P59

A meta-analytical approach to the molecular subtyping of prostate cancer

Eur Urol Suppl 2019;18(8):e3126

McKinney C.

Almac Group, Diagnostics, Belfast, United Kingdom

Introduction & Objectives: Multiple biomarkers and molecular subtypes have been proposed in prostate cancer (PCa) with limited overlap, in part due to the heterogeneity of the disease, data scaling and sample preservation type, all of which impact on variable selection. To reconcile inconsistencies between such studies, and develop robust consensus molecular subtypes, we must harness the full potential of the multiple datasets through appropriate analytical techniques. Such techniques should maintain the integrity and structure of individual datasets as cross-dataset normalization can potentially skew biological signals within datasets (1–3). For use in the meta-omics era, we therefore propose a novel intra-dataset and inter-dataset molecular subtype discovery technique.

Materials & Methods: Three large PCa gene expression datasets were used for the development of the intra-dataset discovery prototype, Taylor et al (n = 131, Affymetrix, FF) (4), TCGA (n = 322, Illumina HiSeq, FF) (5) and Walker et al (n = 322, FFPE, Almac Prostate DSA)(6). From previous work (7), we have developed a bottom-up unsupervised learning approach that uses random gene-sets, with gene expression levels in compositional form. Similarities between samples and cluster stability were assessed using Aitchison's distance, suitable for compositional ratios, (8), and geometric means to represent cluster centres. Clinical significance was assessed using biochemical recurrence (BCR), metastatic-free survival (MFS) and disease-free survival (DFS).

Results: In the Walker et al dataset, three molecular subtypes were characterised (MFS log-rank, p-value: 0.03, BCR log-rank: p-value: 0.005), with a trend towards a separation of Gleason 7 (4+3 vs 3+4) between the groups (chi-squared p-value: 0.089). In the Taylor et al dataset, four molecular subtypes were identified, two poor prognosis and two good prognosis groups (MFS log-rank, p-value: 0.06). In the TCGA dataset, six subtypes were found (DFS log-rank: 0.0004). Association with Gleason 7 subcategories was found (chi-square p-value: 5.428e-08).

Conclusions: From the preliminary results, we have demonstrated that our novel method can derive additional biological and prognostic insights. Our method is also able to perform quality control, prioritizing those samples that maximise intra-cluster similarity and inter-cluster separation. The approach will be released as an R package and can be used to characterise common, robust structures in multiple independent datasets, using any type of -omics data in continuous form eg methylation profiling.

1. Rung et al. Nat Rev, 2013, 14:89-99
2. Fan et al. Pharmacogenomics J, 2010, 10:247-57
3. Tseng et al. Nucleic Acids Res, 2012, 40:3785-99
4. Taylor et al. Cancer Cell, 2010, 18:11-22
5. Abeshouse et al. Cell, 2015, 163:1011-25
6. Walker et al. Eur Urol, 2017, 72:509-18
7. Blayney et al. NAR, 2016, 44:137
8. Aitchison J. Blackburn Press, 2003