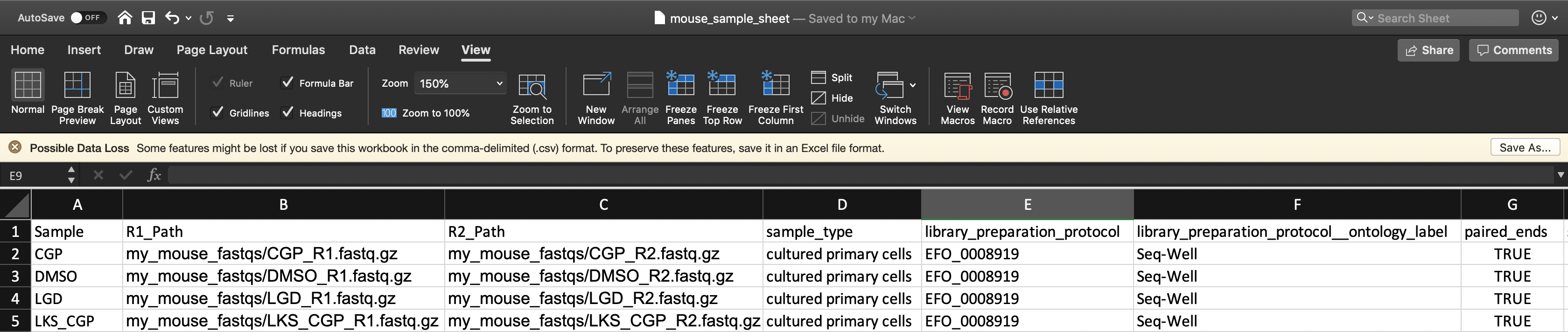
## Formatting Your input\_csv\_file

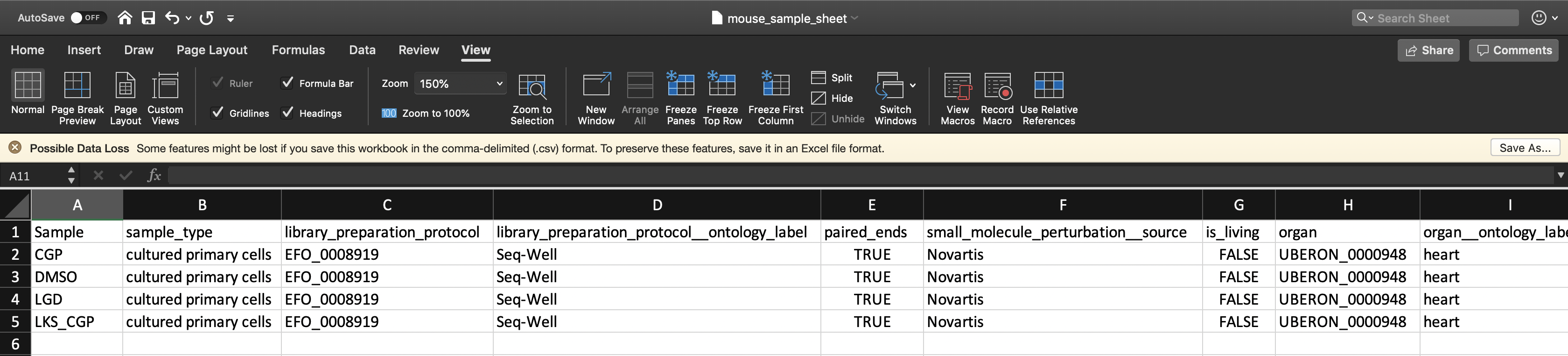
The input CSV file is a user-written comma-separated values (.csv) file with column headers and contains the following in whatever order:

