélène Schroeder, Fabrice Olivier, Véronique Jossa, Guy Jérusalem, Claire Josse, Vincent Bours

Why breast cancer?

- Most frequent cancer type in women (~35% of female cancers)
- First cause of cancer death in women (~35% of cancer deaths)
- ⅛ women will have breast cancer during their lifetime

- Breast cancer involves multiple genes
- Genetic alteration mechanisms are not always well known
- Some of these mechanisms involve antisense lncRNAs

Why antisense lncRNAs?

- Long non-coding RNAs (lncRNAs) = non protein-coding transcripts longer than 200 nt
- Antisense lncRNA or natural antisense transcript (NAT) = lncRNA
  - Sharing the same genomic location as a protein-coding gene
  - Transcribed in the opposite direction
  - Overlapping > 1 exon
- NATs
  - Regulate protein-coding gene expression
  - Overlap more than 50% of sense RNA transcripts
  - Have a lower expression than protein-coding genes
  - Can have an effect in *cis* or in *trans*

Study design

23 ER+/HER- breast cancer patients
- Tumors
- Adjacent healthy tissue

DNA → CGH, RT-qPCR, External microarray data → Validation
- RNA-Seq fold-changes vs. CGH (all genes)
- RNA-Seq fold-changes vs. microarray fold-changes (external study)
- RNA-Seq vs. RT-qPCR (subset of genes)
- Principal Components plot (tumors vs. healthy)

RNA → Strand-specific RNA-Seq → Reads alignment, QC, Gene quantification, Normalization

varRatio analysis:

$$tumoral\ read\ counts\ ratio = \frac{\sum tumor\ read\ counts_{antisense}}{\sum tumor\ read\ counts_{protein\ coding}}$$

$$normal\ read\ counts\ ratio = \frac{\sum normal\ read\ counts_{antisense}}{\sum normal\ read\ counts_{protein\ coding}}$$

$$varRatio = \frac{tumoral\ read\ counts\ ratio}{normal\ read\ counts\ ratio}$$

Differential correlation analysis (DGCA)

Differential expression analysis (DESeq2)

extreme values / p-val < 0.05 / p-val < 0.05 → Lists of antisenses / protein-coding pairs

TCGA Data: 1066 RNA-Seq Breast Cancer samples → Survival analysis

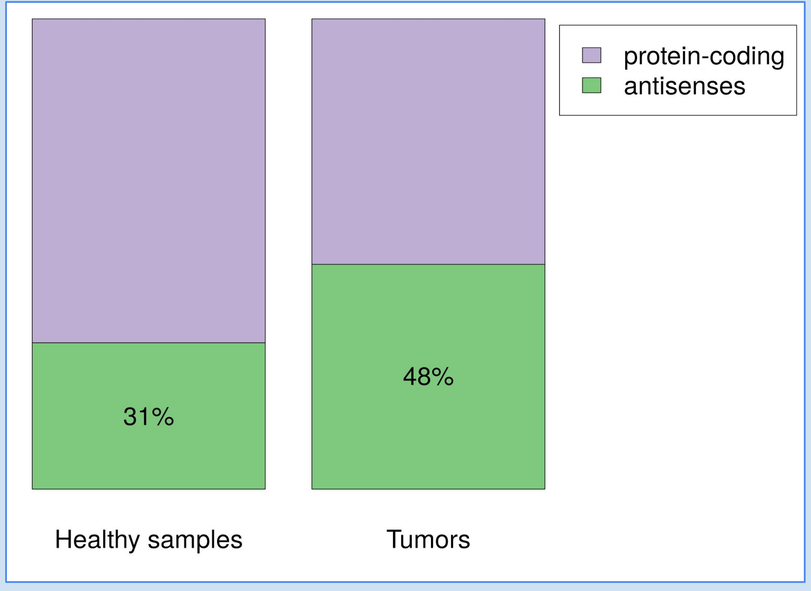
RNA-Seq pipeline | Validation of RNA-Seq data | Antisenses selection methods | Filtering of significant / extreme antisenses | Test for enrichment in survival-associated genes

