**Cancer cell line data in public domain**

**Broad-Novartis CCLE**

http: [www.broadinstitute.org/ccle](http://www.broadinstitute.org/ccle)

Data download:

<http://www.broadinstitute.org/ccle/data/browseData?conversationPropagation=begin>

Download everything except the raw CEL files (they have been converted) unless you want to use the raw files directly or keep them.

The last two files are drug data.

File format descriptions are on the same page with respective links.

Note: need to sign in (free account). If particular types of cancers/cell lines are needed, extract the data from downloaded files.

**NCI60**

<http://discover.nci.nih.gov/cellminer/>

Data download:

<http://discover.nci.nih.gov/cellminer/loadDownload.do>

Download all normalized data set (unless you want to keep the raw data set).

The last dataset is the drug data.

Dataset info: <http://discover.nci.nih.gov/cellminer/datasetMetadata.do>

Cellline info: <http://discover.nci.nih.gov/cellminer/celllineMetadata.do>

Note: If particular types of cancers/cell lines are needed, use “Query Genomic Data” and “Query Drug Data” at the download page.

**Connectivity Map**

<http://www.broadinstitute.org/cmap/>

Data download:

Once login (free account), select “downloads”, and download everything.

The instance file format is self-descriptive.

CEL file format: <http://www.stat.lsa.umich.edu/~kshedden/Courses/Stat545/Notes/AffxFileFormats/cel.html>

Note: If particular types of cancers/cell lines are needed, extract the data from the downloaded files.

**The Cancer Genome Atlas (TCGA)**

<http://cancergenome.nih.gov/>

Data download:

<https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp>

I have not downloaded any data from this site yet and more dataset are being added from what I can tell. You could presumably select All in the “Bulk Download” page and download the ones with Open Access. That’s a lot files though.