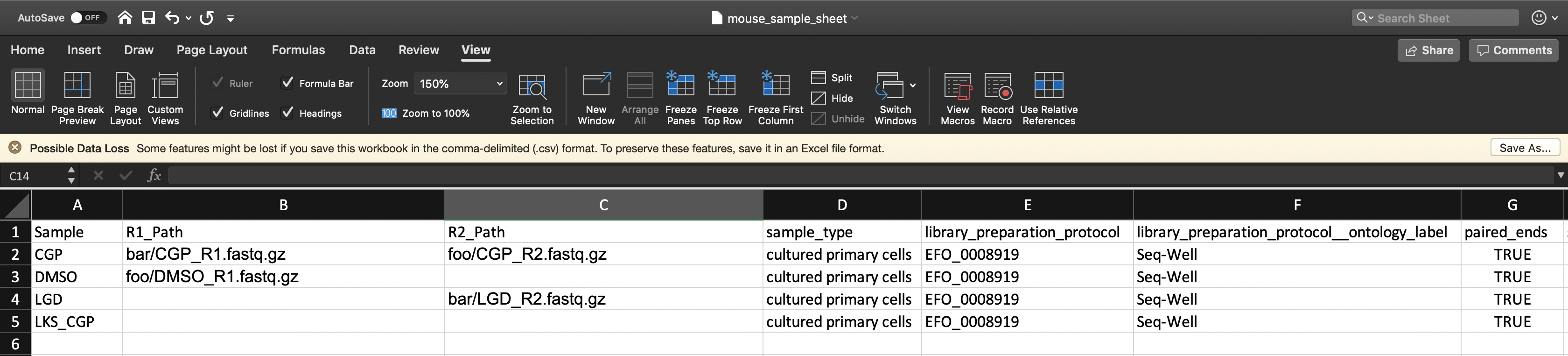
* **(REQUIRED)** the ‘Sample’ column, the sample names that prefix the respective .fastq.gz files/BCL directories and the count matrices outputted by Drop-seq pipeline.
* (OPTIONAL) both ‘R1\_Path’ and ‘R2\_Path’ columns, the paths to fastq.gz files on the bucket. If these columns are not included or if some spreadsheet cells under these columns are left blank, the workflow will check dropseq\_default\_directory by default for unspecified files.
* (OPTIONAL) ‘BCL\_Path’ column, the paths to the BCL directories on the bucket. If not included or if cells are left blank, will check dropseq\_default\_directory by default.
* (OPTIONAL) Other metadata columns that will be appended to the alexandria\_metadata.txt (tab-delimited) file generated after running scCloud. Column labels MUST match the names of the ATTRIBUTE list in the [Alexandria Metadata Convention](https://github.com/ShalekLab/alexandria/blob/master/Docker/metadata_type_map.tsv). Labels outside of this convention will be supported in the future

If made in a spreadsheet manipulation program such as Microsoft Excel, make sure to save your file as a .csv file. 

#### Understanding dropseq\_default\_directory

The use of this variable is not essential and is only meant to help users write their CSV faster. If the snapshot you are using requires dropseq\_default\_directory and you do not wish to use it, just enter an empty string: "".

Refer to the above spreadsheet example. There are four samples which each have two FASTQ reads. All FASTQ files are found in a folder located at the root of the bucket called mouse\_fastqs. Since they are all located in the same directory, one could set mouse\_fastqs as the dropseq\_default\_directory and no longer need to have R1\_Path and R2\_Path columns. 

If the user has R1\_Path and R2\_Path columns but leaves spreadsheet cells left blank, the pipeline will search in the dropseq\_default\_directory for the corresponding sample  Here the pipeline will search the gs://[bucket ID]/mouse\_fastqs directory for any spreadsheet cells left blank; DMSO\_R2.fastq.gz, LGD\_R1.fastq.gz, LKS\_CGP\_R1.fastq.gz, and LKS\_CGP\_R2.fastq.gz. The specific pattern the pipeline searches for is “[Sample Name]\*[R1 or R2]\*.fastq.gz”.