Goal

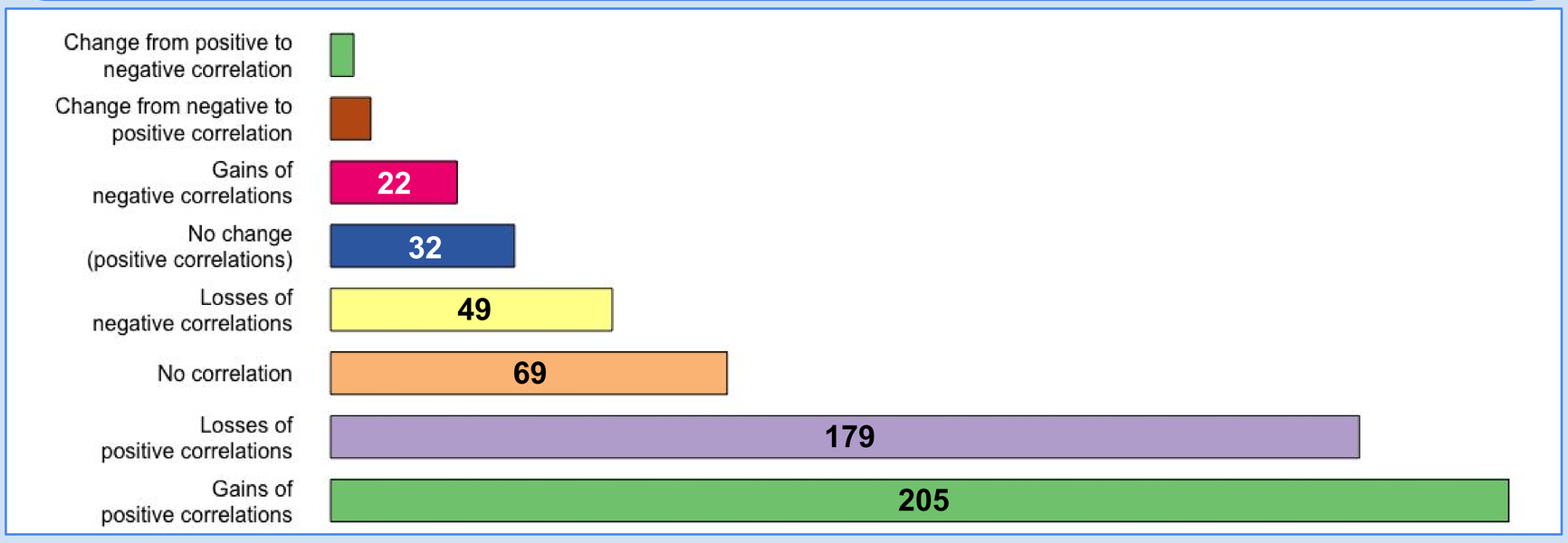
- **Global**: observe the disruption of the antisense lncRNA landscape in breast cancer tumors, at the whole transcriptome scale
- **Local**: use analytical methods to highlight pairs of antisense lncRNAs and overlapping protein-coding genes involved in the breast cancer pathology

Results

1. Global **over-expression** of antisense lncRNAs in breast cancer tumors



2. **Correlations between pairs** of antisense lncRNAs and **overlapping** protein coding genes are disrupted in breast cancer tumors, with **positive correlations** more frequently affected (**3.3** times) than negative ones.



3. Protein coding genes overlapping antisense lncRNAs highlighted by our selection methods are **1.3** times more likely than protein coding genes without antisense overlap to be **associated with survival**.

(survival analysis performed on 1066 TCGA breast cancer samples)

Conclusion

This is the first breast cancer-based, transcriptome-wide, strand-specific RNA-Seq study performed with paired tumor and adjacent tissue samples. Our results show that opposite strand transcription regulation might play a key role in the breast cancer disease, involving several different protein-coding genes and antisenses. Further functional molecular studies will be needed to explore the mechanisms and roles of specific antisenses.

Funding: French Community of Belgium, Belgian Funds for Scientific Research (F.R.S.-FNRS), F.R.S.-FNRS-Télévie, CHU Liège (F.I.R.S.), Wallonia (XSPRELTRIN), CÉCI (F.R.S.-FNRS Grant 2.5020.11)

https://www.biorxiv.org/content/early/2017/08/14/176164