

EVALUATION OF ANALYTICAL METHODS FOR CONNECTIVITY MAP DATA

JIE CHENG¹, QING XIE², VINOD KUMAR², MARK HURLE², JOHANNES M. FREUDENBERG³, LUN YANG²,
PANKAJ AGARWAL²

¹Statistical and Platform Technologies, GlaxoSmithKline R&D, UP4335, 1250 S Collegeville Rd, Collegeville, PA 19426; ²Computational Biology, GlaxoSmithKline R&D, UW2230, 709 Swedeland Road, King of Prussia, PA 19406, USA; ³Computational Biology, GlaxoSmithKline R&D, 3.2085A, 5 Moore Drive, Durham, NC, 27709, USA

Email: Jie.Cheng@gsk.com, Qing.Xie@gsk.com, Vinod.D.Kumar@gsk.com, Mark.R.Hurle@gsk.com,
Johannes.M.Freudenberg@gsk.com, Lun.Y.Yang@gsk.com, Pankaj.Agarwal@gsk.com

Connectivity map data and associated methodologies have become a valuable tool in understanding drug mechanism of action (MOA) and discovering new indications for drugs. However, few systematic evaluations have been done to assess the accuracy of these methodologies. One of the difficulties has been the lack of benchmarking data sets. Iskar et al. (PLoS. Comput. Biol. **6**, 2010) predicted the Anatomical Therapeutic Chemical (ATC) drug classification based on drug-induced gene expression profile similarity (DIPS), and quantified the accuracy of their method by computing the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve. We adopt the same data and extend the methodology, by using a simpler eXtreme cosine (XCos) method, and find it does better in this limited setting than the Kolmogorov-Smirnov (KS) statistic. In fact, for partial AUC (a more relevant statistic for actual application to repositioning) XCos does 17% better than the DIPS method ($p=1.2e-7$). We also observe that smaller gene signatures (with 100 probes) do better than larger ones (with 500 probes), and that DMSO controls from within the same batch obviate the need for mean centering. As expected there is heterogeneity in the prediction accuracy amongst the various ATC codes. We find that good transcriptional response to drug treatment appears necessary but not sufficient to achieve high AUCs. Certain ATC codes, such as those corresponding to corticosteroids, had much higher AUCs possibly due to strong transcriptional responses and consistency in MOA.

1. Introduction

Identifying the correct disease indication for a drug is an important problem and several computational methods have been described [1]. The problem for any practitioner, however, is to assess the precision of these methods. The desired method should provide relatively high confidence that the first few indications that are predicted for a drug contain at least one that will be validated in clinical trials and make a positive impact on patients. One of the most important techniques in the space of drug repositioning is connectivity map (CMAP) [2].

A key contribution of CMAP has been the establishment of a database of cellular expression profiles in response to drug treatment in cell lines such as MCF7. This has enabled both the discovery of drug MOA and new indications [2,3]. Several CMAP hypotheses suggesting

potential therapeutic compounds for new disease indications have been experimentally validated [4-8].

However, despite numerous impressive anecdotal validations, it has proven challenging to quantitatively estimate the accuracy of this technique. A gold standard data set still eludes us in terms of drugs that impact a disease positively. Thus research has turned to the cleaner benchmarking data sets to predict drug relationships. This is with the implicit hope that methods that better predict drug classes will also do better at predicting disease indications for drugs. A useful classification is the Anatomical Therapeutic Chemical (ATC) system, which codes divides drugs into different groups in a hierarchical fashion according to the organ or system on which they act and their therapeutic and chemical characteristics. The ATC level 4 is mostly based on common MOA, and thus has proven useful as a benchmark for comparing similar drugs.

The initial CMAP approach utilized a nonparametric, rank-based Kolmogorov-Smirnov (KS) statistic to connect disease gene expression signatures to drug expression profiles. KS scores are generated based on the location of the genes in the signature (i.e. up and down lists) within the entire ordered list of gene expression changes in response to compound treatment. The disease signatures often come from public repositories of expression profiles, such as Gene Expression Omnibus (GEO) [9]. Compounds from the reference dataset can also be connected with each other using the same type of computation to evaluate the similarity between them.