## Inputs of the dropseq\_scCloud Workflow

#### Basic Usage

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Exposed on SCP** |
| bucket | gsURL of the workspace bucket to which you have permissions, ex: gs://fc-e0000000-0000-0000-0000-000000000000/. SCP locks this value to be the workspace bucket. | No |
| input\_csv\_file | The spreadsheet (comma-separated value file) uploaded in the miscellaneous tab of your study’s Upload/Edit Study Data page. **Formatting must adhere to the criteria!** | Yes |
| reference | Enter the name of the genome to which you wish to align. Supported options include hg19, mm10, hg19\_mm10, or mmul\_8.0.1. Custom json references and hg38 will be supported in the future. | Yes |
| run\_dropseq | Select Yes to run the [Drop-seq pipeline workflow](https://sccloud.readthedocs.io/en/latest/drop_seq.html) which will align your sequencing data and perform quality control. | Yes |
| (OPTIONAL in 75 or default=False) is\_bcl | Select Yes if all of your data is in BCL format, [bcl2fastq](https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq_letterbooklet_15038058brpmi.pdf) will be run to convert your data to fastq.gz. Select No if all of your data is of fastq.gz type. | Yes |
| (OPTIONAL in 75) dropseq\_default\_directory | If all of your sequence data are located within the same folder on your bucket, enter the path to that folder starting from the bucket. Ex: Enter data/mouse\_fastqs for folder mouse\_fastqs located at the following gsURL: gs://[bucket ID]/data/mouse\_fastqs/. If not applicable, view the documentation to learn how you can list paths in the input\_CSV\_file. | Yes |
| dropseq\_output\_directory | Enter the path leading to a bucket folder where you wish all of the Drop-seq outputs (aligned data, count matrices, etc.) to be stored. All folders in this path will be created if they do not exist. Ex: Entering data/20190916/aligned will store all Drop-seq outputted files at gsURL gs://[bucketID]/data/20190916/aligned/ | Yes |
| run\_scCloud | Select Yes if you wish to create run the [scCloud workflow](https://sccloud.readthedocs.io/en/latest/scCloud.html) which will generate metadata, cluster files, coordinate files to be uploaded into the Alexandria for exploration. | Yes |
| scCloud\_output\_directory | Enter the path leading to a bucket folder where you wish all of the scCloud outputs (expression matrix, metadata, cluster, and coordinate files) to be stored. All folders in this path will be created if they do not exist. Ex: Entering data/20190916/analysis will store all Drop-seq outputted files at gsURL gs://[bucketID]/data/20190916/analysis/ | Yes |

#### Advanced Usage

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Exposed on SCP** |
| preemptible, default=2 | Number of request attempts made for a preemptible virtual machine instance before requesting a higher-cost, faster, non-preemptible instance. See [this Google Cloud documentation page](https://cloud.google.com/preemptible-vms/) for details. | Yes |
| zones, default=“us-east1-d us-west1-a us-west1-b” | The ordered list of zone preferences for requesting a Google machine to run the pipeline. See [this Google Cloud documentation page](https://cloud.google.com/compute/docs/regions-zones/) for details. | Yes |
| scCloud\_output\_prefix, default=“sco” | Enter a name you may wish to prefix to your outputted scCloud files so that you can differentiate them from files from a different job. | Yes |
| alexandria\_version, default=“0.1” | Version tag of the [shaleklab/alexandria](https://hub.docker.com/r/shaleklab/alexandria/tags) dockerfile to use. | Yes |
| dropseq\_tools\_version, default=“2.3.0” | Version of the [regevlab/dropseq](https://hub.docker.com/r/regevlab/dropseq/tags) dockerfile to use for the respective version of dropseq-tools. | Yes |
| scCloud\_version, default=“0.8.0:v1.0” | Version of the [regevlab/](https://hub.docker.com/u/regevlab)sccloud dockerfile to use for the respective version of scCloud. | Yes |

#### Drop-seq Pipeline and scCloud Optional Inputs Exposed on Terra

See the [Drop-seq Pipeline workflow documentation](https://sccloud.readthedocs.io/en/latest/drop_seq.html#inputs).  
*NOTE: dropEst does not account for strandedness and therefore its usage is not recommended.* \*NOTE:

See the [scCloud workflow documentation](https://sccloud.readthedocs.io/en/latest/scCloud.html#aggregate-matrix).

## Outputs of the dropseq\_scCloud workflow

When running dropseq and/or scCloud through dropseq\_scCloud, the workflow yields the same outputs as their counterparts.

Explicitly, dropseq\_scCloud\_workflow presents the Single-Cell Portal with the alexandria metadata file (alexandria\_metadata.txt), the dense expression matrix (ends with .scp.expr.txt), and the two coordinate files (end with .scp.X\_diffmap\_pca.coords.txt and .scp.X\_fitsne.coords.txt).