

---

# Multiple kernel learning for integrative clustering in genomic precision medicine

Alessandra Cabassi\*<sup>1</sup> and Paul Kirk†<sup>1</sup>

<sup>1</sup>MRC Biostatistics Unit, University of Cambridge (MRC BSU) – United Kingdom

## Abstract

We present a method to integrate information from diverse 'omics datasets, together with clinical information, in order to define clinically actionable disease subtypes. We show how kernel methods, such as kernel k-means, can be used to summarise the output of Bayesian clustering models (such as Dirichlet process mixture models), and demonstrate how these can be extended to perform integrative clustering using multiple kernel learning. We further extend to the (semi-) supervised setting, in which additional clinical "side information" is available (e.g. survival data), and demonstrate that this can help to guide the clustering toward more relevant stratifications. We apply these methods to cancer datasets, where we combine multiple 'omics datasets and use phenotypic traits as side information in order to identify disease subtypes.

---

\*Speaker

†Corresponding author: paul.kirk@mrc-bsu.cam.ac.uk