Iskar et al [10] provided one of the first quantitative evaluations of CMAP methods. They applied a centered mean normalization approach to preprocess the intensity data in order to eliminate batch-specific effects. The pair-wise **d**rug-**i**nduced gene expression **p**rofile **s**imilarity (DIPS) scores between each pair of drugs in CMAP were then calculated using a method similar to inverse total enrichment score (TES) by Iorio et al [11]. (TES itself is modification of KS.) They used compounds with high chemical similarities, and compounds with shared ATC classification as true positives for their benchmarking. They computed the area under the receiver operator characteristic (ROC) curve (AUC_{0.1}) to measure differences at a low false positive rate (FPR=0.1). This emphasizes early retrieval which is important because for repositioning we are willing to sacrifice some true positives to keep false positives low. The performance of DIPS was shown to be superior to the compound vs. biological control comparison method described by Iorio et al.

In addition to modifications of CMAP data processing workflows, many groups have investigated alternatives to the KS statistic. More recently researchers have extended methods based on Spearman's correlation (EPSA) [12], Fisher's Exact test (EXALT) [13], Wilcoxon rank-sum test (openSESAME) [14], weighted Pearson correlation[15], logistic regression (LRpath) [16], probabilistic categorization (ProbCD) [17], empirical background p-values[18], random set statistic (GRS) [19]. and partially ranked data[20]. In this paper, we explore an eXtreme Cosine method that truncates the middle of the two expression profiles being compared. This focuses

attention on true outliers in both treatments. The cosine is an inner product of two vectors much like Pearson correlation, which has been shown to be superior to GSEA [18].

In this study, we use the ATC classification as the benchmark to compare the eXtreme cosine method (XCos) to other CMAP scoring methods, data processing methods, and signature sizes. Insights from these comparisons will clarify parameter choices, which can then be used in drug repositioning where gold standard benchmarking datasets are more complicated. We score each method using AUC in the early (FPR=0.1 and FPR=0.01) discovery phase. This allows us to determine which compound classes contain robust expression profiles in CMAP data, and which analytical approaches are more accurate at least in this evaluation.

2. Methods

2.1. Data sources and data processing

Small-molecule perturbed genome-wide transcriptional response data were downloaded from the Connectivity Map (CMAP, build 02, <http://www.broadinstitute.org/CMAP/>). These data comprises of 6,100 gene expression instances (treatment vs. vehicle control pairs) from primarily three human cultured cell lines (MCF7, PC3, and HL60) treated with 1,309 bioactive small chemical molecules at varying concentrations. Each instance denotes a treatment and control pair for one small molecule. Each instance has attributes such as perturbagen name, concentration, cell line and batch etc.

Two methods of pre-processing probe level intensities are considered in this paper:

- a) **MC: Mean Centering** CMAP data was obtained directly from P. Bork [personal communication]. The data was generated using the method described by Iskar et al.[10]. Briefly, each compound treatment arrays were grouped based on the cell line and normalized separately using RMA [21]. Vehicle controls from CMAP were discarded and for each batch individual probes for each treatment were mean centered to calculate the average difference values within the batch. The final data consists of 4,849 treatment instances from three cell lines corresponding to 1,144 small molecules.
- b) **Batch DMSO Control (BDC)**: Using controls from within the batch was proposed in the original CMAP paper [2], and we wanted to directly compare MC to it. Probe level data (CEL files) from CMAP was processed using Array Studio (Omicsoft Corporation, Research Triangle Park, NC, USA). Briefly, microarray datasets were grouped based on the cell line. For each microarray dataset, the probe set intensities were normalized using RMA. Next, all scaled probe sets with values less than primary threshold values (set to 64) for all treatments and control samples was set to that threshold value. The intensity values for each probe set are then \log_2 transformed. Finally, the \log_2 intensities of each probe set from all vehicle control samples within the same batch and cell line are averaged and subtracted from the treatment sample to generate the corresponding treatment-to-control

values and this is termed BDC. We filtered the 6,100 instances to the same 4,849 for MC to make results comparable.

