

Identifying subclass-independent master regulators of relapse in pediatric acute myeloid leukemia

Yishai Shimoni Jing He Gonzalo Lopez Paola Nicoletti Mariano Alvarez
Andrea Califano

June 5, 2013

1 introduction

Acute myeloid leukemia (AML) is caused by a maturation arrest and by enhanced proliferation of myeloid precursor cells. This includes precursors of red blood cells, platelets, granulocytes, or monocytes. As a result, AML has very diverse symptoms, diverse presentation, and a diverse gene expression signature. Nevertheless, it is considered as a single disease and most variants are treated similarly. AML incidence has peaks in infants under one year old. It then remains low through adolescence, when it starts to increase from 15 per 10^5 per year in early adulthood to 35 per 10^5 per year at the age of 90 years. In adults, AML may arise de-novo from a somatic mutation, or it may be caused by some underlying syndrome, or by exposure to irradiation or to chemotherapy. The latter causes are much less common in pediatric patients.

AML has been clinically divided into several sub-classes, known as the French-American-British (FAB) classification, based on cytogenic abnormalities and on morphological characteristics of the leukemic cells. These subclasses are denoted M0 to M7. More recently, patient stratification is done by risk-group, which determines the aggressiveness of the treatment the patient receives. Patients with monosomy7, monosomy5, del5q, or FLT3/ITD are considered high risk, while patients presenting t(8;21), Inv(16), NPM mutations (FLT3/ITD negative), or CEBPA mutations are considered to have a more favorable prognosis.

In both classification strategies all patients, except M3 patients, receive similar chemotherapy, consisting of a catarabine/anthracycline backbone. The intensity of the treatment is determined by the risk stratification, by the response to treatment, and by the adverse effects of treatment. Approximately 90% of patients achieve initial remission, but between 30% and 40% experience a relapse of the disease within two years. The molecular or genetic reasons for relapse are largely unknown, and patients who relapse are usually non-responsive to further chemotherapy. Early indications point to a benefit of stem cell transplantation. However, currently, most relapse patients require irradiation and bone marrow transplantation, which have a low success rate. Thus, mortality in relapse patients is close to 80%, and overall mortality is between 30% and 40%.

Pediatric AML is known to be similar to adult AML both in clinical presentation and in the prevalent mutations and cytogenic abnormalities. The main difference between the two categories is thought to be in the abundance of each of the mutations and abnormalities. However, an important advantage of using pediatric AML data is the lack of patients who develop AML as a result of an underlying condition or previous treatment. Additionally, the diversity in environmental exposure is smaller in such patients compared to the general adult population.

We therefore focus on data on pediatric AML from the therapeutically applicable research to generate effective treatment (TARGET) initiative. Pre-treatment gene expression data for 215 patients is available (using Affymetrix gene ST arrays), as well as copy number variation (CNV) data (using Affymetrix SNP 6.0 arrays). Additionally, whole exome sequencing was performed for 22 patients with matching normal blood samples, pre-treatment bone marrow samples, and post-relapse bone marrow samples.

(b)

Figure 1: FAB classes can act as confounding factors. a) Proportions of event-free survival statuses per subclass. The analysis was only performed on subclasses with more than 20 patients. The proportions of patients who have undergone relapse (blue), induction failure (purple), death without remission (green), death (yellow) or censored (red). It can be seen that M5 has the lowest proportion of censored patients, followed by M4, with M2 and M1 having the most favorable prognosis. b) Event-free survival plots for FAB subclasses. The survival curves show that the four clinical subclasses show some separation in prognosis, where M5 patients have the worse prognosis, followed by M4 and M2, and M1 patients having the best prognosis.

In this research we use master regulator analysis methods to identify genes that are responsible for relapse in AML patients. We use multiple reciprocal computational analysis methods to find the candidate genes, and validate these predictions using independent data sets, and using mouse models.

2 Results

2.1 Checking for confounding factors

Master regulator analysis (MARINA) looks at each potential regulatory gene and the set of genes it is predicted to regulate (its regulon). If the expression of the regulon changes in a way that is consistent with the predicted effect of the regulator on the gene (e.g. that a gene that most of the genes that are predicted to be up-regulated are differentially over expressed in the phenotype compared to normal, and most of the genes that are predicted to be down-regulated are differentially under expressed in the phenotype compared to normal), we call such a gene a master regulator (MR) of the change in phenotype. In this context the phenotype of interest are patients who undergo relapse, and the controls are censored patients.

One potential problem with this type of analysis may arise if the control or phenotype patients are not uniformly distributed between the subclasses in the data. In this scenario, MARINA might find the MRs of the subclasses, instead of the MRs of the relapse. Fig. 1a compares the distribution of relapse and censored patients in each of the FAB subclasses that have sufficient numbers of patients (M1, M2, M4, and M5), where we see that indeed there is a bias such that the percentage of relapse patients increases from M1 to M2 to M4 and is highest for M5. A similar result can be seen in Fig. ??, where the survival curves are shown for each of the subclasses. Here, too, class M5 has the worse prognosis, followed by M4.

An additional problem may arise if the subclasses are sufficiently distinct from each other, so that the mechanism leading to relapse may be different in each subclass. This can be manifested in the amount of separation in the gene expression profile of patients belonging to each subclass. Fig. 1b shows a t-SNE dimensional reduction of the gene expression data, colored by their FAB classification, and with markers indicating outcome. This dimensional reduction clearly shows a tendency for samples from the same clinical subclasses to be closer together, and raises the question whether a single analysis for relapse can be performed across all subclasses. However, the subclasses do not separate into distinct clusters, thus still allowing for the possibility that the subclasses represent gradual differences in mechanism rather than distinct disease models.

