Blocking and Randomization to Improve Molecular Biomarker Discovery

Li-Xuan Qin\textsuperscript{1}, Qin Zhou\textsuperscript{1}, Faina Bogomolniy\textsuperscript{2}, Liliana Villafania\textsuperscript{3}, Narciso Olvera\textsuperscript{2}, Magali Cavatore\textsuperscript{3}, Jaya M. Satagopan\textsuperscript{1}, Colin B. Begg\textsuperscript{1}, and Douglas A. Levine\textsuperscript{2}

Abstract
Randomization and blocking have the potential to prevent the negative impacts of nonbiologic effects on molecular biomarker discovery. Their use in practice, however, has been scarce. To demonstrate the logistic feasibility and scientific benefits of randomization and blocking, we conducted a microRNA study of endometrial tumors (\textit{n} = 96) and ovarian tumors (\textit{n} = 96) using a blocked randomization design to control for nonbiologic effects; we profiled the same set of tumors for a second time using no blocking or randomization. We assessed empirical evidence of differential expression in the two studies. We performed simulations through virtual rehybridizations to further evaluate the effects of blocking and randomization. There was moderate and asymmetric differential expression (351/3,523, 10\%) between endometrial and ovarian tumors in the randomized dataset. Nonbiologic effects were observed in the nonrandomized dataset, and 1,934 markers (55\%) were called differentially expressed. Among them, 185 were deemed differentially expressed (185/351, 53\%) and 1,749 not differentially expressed (1,749/3,172, 55\%) in the randomized dataset. In simulations, when randomization was applied to all samples at once or within batches of samples balanced in tumor groups, blocking improved the true-positive rate from 0.95 to 0.97 and the false-positive rate from 0.02 to 0.002; when sample batches were unbalanced, randomization was associated with the true-positive rate (0.92) and the false-positive rate (0.10) regardless of blocking. Normalization improved the detection of true-positive markers but still retained sizeable false-positive markers. Randomization and blocking should be used in practice to more fully reap the benefits of genomics technologies. \textit{Clin Cancer Res}; 20(13); 3371--8. \textcopyright 2014 AACR.

Introduction
Technological advances in genomics profiling provide a plethora of molecular data to discover potential biomarkers for cancer diagnosis and treatment (1, 2). Discovering molecular biomarkers that are accurate and reproducible—the "needles in a haystack"—still remains challenging (3--5). A major source of the challenge comes from the nonbiologic effects in the data that are resulted from the experimental process (6, 7). Nonbiologic effects can be introduced into the data at numerous steps of the experimental process such as sample preparation, array hybridization, and image scanning (8). Previous efforts to remove nonbiologic effects have been mainly focused on \textit{post hoc} data adjustments through a "normalization" step (9, 10).

Careful study planning and sound experimental design offer a preventive opportunity to reduce the level of nonbiologic effects and mitigate their negative impact on biomarker discovery (11, 12). Two basic principles of experimental design are randomization and blocking (13, 14). Random assignment of experimental units to comparison groups eliminates bias due to known and unknown confounders. Randomization has been widely used in many scientific fields, including clinical studies, to determine the efficacy of experimental treatments (15). Arranging experimental units in blocks of similar units can reduce the variance and hence increase the power to detect differences between comparison groups (16). Many profiling platforms come with natural blocks. For example, the Agilent human micro-RNA (miRNA) arrays have eight arrays on each glass slide; the Illumina human gene expression BeadChips have twelve bead-chips on each slide; and the Illumina sequencer has eight lanes on each flow cell. Randomization and blocking have been previously suggested for use in genomic studies (17--19). However, their applications in practice are scarce possibly due to the lack of awareness and the conceived difficulties in logistic planning.

We set out to demonstrate the logistic feasibility and scientific benefits of randomization and blocking in molecular biomarker discovery so as to facilitate their adoption in cancer genomic studies. Toward this end, we profiled the same
set of tumor samples twice with different experimental designs, once using the blocked randomization design and uniform array handling and a second time using no blocking, randomization, or uniform handling. Empirical evidence of differential expression was assessed in both the randomized study and the nonrandomized study. The overall design of our study is illustrated in Fig. 1. This design is general and can be applied to many genomics platforms. In this study, we profiled miRNA expression for a set of 96 endometrial tumors and 96 ovarian tumors using Agilent miRNA arrays (20, 21). We present empirical evidence of confounding nonbiologic effects in the nonrandomized study. We compare the results of differential expression analysis for the randomized study versus that for the nonrandomized study. We consider the impact of a post hoc normalization step to remove nonbiologic effects. In addition to the empirical evaluations, we perform simulation studies to separately examine the effect of randomization and blocking when the true biomarker status is known and array handling is nonuniform. We also use simulations to assess the effect of randomization when samples come in multiple batches and each batch is randomized individually. Finally, we make recommendations on study design for cancer molecular biomarker studies.

**Materials and Methods**

**Patient samples**

Ninety-six high-grade serous ovarian cancer samples and 96 endometrioid endometrial cancer samples were used in our study. These samples were all newly diagnosed,
eventual cryomold embedding and RNA extraction for specimens. Freshly harvested tissue was snap frozen for Tumor microdissection and RNA extraction.

2000 and 2012. Their use in our study was approved by the MSKCC Institutional Review Board.

miRNA microarray hybridization

Human miRNA microarrays (Agilent Technologies), containing 3,523 markers that represent 1,205 human and 142 human viral miRNAs, were used for miRNA expression profiling according to the manufacturer’s protocol. For samples meeting the quality control standards, 200 ng of total RNA was labeled using the miRNA Complete Labeling and Hybridization Kit (Agilent). Labeled RNA was hybridized in Agilent Human miRNA Microarray Release 16.0. Slides were washed and scanned according to the manufacturer’s instructions. Images were quantified using Feature Extraction 10.7.3.1 (Agilent).

Study design for the randomized study

Agilent miRNA arrays have an 8-plex design, where each array slide contains eight individual arrays arranged as two rows and four columns. Therefore, each array slide represents an experimental block. We used the blocked randomization design when assigning arrays to sample groups to remove confounding nonbiologic effects. To achieve the best balance, we further required the blocks to be row and column balanced. That is, there were equal numbers of cases and controls on any row and any column of a slide. A total of six possible configurations achieve row and column balance for a 2 by 4 block such as the Agilent 8-plex array (Fig. 1). The 192 tumor samples used 192 arrays corresponding to 24 array slides. When implementing the blocked randomization design, we assigned the 24 slides to one of the six configurations with equal probabilities and consequently dedicated the arrays to one of the two sample groups. We then assigned each group of arrays to a random permutation of the samples in the corresponding group. We carefully planned our randomized study. All 24 slides were ordered from the same manufacture batch. Their hybridization and production were processed in one run by a single experienced technician.

Study design for the nonrandomized study

RNAs for the same 192 tumors used for the randomized study were rearrayed for a second study that used no blocking or randomization. Care was taken to ensure consistent sample handling for the two studies, and RNAs used for two arrays of the same tumor sample were taken from the same master stock. The 192 arrays for the nonrandomized study were ordered from multiple manufacturer batches. To mimic what would be done in the clinical setting, the arrays in the nonrandomized study were assigned to samples in the order of sample collection, and were processed in multiple runs by two technicians.

Statistical analysis of miRNA array data

Each marker on the Agilent array has multiple replicates (ranging from 10 to 40) on the array. Our data show minimum variation among replicates for the same marker (Supplementary Fig. S1). Hence, we summarized data from replicates for the same marker using the median. Evidence against the null hypothesis of equivalent expression was assessed using the t statistic comparing the endometrial and ovarian sample groups (22). A separate t test was performed for each of the 3,523 markers on the Agilent array and a two-sided P value was calculated. The P values were used to derive a marker set at a given significance level: markers with P values smaller than the significance level were declared differentially expressed, and those having larger P values were declared not differentially expressed. The resulting differential expression status of markers was compared between the randomized study and the nonrandomized study using a Venn diagram. Statistical analyses were conducted using R and Bioconductor.

Simulation study

Our goal for the simulation study was 3-fold: (i) to examine the effect of randomization and blocking when nonuniform handling is used and when the true biomarker status is known; (ii) to assess the effect of randomization and blocking separately; and (iii) to evaluate the effect of randomization when samples come in batches and randomization can only be done within each batch.

Toward this goal, we simulated data for the following scenarios of experimental designs for array-to-sample assignment, each combined with nonuniform array handling:

- a) blocking, randomization of all samples together;
- b) blocking, randomization within balanced batches of samples;
- c) blocking, randomization within unbalanced batches of samples;

miRNA Isolation Kit (Ambion) according to the manufacturer’s instructions. Images were quantified using Feature Extraction 10.7.3.1 (Agilent).
To mimic the characteristics of miRNA array data, we estimated the biologic effects of each sample and the nonbiologic effects of each array using the empirical data from the paired array studies and then simulated data through virtual rehybridizations of samples to arrays. Specific steps of the simulation study are as follows.

1. Biologic effects estimation: We approximated the biologic effects of each sample as its measurements in the randomized data, except that for not differently expressed markers in the ovarian samples mean differences between the two tumor groups were subtracted from the observed data.

2. Nonbiologic effects estimation: We estimated the nonbiologic effects for each array in the nonrandomized data as the difference between this array and its paired array in the randomized data.