We averaged multiple instances for each compound within a cell line and then across cell lines.

The ATC codes were obtained from Iorio et al. [11] and then supplemented by additional annotation.

2.2. *Pair-wise similarity scores*

We used four methods: KS, TES, DIPS, and XCos to compute similarities between drug pairs.

KS: The initial CMap approach utilized a nonparametric, rank-based Kolmogorov-Smirnov (KS) statistic [2].

TES (inverse total enrichment score) is a measure based on the KS statistic as described in Iorio et al.[11]. A key difference is that this does not require the up and down signature to have consistent direction of scores compared to KS.

DIPS: Uses TES on mean centered (MC) data and we used the data as provided (personal communication, P. Bork).

XCos: The Xtreme cosine similarity score is calculated by retaining only the Xtreme probes for each instance after sorting by decreasing fold-change, i.e., only keeping the top N and bottom N probe sets and setting all other probe sets to zero. The cosine similarity between two Xtreme instances can then be calculated as a dot product of the two vectors. This is a variation of a described method [22]. Cosine similarity is much like Pearson correlation except that the vectors are not centered around their individual means. Unlike Euclidian distance, both cosine similarity and Pearson correlation are scale independent and should be more robust for our purpose.

Pair-wise similarity scores of compounds for each of the three cell lines are generated separately and then combined. Similarities between instances of the same compound are excluded and not included in any of the plots.

2.3. *Method nomenclature*

Eight of the nine methods described in this paper follow the SIM_PROC_SIZE nomenclature. SIM describes the similarity method which is one of KS, TES, or XCos (see section 2.2); PROC describes the data processing method which is either MC or BDC (see section 2.1) and the SIZE is the size of the signature which is either 100 or 500. The KS and TES methods were only evaluated with MC (and not with BDC), thus we have 8 total methods. DIPS is the ninth method as described in Iskar et al. [10]. DIPS is most closely related to TES_MC_500 though DIPS uses a

sort order based on detection calls, while our implementation of MC uses a sort order based on fold changes. Moreover, DIPS used only a single ATC for each drug while we used all ATC codes for a drug.

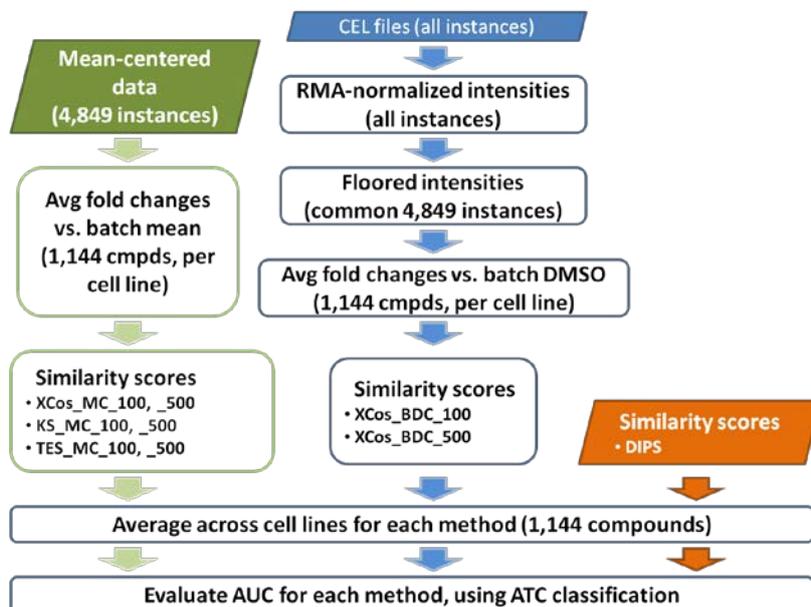


Figure 1. A schematic of the analytical workflow used to generate the AUC. Parallelograms indicate data acquired. The nine measures of similarity scores listed in the three similarity score rectangles were evaluated on the ATC codes.

