Data-driven drug discovery (D4) exploits a comprehensive set of ‘Big’ data in order to provide an efficient path to new drug development. First, we present an in silico chemical genomics approach to predict drug repositioning (DR) candidates for three types of cancer: glioblastoma, lung cancer, and breast cancer. It is based on a recent large-scale dataset (LINCS) of ~20,000 drug-induced expression profiles in multiple cancer cell lines, which provides i) a global impact of transcriptional perturbation of both known targets and unknown off-targets, and ii) rich information on drug’s mode-of-action. Second, we exploited more than a million bioassay dataset for drug virtual screening. It provides, though not complete, extensive information on bioactivity profiles for millions of compounds. We developed a novel method to predict active compounds for a target protein, named BEAR (Bioactive compound Enrichment by Assay Repositioning). The underlying idea of BEAR is to re-use bioassay data for predicting active compounds for a target other than their originally intended targets or purposes, i.e. ‘assay repositioning.’ Our method does not depend on structural information of either target or compound. It is capable of predicting active compounds of novel scaffolds and applicable to thousands of targets.