Research community has been paying great effort to generate large dataset during the past decade through collaboration to enable data-driven analysis. Cancer Cell Line Encyclopedia (CCLE)[[1]](#endnote-1) collected near 1700 human cancer cell lines across 39 different primary cancer types, and conducted extensive profiling of molecular fingerprint, including mutation, gene expression, copy number variation, miRNA expression, protein array and DNA methylation. In addition, DEMETER2[[2]](#endnote-2) and Achilles[[3]](#endnote-3),[[4]](#endnote-4) projects conducted large-scale gene essentiality genetic screen across multiple cell lines using RNAi and CRISPR-Cas9 technologies. Very recently, systematic characterizations of cell line signatures have been extended from genomics, functional genomics to metabolomics. Li et al. profiled 225 metabolites in 928 cell lines from more than 20 cancer types in the CCLE cell panel[[5]](#endnote-5). Combining all these together, we now are able to have a comprehensive view of genomic landscape of human cancer cell lines (Table 1).

With respect to pharmacological response, Genomics of Drug Sensitivity in Cancer (GDSC)[[6]](#endnote-6) measured the sensitivities of near 1000 cancer cell lines to ~250 drugs aiming to identify predictive biomarkers for different drugs. Cancer Therapeutics Response Portal (CTRP)[[7]](#endnote-7),[[8]](#endnote-8) is another large public resource that deposit the response of ~1000 cancer cell lines to over 500 compound that help mine the mechanisms of action. Besides these large valuable public datasets, recently, based on our specific interest in lung cancer, our lab has devised a tiered high-throughput screening strategy on a panel 100 lung cancer cell lines to screen 222 drugs/chemicals, and conducted whole-exome and RNA- sequencing for all the 100 cell lines[[9]](#endnote-9) (Table 2).

DREAM Challenge took initiative at computationally predicting compound pair synergism through crowdsourcing, by providing the first few drug combinatorial effect datasets[[10]](#endnote-10),[[11]](#endnote-11). O'Neil et al. dataset[[12]](#endnote-12) greatly expanded the scale of drug pairs and genetic background to have around 600 unique drug pairs and 30 cell lines. NCI ALMANAC[[13]](#endnote-13) released by far the largest high throughput drug synergy measure dataset, which included 105 FDA-approved anticancer agents covering ~5000 pairs tested in NCI-60 panel (Table 3).

Post-perturbed transcriptomic profile provided an extra layer of knowledge of drug action that can potentially improve drug interaction effect prediction compared with using baseline genetic information along. LINCS project generated gene expression signature of ~100 cancer cell lines that have been exposed ~20000 small molecules[[14]](#endnote-14),[[15]](#endnote-15),[[16]](#endnote-16) (Table 4).

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **# of cell lines** | **# of features** | **Source** |
| Mutation | 1,631 | ~ 1,000,000 | CCLE |
| Gene expression | 1,201 | 16,383 | CCLE |
| Copy number variation | 1,627 | 16,383 | CCLE |
| Protein array | 899 | 214 | CCLE |
| miRNA expression | 954 | 734 | CCLE |
| DNA methylation | 843 | 81,037 | CCLE |
| RNAi | 712 | 17,309 | DEMETER2 |
| CRISPRi | 558 | 16,383 | Achilles |
| Metabolism | 928 | 225 | CCLE |

Table 1. Genomic data summary

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **# of cell lines** | **# of drugs/chemicals** | **# of response experiment** |
| CCLE | 507 | 24 | 11,670 |
| GDSC | 1,065 | 251 | 224,202 |
| CTRP | 1,107 | 545 | 386,547 |
| UTSW | 100 | 222 | 22,200 |

Table 2. Drug and chemical perturbation data summary

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **# of unique drugs** | **# of unique drug pairs** | **# of response experiments** |
| DREAM-7 | 14 | 91 | 91 |
| DREAM-9 | 69 | 169 | 2,199 |
| ALMANAC | 105 | 5,254 | 2,815,653 |
| O’Neil et al. | 38 | 583 | 368,832 |
| UTSW | 33 | 200 | 200 |

Table 3. Drug combination data summary

|  |  |  |
| --- | --- | --- |
| **Source** | **# of cell lines** | **# of perturbation** |
| DREAM-7 | 1 | 14 |
| UTSW | 1 | 33 |
| LINCS - CMap | 98 | 19811 |
| LINCS - L1000FWD | 69 | 20449 |

Table 4. Pharmacogenomic data

Reference

1. . Barretina, Jordi, et al. "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." Nature 483.7391 (2012): 603. [↑](#endnote-ref-1)
2. . McFarland, James M., et al. "Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration." Nature communications 9 (2018). [↑](#endnote-ref-2)
3. . Doench, John G., et al. "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9." Nature biotechnology 34.2 (2016): 184. [↑](#endnote-ref-3)
4. . Meyers, Robin M., et al. "Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells." Nature genetics 49.12 (2017): 1779. [↑](#endnote-ref-4)
5. . Li, Haoxin, et al. "The landscape of cancer cell line metabolism." Nature Medicine 25.5 (2019): 850. [↑](#endnote-ref-5)
6. . Yang, Wanjuan, et al. "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells." Nucleic acids research 41.D1 (2012): D955-D961. [↑](#endnote-ref-6)
7. . Basu, Amrita, et al. "An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules." Cell 154.5 (2013): 1151-1161. [↑](#endnote-ref-7)
8. . Seashore-Ludlow, Brinton, et al. "Harnessing connectivity in a large-scale small-molecule sensitivity dataset." Cancer discovery 5.11 (2015): 1210-1223. [↑](#endnote-ref-8)
9. . McMillan, Elizabeth A., et al. "Chemistry-first approach for nomination of personalized treatment in lung cancer." Cell 173.4 (2018): 864-878. [↑](#endnote-ref-9)
10. . Bansal, Mukesh, et al. "A community computational challenge to predict the activity of pairs of compounds." Nature biotechnology 32.12 (2014): 1213. [↑](#endnote-ref-10)
11. . Menden, Michael Patrick, et al. "A cancer pharmacogenomic screen powering crowd-sourced advancement of drug combination prediction." (2017). [↑](#endnote-ref-11)
12. . O'Neil, Jennifer, et al. "An unbiased oncology compound screen to identify novel combination strategies." Molecular cancer therapeutics 15.6 (2016): 1155-1162. [↑](#endnote-ref-12)
13. . Holbeck, Susan L., et al. "The National Cancer Institute ALMANAC: A comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity." Cancer research 77.13 (2017): 3564-3576. [↑](#endnote-ref-13)
14. . Lamb, Justin, et al. "The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease." Science 313.5795 (2006): 1929-1935. [↑](#endnote-ref-14)
15. . Subramanian, Aravind, et al. "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles." Cell 171.6 (2017): 1437-1452. [↑](#endnote-ref-15)
16. . Wang, Zichen, et al. "L1000FWD: fireworks visualization of drug-induced transcriptomic signatures." Bioinformatics 34.12 (2018): 2150-2152. [↑](#endnote-ref-16)