2.4. AUCs and p-values

Pair-wise similarity scores are evaluated using individual ATC codes at different levels as well as using ATC levels from 1 to 4 for each of the nine methods as listed in Figure 1.

For calculating AUC of a particular ATC level, the positive cases are distinct compound pairs that share any ATC code at this level. All other pairs are considered negative cases. These criteria are effective in handling drugs that have multiple ATC codes. The ROCs in Figure 2 and Table 1 use this method as they count matches across ATC level 4 as positives.

For calculating AUC for a specific ATC code, the only relevant pairs are those have at least one compound with this ATC code. The positive cases are defined as distinct compound pairs that both share this ATC code. The negative cases are the compound pairs with only one compound belonging to this ATC code. Thus, if neither compound of a pair share this ATC code, the pair is excluded from the AUC calculation for this ATC code. Figure 3 uses this as the standard as AUCs are shown for individual ATC codes.

The p value calculation for comparing “paired” partial AUC is based on a bootstrap test [23]. Z is defined as $(pAUC_1 - pAUC_2) / sd(pAUC_1 - pAUC_2)$, where $pAUC_1$ and $pAUC_2$ are the two paired

partial AUCs to be compared and the $sd(pAUC_1-pAUC_2)$ is the standard deviation of the difference between $pAUC_1$ and $pAUC_2$. The standard deviation of the difference between the two AUCs is estimated from the 1,000 bootstrap runs.

2.5. Expression signal strength

The expression signal strength (ESS) is defined as the sum of the absolute values of the \log_2 of the fold changes of the top and bottom N features (or probesets) of a gene expression profile. We first calculated the ESS of every compound expression profile. The ESS values of the same compound were then averaged within a cell line, and then these were averaged across the three cell lines to generate one ESS value per compound. The ESS for a particular ATC code is calculated by averaging all ESS values of the compounds that belongs to this ATC code. Figure 3 plots ESS on the x-axis with N=50.

3. Results

Assessment of methods on 4th level ATC codes

An earlier study showed that DIPS method leads to fewer false positives when compared using a partial AUC value at FPR=0.1 (AUC0.1) counting every pair of drugs which had at least one matching ATC 4th level code as a true positive [10]. Also the AUC0.1 was higher with mean centering (MC) than without mean centering. In this study, we systematically evaluated multiple scoring methods using the same data processing method and AUC measurement. We also suggest and evaluate the performance of the XCos similarity for the expression vectors of pairs of drugs using the top and bottom differentially expressed probes.

XCos_BDC_100 performed best in terms of AUC at FDR=0.1 (see Figure 2 and Table 1). The AUC was **0.0193** and significantly better than the DIPS AUC of 0.016 (two tailed $p = 1.8e-7$). The difference between XCos and DIPS is even larger and more significant at FPR=0.01 ($p < 1e-13$). This may suggest that for early discovery consistent with drug repositioning the XCos with smaller signatures might indeed be better. There are three obvious differences between these two (XCos_BDC_100 and DIPS) methods: *A*) the batch DMSO control (BDC) vs. mean centering (MC), *B*) the size of the signature: 100 vs. 500, and *C*) the method itself: XCos vs. TES. To understand this further, we isolated these three differences.

- A. XCos_BDC_100 had higher AUC0.1 compared with XCos_MC_100 ($p=5e-4$), thus at least for the XCos method, the batch-based DMSO controls are better than mean centering.
- B. The AUC for XCos_BDC_100 is higher than for XCos_BDC_500, but not significant statistically ($p=0.26$). However, the AUC difference for KS_MC_100 compared to KS_MC_500 is significant ($p=6e-6$), thus at least for KS_MC the smaller signatures are better in this comparison.

