

 Open access • Posted Content • DOI:10.1101/540005

MS.liverK: An R package for transcriptome-based computation of molecular subtypes and functional signatures in liver cancer — [Source link](#)

Florent Petitprez, Léa Meunier, Eric Letouzé, Yujin Hoshida ...+9 more authors

Institutions: French Institute of Health and Medical Research, Icahn School of Medicine at Mount Sinai, University of Barcelona, National Institutes of Health ...+2 more institutions

Published on: 07 Feb 2019 - [bioRxiv](#) (W.B. Saunders Ltd)

Related papers:

- [A reference profile-free deconvolution method to infer cancer cell-intrinsic subtypes and tumor-type-specific stromal profiles.](#)
- [Proteomic analysis of breast tumors confirms the mRNA intrinsic molecular subtypes using different classifiers: a large-scale analysis of fresh frozen tissue samples.](#)
- [QSurface: fast identification of surface expression markers in cancers.](#)
- [Identification of modules and functional analysis in CRC subtypes by integrated bioinformatics analysis.](#)
- [De Novo Pathway-Based Classification of Breast Cancer Subtypes](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/ms-liverk-an-r-package-for-transcriptome-based-computation-3jblxesf93>

***MS.liverK*: an R package for transcriptome-based computation of molecular subtypes and functional signatures in liver cancer**

Florent Petitprez^{1,2,3,4}, Léa Meunier⁵, Eric Letouzé⁵, Yujin Hoshida⁶, Augusto Villanueva^{6,7}, Josep Llovet^{6,8}, Snorri Thorgeirsson⁹, Xin Wei Wang⁹, Wolf H Fridman^{2,3,4}, Jessica Zucman-Rossi^{5,10} and Aurélien de Reyniès^{1,*}

¹Programme Cartes d'Identité des Tumeurs (CIT), Ligue Nationale Contre le Cancer, Paris, France

²INSERM UMR_S 1138, Cordeliers Research Centre, Paris, France

³Paris Descartes University, Sorbonne Paris Cité, Paris, France

⁴Pierre and Marie Curie University, Paris, France

⁵INSERM UMR-1162, Paris, France; ⁶Liver Cancer Program, Division of Liver Diseases, Department of Medicine, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁷Division of Haematology and Medical Oncology, Department of Medicine, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁸BCLC, Liver Unit, IDIBAPS, CIBERehd, Hospital Clínic, Universitat de Barcelona, ICREA, Barcelona, Catalonia, Spain

⁹Laboratory of Human Carcinogenesis, CCR, NCI, NIH, Bethesda, MD, USA

¹⁰AP-HP; HEGP, Department of Oncology, Paris, France.

*To whom correspondence should be addressed: reynies@ligue-cancer.net.

Abstract

Abstract

Summary: Liver cancer is a highly heterogeneous disease in terms of etiology, tissue and cellular morphology, tumor molecular characteristics, microenvironment composition and prognosis. Several studies, based on tumor gene-expression profiling (GEP) data, have dissected the molecular heterogeneity of liver cancer. They resulted in various tools, either delineating homogeneous tumor subtypes or calculating molecular scores of prognostic or biological functions. Here, we present *MS.liverK*, an easy-to-use R package providing a comprehensive implementation of these tools, for research use.

Availability and implementation: The *MS.liverK* R package is available from GitHub (<https://github.com/cit-bioinfo/MS.liverK>).

1. Introduction

Primary liver cancer is the third most deadly cancer worldwide, with Hepatocellular carcinomas (HCC) accounting for 85%-90% of the cases. The molecular heterogeneity of liver cancers has motivated numerous transcriptome-based studies, yielding gene-signatures and algorithms for molecular subtyping, prognostic prediction and functional scores computation. The use of these tools is of great interest for researchers dedicated to liver cancer study, but is very limited due to their dispersion.

Here, we introduce *MS.liverK* (Molecular Signatures in Liver Cancer), an R package reimplementing these tools. *MS.liverK* takes as input a (log₂ scale) transcriptome matrix of liver cancer samples, either microarray- or RNA-Seq-derived. It also includes graphical functions, allowing to easily visualize the outputs.

2. MS.liverK

After a careful review of the literature, we identified 5 GEP-based molecular subtyping systems of HCCs published in the last 15 years (Lee *et al.*, 2004; Boyault *et al.*, 2007; Yamashita *et al.*, 2008; Chiang *et al.*, 2008; Hoshida *et al.*, 2009). Lee classification (Lee *et al.*, 2004) was established on a belgo-chinese cohort of 91 HCC samples; it is made of 2 classes (A/B), related to proliferation and survival. Lee subtypes have been further refined in 4 groups using Alpha-fetoprotein (AFP) marker, improving association to survival. Boyault classification (Boyault *et al.*, 2007) was established on a French cohort of 57 HCCs including patients from various geographic origins; it is made of 6 classes (G1 to G6) related to TP53 and CTNNB1 mutations, proliferation, HBV infection and prognosis. Yamashita classification (Yamashita *et al.*, 2008) is based on two markers, EPCAM and AFP, and defines four groups related to differentiation and prognosis. Chiang classification (Chiang *et al.*, 2008) was built using 91 HCV infected HCCs; it identifies 5 classes related to CTNNB1 mutation, proliferation, inflammation and polysomy of the chromosome 7. Hoshida classification (Hoshida *et al.*, 2009) is based on the meta-analysis of 9 cohorts totaling 603 HCC samples; it contains 3 classes (S1/S2/S3) related to proliferation, differentiation and survival. MS.liverK implements all of the above classification systems from transcriptomic data; it also includes a conversion function to allow using different GEP platforms (Fig. S1A).

To ensure classifiers were adequately assigning classes to samples, we ran them on the data from the samples used by the four teams to define their subgroups. We then compared MS.liverK classes to the ones they originally established and found a very good correspondence (Fig. S1B): Chi-squared test p-values for Lee, Boyault, Chiang and Hoshida and Roessler classifications were all $< 2.2 \times 10^{-16}$, with accuracy rate of respectively 94.5%, 98.2%, 100%, 93.8% and 90.0%.

MS.liverK implements the prognostic prediction algorithm published by (Nault *et al.*, 2013), which is so far the most extensively validated prognostic score for hepatocellular carcinomas (HCC).

Several teams have published gene-signatures related to biological functions/pathways involved in liver cancer oncogenesis. These gene-signatures relate to TGFB1 signalling (Coulouarn *et al.*, 2008), MET pathway (Kaposi-Novak *et al.*, 2006), stemness (Oishi *et al.*, 2012; Yamashita *et al.*, 2008), EPCAM (Yamashita *et al.*, 2008) and hypoxia (van Malenstein *et al.*, 2010). MS.liverK computes all the corresponding scores.

We applied MS.liverK to the TCGA LIHC RNA-seq dataset (n=286). The graphical output shows a striking concordance across the various subtyping systems, the prognostic score and the functional scores (Figure 1). This observation is expected, given that molecular subtypes should represent homogeneous types of tumors with specific biological and prognostic characteristics.

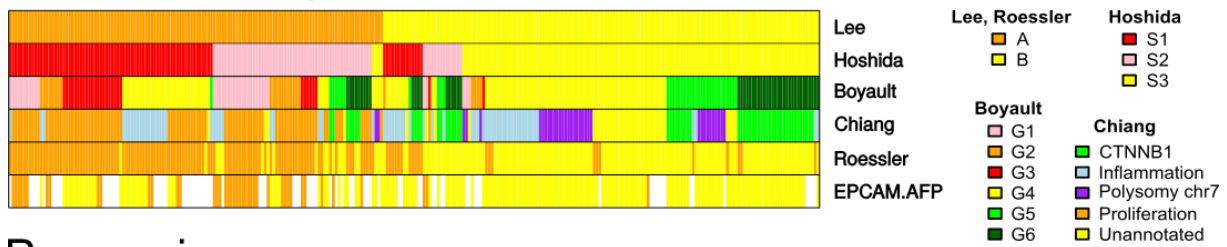
3. Software implementation and use

The package has been entirely written using R language. It includes pre-packaged data from Gene Expression Omnibus (GEO), accession number GSE20238 (Mínguez *et al.*, 2011), to be used as example data. A vignette describing the use of the package is available on GitHub (<https://github.com/FPetitprez/MS.liverK/blob/master/vignettes/vignette.pdf>). On a standard computer, the computation of all subtypes and scores takes a couple of minutes.

References

- Boyault,S. *et al.* (2007) Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology*, **45**, 42–52.
- Chiang,D.Y. *et al.* (2008) Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Res.*, **68**, 6779–6788.
- Coulouarn,C. *et al.* (2008) Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. *Hepatology*, **47**, 2059–2067.
- Hoshida,Y. *et al.* (2009) Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.*, **69**, 7385–7392.
- Kaposi-Novak,P. *et al.* (2006) Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype. *J. Clin. Invest.*, **116**, 1582–1595.
- Lee,J.-S. *et al.* (2004) Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology*, **40**, 667–676.
- van Malenstein,H. *et al.* (2010) A seven-gene set associated with chronic hypoxia of prognostic importance in hepatocellular carcinoma. *Clin. Cancer Res.*, **16**, 4278–4288.
- Mínguez,B. *et al.* (2011) Gene-expression signature of vascular invasion in hepatocellular carcinoma. *J. Hepatol.*, **55**, 1325–1331.
- Nault,J.-C. *et al.* (2013) A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology*, **145**, 176–187.
- Oishi,N. *et al.* (2012) Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma. *Hepatology*, **56**, 1792–1803.
- Yamashita,T. *et al.* (2008) EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. *Cancer Res.*, **68**, 1451–1461.

Molecular subtypes



Prognosis



Functional scores

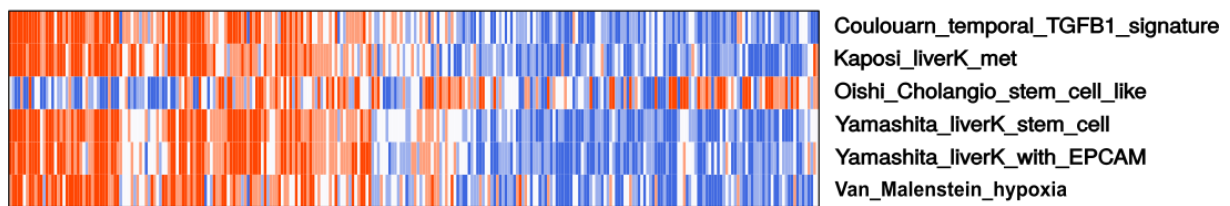


Fig. 1. MS.liverK graphical output. Graphical output of MS.liverK illustrated on the TCGA LIHC transcriptome dataset. Columns represent samples, rows represent molecular subtypes (upper panel), prognostic scores (middle panel) and functional scores (lower panel).