Prediction of cytogenetic abnormalities with gene expression profiles

Yiming Zhou, Qing Zhang, Owen Stephens, Christoph J. Heuck, Erming Tian, Jeffrey R. Sawyer, Marie-Astrid Cartron-Mizeracki, Pingping Qu, Jason Keller, Joshua Epstein, Bart Barlogie, and John D. Shaughnessy Jr

Cytogenetic abnormalities are important clinical parameters in various types of cancer, including multiple myeloma. We developed a model to predict cytogenetic abnormalities in patients with multiple myeloma using gene expression profiling and validated it by different cytogenetic techniques. The model has an accuracy rate up to 0.89. These results provide proof of concept for the hypothesis that gene expression profiling is a superior genomic method for clinical molecular diagnosis and/or prognosis. (Blood. 2012;119(21):e148-e150)

Introduction

Multiple myeloma (MM), a neoplasm of plasma cells, is characterized by complex chromosomal abnormalities, including structural and numerical rearrangements.1,2 The cytogenetic abnormalities (CAs) that are a hallmark of MM and other cancers are commonly used as clinical parameters for determining disease stage and guiding therapy decisions for patients.3,4 Traditional cytogenetic techniques, including FISH, metaphase karyotyping, and the recently developed array-based comparative genomic hybridization (aCGH), are widely used to detect chromosomal aberrations and gene copy-number changes. These methods, however, are expensive, time-consuming, or both. We describe here a virtual CA (vCA) model that uses gene expression profiling (GEP) to predict CA. The rationale for the model is that disease-associated alterations of genomic regions should, in some way, alter (“drive”) expression levels of target genes within the regions or nearby; otherwise, the genomic alterations would not contribute to the disease. Therefore, we thought it reasonable to hypothesize that the driving alterations should be predictable via the alteration of expression levels of the genomic region’s target genes. We provide proof of concept that GEP offers a superior data source for molecular diagnosis and/or prognosis.

Methods

Study subjects

Bone marrow aspirates were obtained from patients newly diagnosed with MM, who were subsequently treated on National Institutes of Health-sponsored clinical trials. Patients provided samples under Institutional Review Board–approved informed consent in accordance with the Declaration of Helsinki, and records are kept on file. Myeloma plasma cells were isolated from heparinized bone marrow aspirates with an autoMACS device (Miltenyi Biotec) using CD138-based immunomagnetic bead selection, as previously described.5

DNA isolation and aCGH

High-molecular-weight genomic DNA was isolated from aliquots of CD138-enriched plasma cells with the use of the QIAamp DNA mini kit (QIAGEN). Tumor- and sex-matched reference genomic DNA (Promega) was hybridized to the Agilent 244K aCGH array according to the manufacturer’s instructions (Agilent Technologies).

Interphase FISH

Bone marrow aspirates from patients with MM were first subjected to Ficoll-Hypaque gradient-centrifugation separation to remove erythrocytes. Copy-number changes in myeloma plasma cells were detected by triplicecolor interphase FISH analysis of chromosome loci, as previously described.3 Bacterial artificial chromosome clones specific for 1q21 (CKS1B), 1p13 (AHCYL1), 13q14 (D13S31), and 13q34 (D13S285) were obtained from BACPAC Resources Center and labeled with Spectrum Red– or Spectrum Green–conjugated nucleotides via nick translation (Vysis). At least 100 myeloma cells stained with immunoglobulin light-chain antibody (κ or λ) conjugated with 7-aminomethylcoumarin-3-acetic acid were counted for copies of each probe. The threshold of significant abnormality was set at 2.5 for amplification of 1q and at 1.5 for deletion of 1p and chr13, according to the distribution of the FISH signals (supplemental Figure 1; see the Supplemental Materials link at the top of the article).