3. Array-to-sample assignment: For each simulation, samples are randomly shuffled, and arrays are reassigned to samples according to an experimental design:
   a) Blocking, randomization of all samples together: four arrays from each slide are randomly selected and assigned to one tumor group, and the rest to another group.
   b) Blocking, randomization within balanced batches: twenty-four slides are randomly allocated to five batches each with equal number of samples from the two tumor groups; four arrays of each slide are randomly selected and assigned to one group, and the rest to another.
   c) Blocking, randomization within unbalanced batches: twenty-four slides are randomly allocated to five batches each with unequal number of samples from the two tumor groups (with the ratio of the two tumor types 1:3 or 3:1); four arrays of each slide are randomly selected and assigned to one group, and the rest to another.
   d) No blocking, randomization of all samples together: ninety-six arrays across all array slides are randomly selected and assigned to one tumor group, and the rest to another.
   e) No blocking, randomization within balanced batches: twenty-four slides are randomly allocated to five batches each with equal number of samples from the two tumor groups; half arrays across slides in each batch are randomly selected and assigned to one group, and the rest to another.
   f) No blocking, randomization within unbalanced batches: twenty-four slides are randomly allocated to five batches each with unequal number of samples from the two tumor groups; half arrays across slides in each batch are randomly selected and assigned to one group, and the rest to another.
   g) Blocking, no randomization: four arrays on the first row of each slide are assigned to one tumor group, and the rest to another group.
   h) No blocking, no randomization: the allocation of arrays to the tumor groups is kept the same as the nonrandomized study.
   i) Complete confounding of sample groups and array slides: twelve array slides are randomly assigned to one group, and the rest slides to another group.

4. Virtual hybridization: Simulated data were generated by summing the biologic effects of each sample and the nonbiologic effects of its reassigned array.

5. Analysis of each simulated dataset: Simulated data were analyzed for differential expression similar to the analysis of the nonrandomized empirical study.

6. Summary of simulated datasets: Steps 3 to 5 were repeated to generate 100 datasets for each experimental design. True-positive rate (TPR) and false-negative rate (FNR) are calculated and averaged across the 100 simulated datasets.

Results and Discussion

Empirical evaluation of molecular biomarkers in the randomized study

We compared miRNA expression levels between endometrial and ovarian tumors for each of the 3,523 markers on the Agilent array using the randomized data. Among them, 351 markers (10%) were differentially expressed at a P value cutoff of 0.01 (Fig. 2A and Supplementary Table S1). Among the 351 markers, 210 are overexpressed in ovarian tumors, six of which have a log, fold-change (L2FC) more than 1 with a maximum of 1.5; 141 are underexpressed, 13 of which have a L2FC more than 1 with a maximum of 3.3. By chance, 35 markers are expected to have a P value less than 0.01. The level of differential expression is moderately abundant in our data and the pattern of differential expression is not symmetric around 0 (Fig. 2B).

The six markers that are upregulated in ovarian tumors and have a L2FC more than 1 represent four unique microRNAs (miRNA), and the 13 downregulated markers represent eight unique miRNAs (Supplementary Table S2). miR-224, a gene located on the X-chromosome and thought to be active in mammalian ovaries, is the most upregulated in ovarian tumors with a L2FC of 1.5 (23). The other three upregulated miRNAs in ovarian tumors are miR-130a, miR-221, and its parologue miR-222, which all have been previously reported to be oncogenic (24–26).
miR-449a, which has been shown to be involved in vertebrate multiciliogenesis, is the most upregulated in endometrial tumors with a L2FC of 3.3 (27, 28). The other seven downregulated miRNAs are miR-9, miR-34a/b/c, miR-196b, miR-375, and miR-449b, which have been implicated in a number of cancer types (28–34).

We experimentally validated two not differentially expressed miRNAs and eight differentially expressed miRNAs (two upregulated in ovarian tumors and six downregulated) using qRT-PCR. The differential expression status was confirmed for all 10 miRNAs (Fig. 2C).

Empirical evaluation of molecular biomarkers in the nonrandomized study

Figure 3A and B shows the distribution of the nonrandomized data without normalization in comparison with the distribution of the randomized data. It indicates clear nonbiologic effects in the nonrandomized data. When comparing miRNA expression between endometrial and ovarian tumors in the nonrandomized data, 1,934 markers (55%) were called differentially expressed at a \( P \) value cutoff of 0.01. Among these markers, 181 were called differentially expressed in the randomized data.
Clinical Cancer Research

Figure 3. A, boxplot of the randomized data. B, boxplot of the nonrandomized data before normalization. C, Venn diagram comparing differentially expressed markers identified by the nonrandomized data before normalization and those by the randomized data. D, Venn diagram comparing differentially expressed markers identified by the nonrandomized data after quantile normalization and those by the randomized data.

(185/351, 53%); 1,749 were called not differentially expressed in the randomized data (1,749/3,172, 53%; Fig. 3C). With quantile normalization, the number of differentially expressed markers identified in the nonrandomized data was reduced to 708. Among them, 328 were differentially expressed (328/351, 93%) and 380
were not differentially expressed (380/3,172, 12%) in the randomized data (Fig. 3D).

Simulation study

The results of the simulation study are summarized in Table 1 and graphically displayed in Supplementary Fig. S2.

Not surprisingly, blocking and randomization applied to all samples at once led to a best accuracy of biomarker detection with the highest TPR (0.97, 340/351) and lowest FPR (0.002, 3/3,172). This level of accuracy was equally, if not slightly better, achieved by blocking and randomization applied to sample batches that were balanced in sample groups (TPR = 0.97, FPR = 0.001 = 3/3,172). The next best level of biomarker detection accuracy was achieved when the array-to-sample assignment used randomization across all samples (TPR = 0.95 = 333/351, FPR = 0.02 = 63/3,172) or randomization within balanced batches of samples (TPR = 0.95, FPR = 0.01) but with no blocking. When randomization was used for samples that came in batches unbalanced in sample groups, the accuracy of biomarker detection had a slightly worse TPR (0.92, 323/351) and a much worse FPR (0.10–0.17, 312–539/3,172) regardless of whether blocking was used within each batch. As expected, complete confounding of sample group and array slide resulted in a low TPR (0.89 = 312/351) and a high FPR (0.20 = 634/3,172).

When no randomization was used, our simulation showed a high accuracy of biomarker detection (TPR = 0.97, FPR = 0.001) with blocking, and a low detection accuracy without blocking (TPR = 0.84 = 295/351, FPR = 0.07 = 222/3,172). We note that this result depends on the actual level of confounding between sample group and array slide in the data. In this simulation, we used the array-to-sample assignment in the empirical nonrandomized study, where the sample group was partly confounded with array slides with the number of endometrial tumors being 0, 1, 2, 3, 5, 6, and 7 for 4, 2, 1, 4, 2, 7, and 4 array slides, respectively (Supplementary Table S3).

Regardless of the experimental design, the post hoc normalization step, using quantile normalization in our study, improved the detection of true-positive markers to an almost perfect level and at the same time still had false-positive markers with a FPR often around 0.20, corresponding to a FDR of about 64%.

We used P values at a cutoff of 0.01 to call significance in our study. We tested other P value cutoffs and observed similar results (results not shown). We also repeated our analysis using the FDR to adjust for multiple comparisons when calling significant markers, and observed similar results on the relative ranking of the experimental designs (results not shown).

Taken together, the results from our simulation study further underline the importance of blocking and randomization for accurate detection of disease relevant markers. Even blocking alone, or randomization alone, or randomization within balanced batches alone offers significant benefits in accurate biomarker detection. Post hoc data adjustments through normalization improve the identification of true-positive markers, but still possess a large number of false-positive markers.

Conclusions

We have demonstrated that blocking and randomization are valuable for miRNA array studies of tumor samples. We also showed, through both empirical and simulated studies, that blocking, randomization across all samples, or randomization within balanced sample batches can effectively remove confounding nonbiologic effects and ensure the accuracy of detecting disease relevant markers with both a high TPR and a low FPR. We observed that post hoc normalization to the data improves the detection of true-positive markers and at the same time retains a large number of false-positive markers in our study.

In practice, we recommend use of the block-type design of profiling platforms, placing equal numbers of samples from each comparison group in each block to the extent possible. If blocking cannot be done, one should randomize...
the samples altogether if they are available (e.g., in a small-scale study or a retrospective study using banked tissues), or batch samples so that each batch is balanced in comparison groups and then randomize within each batch.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: L.-X. Qin, D.A. Levine
Development of methodology: L.-X. Qin
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Villafania, M. Cavatore, D.A. Levine
Analysis and interpretation of data (e.g., statistical analysis, bioinformatics, computational analysis): L.-X. Qin, Q. Zhou
Writing, review, and/or revision of the manuscript: L.-X. Qin, J.M. Satagopan, C.B. Begg, D.A. Levine

References
Blocking and Randomization to Improve Molecular Biomarker Discovery

Li-Xuan Qin, Qin Zhou, Faina Bogomolniy, et al.

Clin Cancer Res 2014;20:3371-3378. Published OnlineFirst May 1, 2014.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-3155

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/05/01/1078-0432.CCR-13-3155.DC1

Cited articles
This article cites 26 articles, 7 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/20/13/3371.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/20/13/3371.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/20/13/3371.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.