

## RESEARCH HIGHLIGHT

# From small studies to precision medicine: prioritizing candidate biomarkers

Daniel P Gaile\* and Jeffrey C Miecznikowski

See related research, <http://genomemedicine.com/content/5/11/103>

### Abstract

There are still many open questions in data-analytic research pertaining to biomarker development in the era of personalized/precision medicine and big data. Among them is the question of what constitutes best practice for the extraction of prioritized lists of candidate biomarkers from smaller studies that are 'hypothesis generating' in nature. A recent comparison of methods to detect patient-specific aberrant expression events in small- to medium-sized (10 to 50 samples) studies provides results that favor the use of outlying degree methods.

### Developing biomarkers for precision medicine

The concept behind personalized/precision medicine is intuitive: patients are better modeled by a subgroup of patients that are most like them, rather than a larger, more general population of patients [1]. Conceivably relevant biomarkers can be used to define subgroups of patients, and a patient's subgroup affiliation can be incorporated into medical decisions. Biomarkers such as prostate-specific antigen and specific mutations in genes encoding BRCA1/BRCA2 (which increase breast and ovarian cancer risk), have been utilized in clinical practice for some time, so the concept of personalized/precision medicine is not a new one. However, the use of genetic testing and molecular diagnostics will almost certainly grow in the coming years [2]. Furthermore, expectations regarding the level of precision for such tools will likely be increased by the perception that big data (for example, clinical databases, high-throughput experimental datasets, bioinformatic resources) can be translated into clinically relevant and useful information.

If finding a biomarker can be compared to finding a 'needle in a haystack', then big data promises more needles, but with the challenge of substantively more haystacks. For example, the high-throughput 'omics component of big data has demonstrated, across multiple studies, the potential for 'extreme genomic heterogeneity' [3]. Such heterogeneity constitutes a potentially rich source of candidate biomarkers. Screening for biomarkers as covariates within classic statistical models requires that error rates be controlled in a manner that accounts for test multiplicity. Similarly, classic predictive model building approaches must minimize the potential for over-fitting models. Personalized-medicine-based clustering approaches, wherein patients are clustered into subgroups using statistical clustering heuristics, cannot escape this multiplicity issue either. Rather, subgroup approaches, such as 'patients like me' [4], must instead be considered as 'patients like me with respect to traits that are relevant to my particular medical decision' (for instance, treatment choice). Clustering heuristics require the quantification of the 'distance' between patients, where the distance is preferably measured with respect to a clinically relevant set of covariates. However, with data available for so many patient covariates (for example, clinical, lifestyle, genetic or genomic factors), understanding which covariates are truly relevant is a major challenge. Inclusion of clinically irrelevant covariates will increase the variability of distance estimates and reduce the efficiency of clustering heuristics. Both classic statistical modeling and contemporary cluster-based approaches to precision medicine will be optimized if their inputs consist of refined, rather than expansive and diluted, lists of candidate biomarkers.

The refinement of a set of candidate biomarkers can be achieved through many different pipelines. Although all pipelines aim to end with the types of controlled studies that merit approval from, for example, the US Food and Drug Administration [5], they may originate in different ways, with data-mining 'hypothesis-generating'

\* Correspondence: [dpgaile@buffalo.edu](mailto:dpgaile@buffalo.edu)  
Department of Biostatistics, SUNY University of Buffalo, 433 Kimball Tower,  
3435 Main Street, Buffalo, NY 14214, USA

approaches having become more prevalent in recent years. There are numerous open and important data analytic research questions associated with all stages of biomarker development [6]. These include questions related to the analysis of hypothesis-generating pilot studies. Clearly, the identification of better candidate biomarkers at the beginning of the development pipeline will prove beneficial in the later stages of the process. Here, we discuss a recent comparison study [7] of different methods that generate prioritized lists of candidate biomarkers using data from small, hypothesis-generating studies.

### Identification of aberrant features

The study by Bottomly and colleagues [7] provides recommendations regarding the use of existing tools for the identification of patient/sample aberrant feature values, such as gene expression values in small- to medium-sized (10 to 50 samples) single-arm (all patients part of the same disease subgroup with no control group) studies. The study compared the  $Z$  score,  $R$  score, and both weighted and unweighted variants of the outlying degree (OD) metric. Each of these metrics measures the extent to which an observation can be considered an aberration (that is, an outlier) with respect to the distribution of the balance of the observations. The  $Z$  score measures an observation's deviation from the mean in standard deviations, the  $R$  score measures an observation's deviation from the median in median absolute deviations, and the (weighted) outlying degree is the (weighted) sum of the  $k$  smallest absolute deviations involving a particular observation (where  $k$  is the OD tuning parameter). Bottomly *et al.* provided simulation-based evidence suggesting the superiority of the OD heuristic over the other heuristics.

The most important aspect of the Bottomly *et al.* article is the study of outlying degree-based heuristics. Their simulation studies suggest the superiority of the OD heuristic over the  $Z$ -score and  $R$ -score heuristics and support the practice of setting the OD tuning parameter,  $k$ , equal to half the sample size. Additionally, the authors proposed two specific weighting schemes and presented evidence that suggests that the weighted variants of the OD heuristic can provide an improved performance if the hybridization characteristics of a substantial portion of the gene expression assays are profoundly affected in a few samples. These conclusions were drawn primarily from simulations in which the univariate distribution of simulated expression values was closely approximated by what might be expected for Affymetrix mRNA expression array data using a Robust Multi-array Average normalization approach. Their simulation design did not model the underlying expression covariance structure; however, the design was computationally feasible and its parameters (that is, a normal distribution

with a mean of 7 and a variance of 1, or a  $t$  distribution with a non-centrality parameter of 7 and 15 degrees of freedom) were easily defined and interpretable.

The Bottomly *et al.* study also included results from the analysis of Affymetrix expression data from 12 pediatric acute B-cell lymphoblastic lymphoma patients. Most notably, the application of the OD and  $Z$ -score heuristics to the Affymetrix expression scores for a particular sample returned a prioritized list that contained some biologically compelling candidate probe sets. Specifically, the sample was known to have had a single small interfering RNA (siRNA) hit and the prioritized list contained genes that would plausibly be dysregulated by such an event. These results demonstrate that the heuristics have the potential to deliver prioritized lists that contain at least some promising candidate aberrant expression values. However, the analysis of both simulated and real data suggests that the false-discovery rates of prioritized lists may be still be large, even within the top portion of the prioritized lists.

### From aberrant feature to candidate biomarker

The heuristics studied by Bottomly *et al.* are hypothesis generating in nature and their value lies in their ability to provide a prioritized list in which biologically and/or clinically meaningful expression aberrations 'rise to the top'. In order to be a valuable biomarker, a candidate assay must demonstrate properties to discriminate clinically relevant patient subgroups. In studies of the type considered by Bottomly *et al.*, the prioritized lists that are generated have unknown statistical significance and must be interrogated and refined by incorporating information external to the original experimental data. Since the statistical significance of the prioritized lists cannot be determined using the original data, the ability of the list to provide biologically/clinically relevant aberrant expression patterns can only be assessed in subsequent studies. For example, members of the prioritized OD candidate list could be interrogated for biological importance, their assay values could be validated by other means (such as quantitative PCR), and their behavior within publicly available datasets could be analyzed. The most promising candidates could be carried forward from these additional analyses and assayed and tested within a properly powered validation study that includes suitable control groups.

Biomarker study designs involving large-scale clinical samples (for instance, sample sizes in the thousands) are becoming more prevalent [3]. Indeed, such sample sizes may be required to validate and properly model biomarkers of reasonable effect size. On the other hand, smaller single-arm pilot studies may still contain invaluable information that can motivate the interrogation of other publicly available data sources, as well as guide the design and implementation of future studies. Bottomly

and colleagues have provided a useful and timely comparison of methods for extracting such information in the context of a common class of small-sample studies.

#### Abbreviations

OD: Outlying degree; siRNA: Small interfering RNA.

#### Competing interests

The authors declare that they have no competing interests.

Published: 29 November 2013

#### References

1. Alemi F, Erdman H, Griva I, Evans CH: **Improved statistical methods are needed to advance personalized medicine.** *Open Transl Med J* 2009, **1**:16–20.
2. De Sa J, Carlson B, Caputa N, Vojta D, Sandy L, Stevens S: **Growth of molecular diagnostics and genetic testing in the USA, 2008–2011: analysis and implications.** *Pers Med* 2013, **10**:785–792.
3. Ku CS, Roukos DH: **From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine.** *Expert Rev Med Devices* 2013, **10**:1–6.
4. *Patientslikeme.* <http://www.patientslikeme.com>.
5. McShane L, Cavenagh M, Lively T, Eberhard D, Bigbee W, Williams P, Mesirov J, Polley M-Y, Kim K, Tricoli J, Taylor J, Shuman D, Simon R, Doroshow J, Conley B: **Criteria for the use of omics-based predictors in clinical trials: explanation and elaboration.** *BMC Med* 2013, **11**:220.
6. Feifang H: **Statistical issues in trial design and personalized medicine.** *J Clin Invest* 2012, **2**:121–124.
7. Bottomly D, Ryabinin PA, Tyner JW, Chang BH, Loriaux MM, Druker BJ, McWeeney SK, Wilmot B: **Comparison of methods to identify aberrant expression patterns in individual patients: augmenting our toolkit for precision medicine.** *Genome Med* 2013, **5**:103.

doi:10.1186/gm507

**Cite this article as:** Gaile and Miecznikowski: From small studies to precision medicine: prioritizing candidate biomarkers. *Genome Medicine* 2013 **5**:104.