Cyto genetics

Bone marrow was processed for chromosome studies by standard techniques. A direct-harvest, 24-hour unsynchronized culture and a 48-hour synchronized culture were used on most specimens. Colcemid (0.05 μg/mL) was added for 1 hour. For the purpose of cytogenetic examination, an effort was made to examine at least 20 metaphases, with the application of Giemsa banding techniques. The presence of CA required the detection of at least 2 abnormal metaphases in cases of hyperdiploidy and translocations, whereas at least 3 metaphases with clonal abnormalities were required in cases of whole and partial chromosome deletions.

RNA purification and microarray hybridization

RNA purification, cDNA synthesis, cRNA preparation, and hybridization to the Human Genome U133Plus 2.0 GeneChip microarray (Affymetrix) were performed as previously described.

Data analyses

A modified Lowess algorithm was used to normalize aCGH data.11 Statistically, altered regions were identified with the use of a circular binary
Results and discussion

We determined genome-wide gene expression profiles and DNA copy numbers in purified plasma cell samples obtained from 92 newly diagnosed MM patients, using the Affymetrix GeneChip and the Agilent aCGH platforms, respectively. The details of the study subjects, procedures, and data analyses are provided in "Methods." DNA copy number-sensitive genes were determined by PCC of gene expression levels and the copy numbers of the corresponding DNA loci. Applying the criterion of PCC more than 0.35, which kept the false discovery rate to less than or equal to 5%, we identified 1114 copy number-sensitive genes (supplemental Table 1).

On the basis of these copy number-sensitive genes, we developed a vCA model for predicting CAs in MM patients by GEP. The model focuses particularly on the uneven chromosomes, as well as the 1p, 1q, and 6q segments, which are the most commonly altered chromosome regions in myeloma plasma cells.

The reference CAs of a given chromosome region were determined by the mean values of signals of aCGH probes located in that region. We set the cutoff values at 0.45 for amplification and −0.45 for deletion, as there were only 1% greater than 0.45 on the basis of the absolute signals of probes located in chromosomes 2, 4, 10, and 12, which are the most stable chromosomes in myeloma cells. The values of reference CAs could be used to distinguish among amplification, deletion, and normal.

The predicted CAs (pCAs) of a given chromosome region were determined by the following procedures. First, we calculated the mean expression levels of copy number-sensitive genes within the region. Then, by training the model in a GEP dataset with 92 MM samples, we set the cutoff value of the mean expression levels of copy number-sensitive genes for each chromosome region to obtain pCAs that were most consistent with reference CAs in terms of the Matthews correlation coefficient, a measure of the quality of binary (2-class) classifications.

The mean prediction accuracy was 0.88 (range, 0.59-0.99; Table 1; supplemental Table 2) when the model was applied to the training dataset. To check for overfitting in the vCA model, we applied the model to an independent dataset of 23 MM samples for which both GEP and aCGH data were available. The mean prediction accuracy was 0.89 (range, 0.74-1.00; Table 1; supplemental Table 3), which indicated that overfitting was negligible if present at all.

We validated the model with a FISH dataset compiled from 262 independent MM samples for which both FISH records and GEP data were available. All 262 MM samples had been tested with 1p (AHCRYL1) and 1q (CKS1B) probes. Of these samples, 195 had also been tested with chromosome 13 probes (D13S31 and D13S285). The cutoff value was set at 2.5 for amplification of 1q and at 1.5 for deletion of 1p and chr13, according to the distribution of the FISH signals (supplemental Figure 1). Applying the vCA model to the GEP data, we determined pCA for the 262 samples. The pCA results were well matched with the FISH reports. The mean prediction accuracy was 0.87 (range, 0.82-0.90; Table 1; supplemental Table 4).

