

# Kidney renal clear cell carcinoma expression analyse

Bofill A,<sup>\*1</sup>, Castillo S,<sup>\*1</sup> and Pérez A<sup>\*1</sup>

<sup>\*</sup>Msc in Bioinformatics for Health Sciences, Pompeu Fabra University

**ABSTRACT** The abstract should be written for people who may not read the entire paper, so it must stand on its own. The impression it makes usually determines whether the reader will go on to read the article, so the abstract must be engaging, clear, and concise. In addition, the abstract may be the only part of the article that is indexed in databases, so it must accurately reflect the content of the article. A well-written abstract is the most effective way to reach intended readers, leading to more robust search, retrieval, and usage of the article.

**KEYWORDS** Kidney; Carcinoma; Bioconductor; ASS. Sergio Castillo ...

## Your Abstract

In addition to the guidelines provided in the example abstract above, your abstract should:

- provide a synopsis of the entire article;
- begin with the broad context of the study, followed by specific background for the study;
- describe the purpose, methods and procedures, core findings and results, and conclusions of the study;
- emphasize new or important aspects of the research;
- engage the broad readership of GENETICS and be understandable to a diverse audience (avoid using jargon);
- be a single paragraph of less than 250 words;
- contain the full name of the organism studied;
- NOT contain citations or abbreviations.

## Introduction

For the introduction, authors should be mindful of the broad readership of the journal. The introduction should set the stage for the importance of the work to a generalist reader and draw the reader in to the specific study. The scope and impact of the work should be clearly stated. In individual organisms where a mutant is being studied, the rationale for the study of that mutant must be clear to a geneticist not studying that particular organism. Similarly, study of particular phenotypes should be justified broadly and not on the basis of interest for that organism alone. General background on the importance

of the genetic pathway and/or phenotype should be provided in a single, well-reasoned paragraph near the beginning of the introduction.

Authors are encouraged to:

- cite the supporting literature completely rather than select a subset of citations;
- provide important background citations, including relevant review papers (to help orient the non-specialist reader);
- to cite similar work in other organisms.

## Materials and Methods

### Sample filtering

The kidney Renal Clear-Cell Carcinoma RNA-seq data used in this study has been extracted from The Cancer Genome Atlas [TCGA](#). The data initially comprised 542 tumours and 72 normal samples. Each one of the samples have a total of 20115 genes. We have applied several filters to reduce the data samples:

- *Library size filtering*: In order to obtain a more reliable data, we need to filter out the samples with a low sequencing depth (< 45 Millions per read). We obtain a filtered data of: 298 tumor and 48 normal samples.
- *Paired data filtering*: The next step is pairing the normal and tumor data. It allows the reduction of our set and to obtain a more accurate information. This filter results in 38 normal and 38 tumor samples.

We also need to apply filters to reduce the gene sets:

- *expression levels filter*: To filter out the genes with a low expression, we use a cut-off of 1 log Counts per Millions (logCPM). With this filter we reduce the number of genes from 20.115 to 11.742 genes. Manuscripts submitted to

GENETICS should contain a clear description of the experimental design in sufficient detail so that the experimental analysis could be repeated by another scientist

### **Statistical and Data Analysis**

Analyses has been performed with R. See more information about that in: [www.r-project.org](http://www.r-project.org). All packages used in this Project could be viewed on the DONDEEEEEEEEEEE.

It is important to indicate what statistical analysis has been performed; not just the name of the software and options selected, but the method and model applied. In the case of many genes being examined simultaneously, or many phenotypes, a multiple comparison correction should be used to control the type I error rate, or a rationale for not applying a correction must be provided. The type of correction applied should be clearly stated. It should also be clear whether the p-values reported are raw, or after correction. Corrected p-values are often appropriate, but raw p-values should be available in the supporting materials so that others may perform their own corrections. In large scale data exploration studies (e.g. genome wide expression studies) a clear and complete description of the replication structure must be provided.

### **Data Availability**

At the end of the Materials and Methods section, include a statement on reagent and data availability. Please read the Data and Reagent Policy before writing the statement. Make sure to list the accession numbers or DOIs of any data you have placed in public repositories. List the file names and descriptions of any data you will upload as supplemental information. The statement should also include any applicable IRB numbers. You may include specifications for how to properly acknowledge or cite the data.

For example: Strains are available upon request. File S1 contains detailed descriptions of all supplemental files. File S2 contains SNP ID numbers and locations. File S3 contains genotypes for each individual. Sequence data are available at GenBank and the accession numbers are listed in File S3. Gene expression data are available at GEO with the accession number: GDS1234. Code used to generate the simulated data is provided in file S4.

### **Results and Discussion**

The results and discussion should not be repetitive. The results section should give a factual presentation of the data and all tables and figures should be referenced; the discussion should not summarize the results but provide an interpretation of the results, and should clearly delineate between the findings of the particular study and the possible impact of those findings in a larger context. Authors are encouraged to cite recent work relevant to their interpretations. Present and discuss results only once, not in both the Results and Discussion sections. It is sometimes acceptable to combine results and discussion. The text should be as succinct as possible. Heed Strunk and White's dictum: "Omit needless words!"

## **Additional guidelines**

### **Numbers**

In the text, write out numbers nine or less except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numbers for those larger than nine, except as the first word of a sentence; however, try to avoid starting a sentence with such a number.

### **Units**

Use abbreviations of the customary units of measurement only when they are preceded by a number: "3 min" but "several minutes". Write "percent" as one word, except when used with a number: "several percent" but "75%." To indicate temperature in centigrade, use ° (for example, 37°); include a letter after the degree symbol only when some other scale is intended (for example, 45°K).

### **Nomenclature and Italicization**

Italicize names of organisms even when the species is not indicated. Italicize the first three letters of the names of restriction enzyme cleavage sites, as in HindIII. Write the names of strains in roman except when incorporating specific genotypic designations. Italicize genotype names and symbols, including all components of alleles, but not when the name of a gene is the same as the name of an enzyme. Do not use "+" to indicate wild type. Carefully distinguish between genotype (italicized) and phenotype (not italicized) in both the writing and the symbolism.

### **In-text Citations**

Add citations using the `\citep{}` command, for example (?) or for multiple citations, (??)

## **Examples of Article Components**

The sections below show examples of different header levels, which you can use in the primary sections of the manuscript (Results, Discussion, etc.) to organize your content.

### **First level section header**

Use this level to group two or more closely related headings in a long article.

#### **Second level section header**

Second level section text.

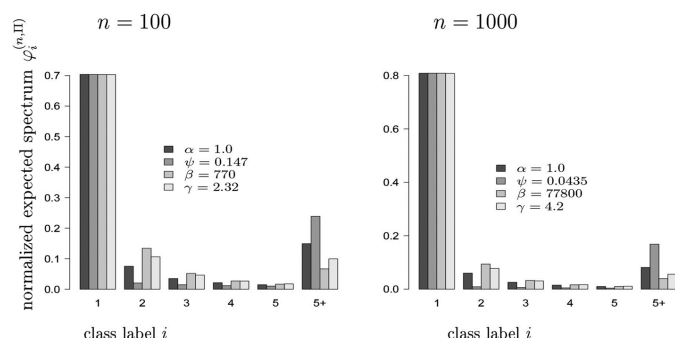
**Third level section header:** Third level section text. These headings may be numbered, but only when the numbers must be cited in the text.

### **Figures and Tables**

Figures and Tables should be labelled and referenced in the standard way using the `\label{}` and `\ref{}` commands.

#### **Sample Figure**

Figure 1 shows an example figure.

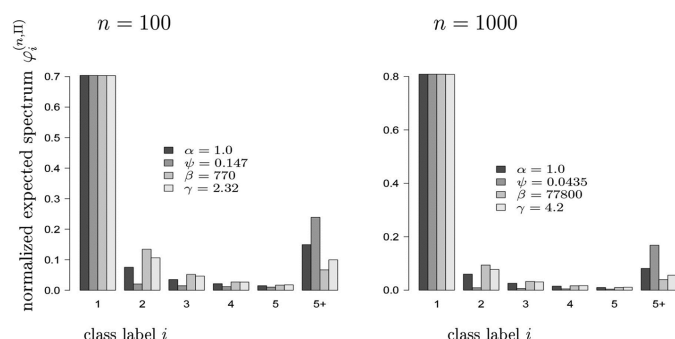


**Figure 1** Example figure from [10.1534/genetics.114.173807](https://doi.org/10.1534/genetics.114.173807). Please include your figures in the manuscript for the review process. You can upload figures to Overleaf via the Project menu. Upon acceptance, we'll ask for your figure files to be uploaded in any of the following formats: TIFF (.tiff), JPEG (.jpg), Microsoft PowerPoint (.ppt), EPS (.eps), or Adobe Illustrator (.ai). Images should be a minimum of 300 dpi in resolution and 500 dpi minimum if line art images. RGB, CMYK, and Grayscale are all acceptable. Halftones should be high contrast with sharp detail, because some loss of detail and contrast is inevitable in the production process. Figures should be 10-20 cm in width and 1-25 cm in height. Graph axes must be exactly perpendicular and all lines of equal density. Label multiple figure parts with A, B, etc. in bolded type, and use Arrows and numbers to draw attention to areas you want to highlight. Legends should start with a brief title and should be a self-contained description of the content of the figure that provides enough detail to fully understand the data presented. All conventional symbols used to indicate figure data points are available for typesetting; unconventional symbols should not be used. Italicize all mathematical variables (both in the figure legend and figure), genotypes, and additional symbols that are normally italicized.

## Table

Table 1 shows an example table. Avoid shading, color type, line drawings, graphics, or other illustrations within tables. Use tables for data only; present drawings, graphics, and illustrations as separate figures. Histograms should not be used to present data that can be captured easily in text or small tables, as they take up much more space.

Tables numbers are given in Arabic numerals. Tables should not be numbered 1A, 1B, etc., but if necessary, interior parts of the table can be labeled A, B, etc. for easy reference in the text.



**Figure 2** Example movie (the figure file above is used as a placeholder for this example). GENETICS supports video and movie files that can be linked from any portion of the article - including the abstract. Acceptable formats include .asf, avi, .wav, and all types of Windows Media files.

**Table 1** Students and their grades

Student	Grade <sup><i>a</i></sup>	Rank	Notes
Alice	82%	1	Performed very well.
Bob	65%	3	Not up to his usual standard.
Charlie	73%	2	A good attempt.

<sup>*a*</sup> This is an example of a footnote in a table. Lowercase, superscript italic letters (a, b, c, etc.) are used by default. You can also use \*, \*\*, and \*\*\* to indicate conventional levels of statistical significance, explained below the table.