C. In terms of method itself, XCos outperformed KS ($p=0.008$) with mean centering and 100 probe signatures.

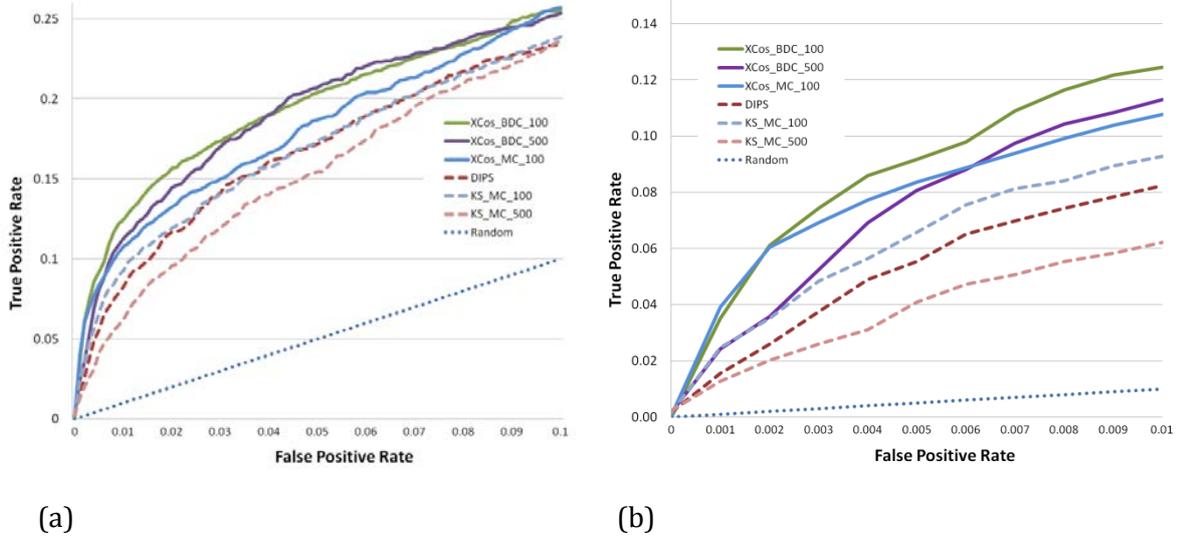


Figure 2: Comparison of the different scoring, data processing methods and signature sizes. Drugs with at least one matching ATC 4th level code are counted as true positives. The two TES scores track KS quite closely so are not shown for clarity. **a)** AUC0.1: Partial ROC curve at the FPR = 0.1. **b)** AUC0.01: Partial ROC curve at the FPR = 0.01.

Table 1: Partial AUCs from multiple scoring methods. Drugs with at least one matching ATC 4th level code are counted as true positives.

Method	AUC0.1: Partial AUC @FPR= <u>0.1</u>	AUC0.01: Partial AUC @FPR= <u>0.01</u>
KS_MC_100	0.01655	6.06e-4
KS_MC_500	0.01503	3.79e-4
TES_MC_100	0.01663	6.12e-4
TES_MC_500	0.01484	3.82e-4
XCos_MC_100	0.01789	7.73e-4
XCos_MC_500	0.01738	6.84e-4
XCos_BDC_100	0.01926	8.56e-4
XCos_BDC_500	0.01898	7.20e-4
DIPS	0.01642	5.14e-4

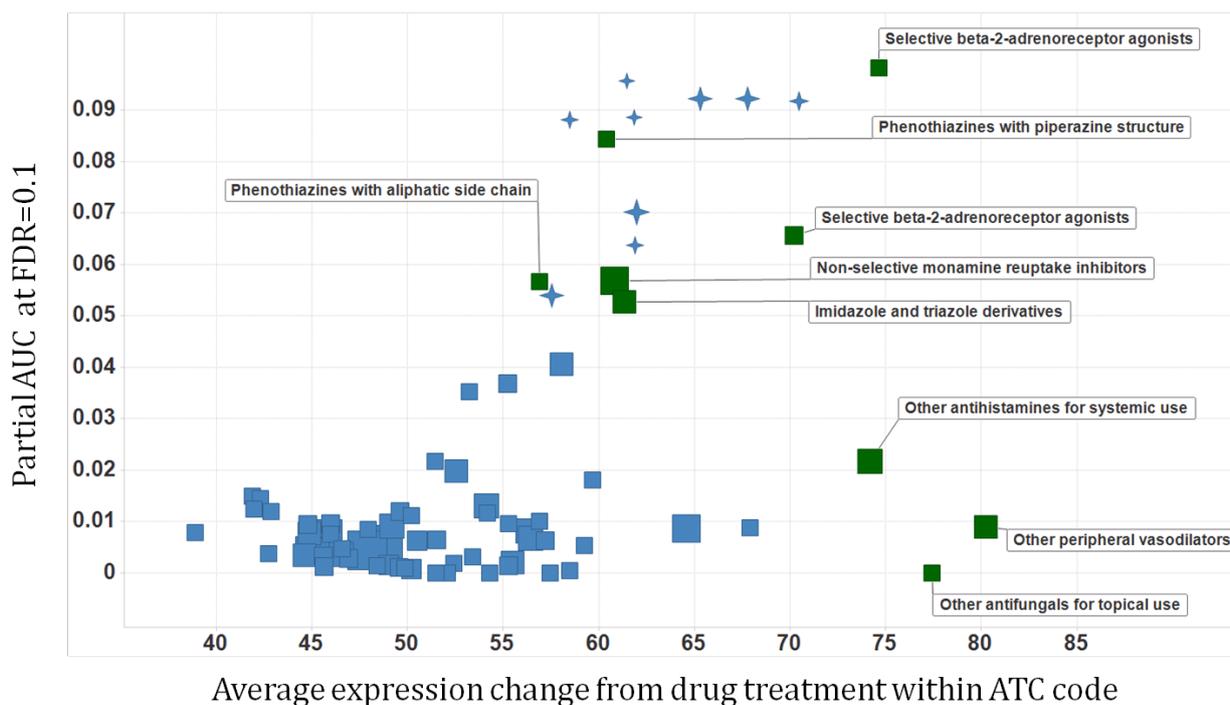


Figure 3. Relationship between AUC0.1 (for XCos_BDC_100) and the average expression change from drug treatment within an ATC level code. ATC codes which primarily describe corticosteroids are indicated by crosses, all other ATC codes are shown as rectangles. Descriptions are provided for ATC codes of interest shown in green rectangles. Points are sized by the number of compounds in the ATC code. All ATC level 4 codes with at least 5 compounds are shown. If all 100 probesets had a uniform absolute fold change of 1.414, it would correspond to an expression level of 50 on the x-axis.

All the p-values were computed as described in the methods. In fact, from Figure 2a and Table 1 the trends mentioned above are quite apparent as well and the AUC0.1 for XCos_BDC_100 is statistically significantly different from the AUC0.1 for all the MC methods in Figure 2a. The above trend in terms of AUC0.1 comparisons on different methods could not be observed on the overall AUCs (data not shown). For overall AUCs, we observed that mean centering outperforms batch-based DMSO controls at least for the XCos method. We also noticed that TES is quite similar to KS and thus not shown in Figure 2 for readability.

Differences amongst ATC codes

The specific ATC codes at level 4 compared to the generic ATC level 1 codes provide more accurate classifiers; in fact, the classification at ATC level 1 is quite close to random (data not shown). Figure 3 displays the heterogeneity in the AUC measures for ATC level 4 codes using XCos_BDC_100. The ATC codes with the strongest signal are dominated by corticosteroids, β 2-adrenoreceptor agonists, and phenothiazines. We note a large number of related corticosteroid-related ATC codes with high AUC0.1. On investigation, these are compounds with same MOA

but grouped into different ATC codes based on strength, anatomy, and formulation (inhaled, oral or topical).

This figure also shows the dependence of the AUC_{0.1} on the average change in expression due to compound treatment for a given ATC class. It seems intuitively obvious that if the expression change is low, the analytical methods cannot detect similarity. In addition, we observed that the poorly detected ATC codes with high expression changes (those labeled as starting with “Other”) are often collections of miscellaneous compounds that are unlikely to have common MOAs.

4. Discussion

Numerous methods have been proposed to identify related transcriptional profiles for CMAP readouts. They differ mostly by the underlying similarity measure, some of which are quite simple and have been known for decades, while other, more complex methods rely on powerful computing. Surprisingly, the XCos similarity score, which simply measures the cosine of two signatures, outperforms the standard, Kolmogorov-Smirnov (KS)-based CMAP method (Figure 2). Furthermore, the similarity between related signatures appears to be driven by the genes that change the most between treatment and control. Both XCos and KS scoring methods based on the top 100 features more accurately predicted ATC codes than the ones based on the top 500 features. Of course, both these signature sizes are arbitrary and the optimal signature size should be further explored. Flexible signature sizes, however, have also been explored recently [24]. Finally, the preprocessing method used to compute the signatures plays a significant role as well. We find that mean centering does not improve the similarity scores in comparison to batch based DMSO controls – at least for the XCos method. This contrasts with the earlier results[10], and the reason is not evident; however, possible explanations include our not using probeset detection calls and DIPS comparison not using batch-matched DMSO controls. Moreover, we did not restrict a drug to have exactly one ATC code as required by DIPS [10].

It should be noted that these conclusions should be considered preliminary as they are limited by the use of ATC codes as a “gold standard”. Multiple ATC codes per compound can lead to errors and redundant ATC codes may inflate AUCs. Furthermore, many ATC codes do not properly characterize MOAs (e.g. “other peripheral vasodilators”, Figure 3).

Another limitation may be that the averaging over multiple cell lines averages biological variation for compounds that may have differential responses in the three cell lines. On the other hand, using all available data may lead to more “stable” compound-specific signatures.

Future work should explore additional accuracy measures, as even AUC_{0.1} and AUC_{0.01} have too many false positives to be useful in terms of number of hypotheses that can be experimentally validated. It should also compare more methods and isolate the impact of each parameter completely across multiple methods. As indicated in Figure 3, some ATC codes lead to high AUC numbers regardless of the method used i.e. some drug classes are really easy to find

with expression profiles. To ensure that such high performing ATC codes do not skew the overall comparison, future work should include a comparison of methodologies focusing only on the more “difficult” ATC codes.

A key challenge for drug repositioning is to develop a gold standard benchmarking data set that will not necessitate the extrapolation of results from drug MOA. With some expert curatorial effort FDA approved indications could be mapped to a disease ontology. However, it is not evident as to what constitutes matching disease signatures as we would also need to determine which of those drugs are disease modifying as opposed to those providing symptomatic relief and not expected to match as true positives. We believe that quantitative assessment of repositioning methodologies is a must, if computational biology is to make a more compelling case for its utility in this field.

5. Acknowledgments

We would like to thank Drs Aaron Spivak, Hui Huang, Kwan Lee, and Philippe Sanseau for feedback and comments.

References

1. J. T. Dudley, T. Deshpande, and A. J. Butte, "Exploiting drug-disease relationships for computational drug repositioning," *Brief. Bioinform.* **12**, 303-311 (2011).
2. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub, "The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease," *Science* **313**, 1929-1935 (2006).
3. X.A.Qu and Rajpal D.K. Applications of Connectivity Map in Drug Discovery and Development. *Drug Discov Today* . 2012.
4. M. Chang, S. Smith, A. Thorpe, M. J. Barratt, and F. Karim, "Evaluation of phenoxybenzamine in the CFA model of pain following gene expression studies and connectivity mapping," *Mol. Pain* **6**, 56 (2010).
5. S. Claerhout, J. Y. Lim, W. Choi, Y. Y. Park, K. Kim, S. B. Kim, J. S. Lee, G. B. Mills, and J. Y. Cho, "Gene expression signature analysis identifies vorinostat as a candidate therapy for gastric cancer," *PLoS. One.* **6**, e24662 (2011).

6. J. T. Dudley, M. Sirota, M. Shenoy, R. K. Pai, S. Roedder, A. P. Chiang, A. A. Morgan, M. M. Sarwal, P. J. Pasricha, and A. J. Butte, "Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease," *Sci. Transl. Med.* **3**, 96ra76 (2011).
7. Y. Ishimatsu-Tsuji, T. Soma, and J. Kishimoto, "Identification of novel hair-growth inducers by means of connectivity mapping," *FASEB J.* **24**, 1489-1496 (2010).
8. S. D. Kunkel, M. Suneja, S. M. Ebert, K. S. Bongers, D. K. Fox, S. E. Malmberg, F. Alipour, R. K. Shields, and C. M. Adams, "mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass," *Cell Metab* **13**, 627-638 (2011).
9. T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva, "NCBI GEO: archive for functional genomics data sets--10 years on," *Nucleic Acids Res.* **39**, D1005-D1010 (2011).
10. M. Iskar, M. Campillos, M. Kuhn, L. J. Jensen, V. van Noort, and P. Bork, "Drug-induced regulation of target expression," *PLoS. Comput. Biol.* **6**, (2010).
11. F. Iorio, R. Bosotti, E. Scacheri, V. Belcastro, P. Mithbaekar, R. Ferriero, L. Murino, R. Tagliaferri, N. Brunetti-Pierri, A. Isacchi, and D. di Bernardo, "Discovery of drug mode of action and drug repositioning from transcriptional responses," *Proc. Natl. Acad. Sci. U. S. A* **107**, 14621-14626 (2010).
12. J. D. Tenenbaum, M. G. Walker, P. J. Utz, and A. J. Butte, "Expression-based Pathway Signature Analysis (EPSA): mining publicly available microarray data for insight into human disease," *BMC. Med. Genomics* **1**, 51 (2008).
13. Y. Yi, C. Li, C. Miller, and A. L. George, Jr., "Strategy for encoding and comparison of gene expression signatures," *Genome Biol.* **8**, R133 (2007).
14. A. C. Gower, A. Spira, and M. E. Lenburg, "Discovering biological connections between experimental conditions based on common patterns of differential gene expression," *BMC. Bioinformatics.* **12**, 381 (2011).
15. J. M. Engreitz, R. Chen, A. A. Morgan, J. T. Dudley, R. Mallelwar, and A. J. Butte, "ProfileChaser: searching microarray repositories based on genome-wide patterns of differential expression," *Bioinformatics.* **27**, 3317-3318 (2011).
16. M. A. Sartor, G. D. Leikauf, and M. Medvedovic, "LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data," *Bioinformatics.* **25**, 211-217 (2009).
17. R. Z. Vencio and I. Shmulevich, "ProbCD: enrichment analysis accounting for categorization uncertainty," *BMC. Bioinformatics.* **8**, 383 (2007).

18. S. W. Tanner and P. Agarwal, "Gene Vector Analysis (Geneva): a unified method to detect differentially-regulated gene sets and similar microarray experiments," *BMC. Bioinformatics.* **9**, 348 (2008).
19. J. M. Freudenberg, S. Sivaganesan, M. Phatak, K. Shinde, and M. Medvedovic, "Generalized random set framework for functional enrichment analysis using primary genomics datasets," *Bioinformatics.* **27**, 70-77 (2011).
20. M. R. Segal, H. Xiong, H. Bengtsson, R. Bourgon, and R. Gentleman, "Querying genomic databases: refining the connectivity map," *Stat. Appl. Genet. Mol. Biol.* **11**, (2012).
21. R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics.* **4**, 249-264 (2003).
22. A. Ben-Hur and I. Guyon, "Detecting stable clusters using principal component analysis," *Methods Mol. Biol.* **224**, 159-182 (2003).
23. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, and M. Muller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC. Bioinformatics.* **12**, 77 (2011).
24. D. Shigemizu, Z. Hu, J. H. Hung, C. L. Huang, Y. Wang, and C. DeLisi, "Using functional signatures to identify repositioned drugs for breast, myelogenous leukemia and prostate cancer," *PLoS. Comput. Biol.* **8**, e1002347 (2012).