A similar analysis was done for each one of the clinical variables that were available, where it was found that expression is biased by the following variables: t(8;21), t(9;11)(p22;q23), inv(16), complexity, FLT/ITD3 positive, NPM mutation and CEBPA mutation (plots available in file `tsne_clinical_variables.pdf`). However, these variables are all taken under consideration when determining the FAB classification and are therefore controlled for when selecting FAB classifications as the classifiers.

A more current method of patient stratification is by risk group, as determined by the several criteria, as described below. Low risk patients are patients with one of the following: t(8;21), Inv(16), NPM mutations

(b)

Figure 2: Dimensional reduction of gene expression data, colored by FAB classification. T-SNE was used to visually display the expression data. Each point represents a sample, and samples that are close have a similar gene expression profile. The shape of the markers represents the event-free status (square - relapse, circle-censored, triangle - death without remission). a) Samples are colored by their FAB classification FAB classes M2, M4, and M5 exhibit similarity, but they do not form distinct clusters; b) Samples are colored by their risk group (low, standard and igh risk are red, green and blue, respectively). The groups present some separation, but do not offer better separation than the FAB classifications.

Figure 3: Event-free survival curves for risk groups. The survival curves show good separation between risk groups, especially good prognosis of the low risk group. The separation between the high and standard risk groups

and FLT3/ITD negative, or CEBPA mutations. High risk patients with monosomy7, monosomy5, del5q, or FLT3/ITD positive. A t-SNE analysis colored by the risk groups (Fig. ??) shows that the risk groups provide high separation and high consistency in their gene experssion pattern, but do not provide further separation compared to the FAB classification. Furthermore, Kaplan-Meier survival curves of the risk groups (Fig. 3) show that the separation between the low and standard risk groups is not large. We therefore use FAB classifications for the rest of the analysis.

MARINA analysis was run using the R package `marina.package.version('marina')` on each FAB subclass, where control patients were censored patients, and the phenotype was defined by relapse patients. We excluded patients who were reported as "Death", "Death without remission", or "Induction failure". The regulon that we use was created by running ARACNe on the same expression data. Additionally, in order to allow estimating the reproducibility of the results we randomly choose a training and validation subsets in the ratio of 3:1, respectively.

The analysis sorts all genes according to their differential expression in relapse patients compared to censored samples. Then, for each gene in the list of potential regulators a list is created of the target genes it downregulates and the genes it upregulates. The gene-set enrichment is calculated for each set and the sum of these enrichment scores represents the change in activity of that regulator between the two phenotypes. Fig. 4 shows the top 30 predicted master regulators (MRs) for each subclass, showing, from left to right: the calculated p-value; the location of the activated genes (red) and repressed genes (blue) in the ranked list of genes (ranked by differential expression), the change in activity (blue - low to red - high), the differential expression, and the rank of the change in expression.

The results from all the marina analysis were combined using Stauffer's method, and after Bonferoni correction we are left with the follwing master regulators: Clustering the validation data using the activity of the significant master regulators (Fig. 5A) reveals 4 clusters which separate the classes by their relapse, but not by any other clinical information (such as the FAB classification or their risk group). This can also be seen by the survival curves for each one of these clusters (Fig. 5B).

(b)

(d)

Figure 4: MARINA results for each FAB class with sufficient samples.

(b)

(d)

Figure 5: Hierarchical clustering and survival plots of the training samples according to the virtual proteomics of the top master regulators, where the separation into clusters was performed using the top 4 branches of the hierarchical clustering. a) Hierarchical clustering using activity of TF MRs; b) relapse plots for the groups defined by the hierarchical clustering of TF MRs; c) a) Hierarchical clustering using activity of signaling MRs; b) relapse plots for the groups defined by the hierarchical clustering of signaling MRs

The above analysis was performed to identify transcriptional master regulators. Gene expression of the downstream genes of TFs may be a good proxy for their activity levels, but they are normally difficult targets for drug therapy, and they also tend to be poor biomarkers. We therefore ran the same analysis using signaling molecules. The results for signaling molecules may be more noisy, but they have a better chance to be relevant drug targets or biomarkers.

This analysis provides an additional master regulators: The analysis using the signaling molecules also allows stratification of the validation samples into 4 groups that have distinct relapse curves. (Fig. 5C,D)

To try and get results that are more consistent across other datasets it is possible to run the same analysis using regulons that were derived from other datasets. One of the most relevant datasets may be the AML data provided by the cancer genome atlas (TCGA). This data consists of gene expression data of RNA-seq data for 179 adult AML patients. Although this data-set involves a different platform and adult patients, since the diseases are considered clinically similar it is hypothesised that the results may be similar, and that any result that comes up in both datasets may be more relevant. We therefore performed a similar analysis using a regulon derived from the TCGA data.

This provides a similar list of TF and signaling MRs. The overlapping MRs are: for the TFs and for signaling molecules.

Finally, we would like to see whether these MRs also apply for an independent validation data set. One way to try this is using the TCGA data for adult AML patients. Although there are some known age-related differences in gene expression, the diseases are clinically very similar, which leads to the hypothesis that the relapse mechanism should be similar in both diseases. We therefore use the gene expression data of the master regulators determined above to classify the samples in the TCGA dataset.