In a further validation of the vCA model, we compiled a set of cytogenetic data generated by conventional metaphase karyotyping that included 533 independent MM samples for which both karyotype records and GEP data were available. Applying the vCA model to the GEP data, we determined the pCA for the 533 samples. The pCA results were matched to the karyotype reports with a mean prediction accuracy of 0.65 (range, 0.36-0.77; Table 1; supplemental Table 5). The consistency of the matching was lower than those of pCA versus aCGH and pCA versus FISH. This prediction underperformance may be because karyotyping is limited by the low proliferation rate of terminally differentiated plasma cells in vitro and also by only detecting the cytogenetic information for cells at metaphase, thus missing a considerable amount of information regarding the copy number of DNA in a tumor cell population. We hypothesize that FISH reports would also not match karyotype records well. To test this hypothesis, we compared the FISH and karyotype data for the 262 samples for which both records were available. Indeed, the prediction accuracies between FISH and karyotype records for chr1q21, chr1p13, and chr13 were 0.76, 0.83, and 0.60, respectively (supplemental Table 6), which were similar to the prediction accuracies between pCA and karyotype (0.72, 0.75, and 0.64, respectively; supplemental Table 5).

The vCA method does not effectively predict chr17p deletion. Our aCGH data (supplemental Figure 2) demonstrated that TP53 is the most common locus deleted in patients with del17p. Although TP53 is a copy number sensitive gene, its expression is also controlled by mutations, resulting in a wide range of expression levels, even in patients with 2 copies of the gene, hence a low predictive power (supplemental Figure 3). The frequency of TP53 gene deletion detected by FISH in MM patients is approximately 10%14,15 (7% in our aCGH data). Nevertheless, the frequency of TP53 mutations in MM detected by whole-genome sequencing is approximately 8%.16 Therefore, patients without chr17p deletion can have very low TP53 expression (supplemental Figure 3). Because the main merit of FISH for chr17p is detection of TP53 levels, the TP53 expression level from GEP is the relevant prognostic parameter.

We previously reported that GEP predicts chromosome translocations.5,8 Here we demonstrate that GEP can predict other
chromosomal abnormalities, providing proof of concept that GEP yields most of the relevant prognostic information obtained through traditional karyotyping methods and aCGH. Despite its inability to always accurately distinguish changes in gene expression resulting from deletions, amplifications, or mutations, analysis of GEP data alone offers a superior prognostic tool for MM and potentially other malignancies.

Acknowledgments
This work was supported by the National Cancer Institute (grant CA055819).

Authorship
Contribution: Y.Z., Q.Z., B.B., and J.D.S. designed research; Y.Z. and Q.Z. performed the analysis; O.S., E.T., and J.R.S. prepared biologic samples and performed experiments; M.-A.C.-M., P.Q., and J.K. compiled the data; B.B. supervised the clinical study; and Y.Z., Q.Z., C.J.H., J.E., and J.D.S. wrote the manuscript.

Conflict-of-interest disclosure: Y.Z., Q.Z., B.B., and J.D.S. have filed patents on technology reported in this manuscript. J.D.S. is a founder and has an ownership stake in Signal Genetics LLC, a biotechnology company that has licensed said technology from the University of Arkansas for purposes of commercial development. J.D.S. holds patents or has submitted patent applications on the use of GEP in cancer medicine. Y.Z., P.Q., and B.B. are coinventors on patents and patent applications related to the use of GEP in cancer medicine. The remaining authors declare no competing financial interests.

Correspondence: Qing Zhang, 4301 W Markham St, Little Rock, AR 72205; e-mail: qzhang@uams.edu; and John D. Shaughnessy Jr, 4301 W Markham St, Little Rock, AR 72205; e-mail: jdsjr@me.com.

References
Prediction of cytogenetic abnormalities with gene expression profiles

Yiming Zhou, Qing Zhang, Owen Stephens, Christoph J. Heuck, Erming Tian, Jeffrey R. Sawyer, Marie-Astrid Cartron-Mizeracki, Pingping Qu, Jason Keller, Joshua Epstein, Bart Barlogie and John D. Shaughnessy Jr

Updated information and services can be found at:
http://www.bloodjournal.org/content/119/21/e148.full.html

Articles on similar topics can be found in the following Blood collections
- e-Blood (154 articles)
- Lymphoid Neoplasia (2506 articles)
- Multiple Myeloma (361 articles)

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml