Available online at www.annclinlabsci.org

Bioinformatics Analysis to Determine Prognostic Mutations of 72 *de novo* Acute Myeloid Leukemia Cases from the Cancer Genome Atlas (TCGA) with 23 Most Common Mutations and no Abnormal Cytogenetics

Kerry J. Welsh, Elena Nedelcu, Amer Wahed, Yu Bai, Amitava Dasgupta, and Andy Nguyen

Department of Pathology and Laboratory Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA

Part of the contents in the manuscript has been presented at the 135th Meeting of the Association of Clinical Scientists, Salt Lake City, Utah, May 13-16, 2015.

Abstract. ***Objectives.*** Up to 40% of acute myeloid leukemia (AML) patients have normal cytogenetics (CN-AML) but they may have gene mutations. An important issue in the treatment of CN-AML is how gene mutation patterns may help with patient management. The Cancer Genome Atlas (TCGA) database has data from 200 cases of de novo AML including cytogenetics, gene mutations, and survival duration (prognosis). ***Methods.*** Cases with the most common mutations and no cytogenetic abnormalities were selected from the TCGA. Unsupervised neural network analysis was performed to group them into clusters according to their pattern of mutations and survival. ***Results.*** 72 cases of CN-AML with the 23 most common mutations were obtained from TCGA. Clustering was found to be based on 6 mutations, with the following prognostic groups: (a) good: NPM1, CEBPA, or TET2, (b) intermediate: NPM1/DNMT3A, or other mutations, (c) poor: RUNX1, FLT3-ITD, FLT3-ITD/NPM1, or FLT3-ITD/CEBPA. Some discrepancy between our results and those from previous studies is most likely due to inclusion of AML cases transformed from myeloproliferative neoplasms or myelodysplastic syndrome in previous studies. ***Conclusions.*** This study provides further molecular characterization and prognostic data most specific for the de novo subgroup of CN-AML patients.

Keywords: acute myeloid leukemia, The Cancer Genome Atlas, prognostic mutations

Introduction

AML is a heterogeneous malignancy with many karyotypic and molecular abnormalities. The discovery of the recurrent karyotype abnormalities in AML such as the t(15;17) has been invaluable in obtaining more accurate prognostic information, the development of specific therapies, and molecular monitoring. The current WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues [1] includes nine recurrent karyotypic abnormalities in AML, which are critical for initial diagnosis and treatment. However, approximately half of AML patients have no karyotype abnormality (CN-AML) [2]. This group of AML cases is presumably heterogeneous in all respects, and molecular monitoring is not possible unless there is an associated mutation. Recently it has been demonstrated that mutations of FLT3-ITD, NPM1, and CEBPA genes are preferentially found in CN-AML with significant prognostic association. It is important to note that many CN-AML cases do not possess those three mutations but have other mutations. There have been efforts to detect a large number of selected gene mutations in AML cases using next generation sequencing studies to expand the prognostic tools [2-8]. However, a challenge in exploring the prognostic value of mutations is that once cells become cancerous due to mutation, they are prone to even greater rates of mutation as their self-control and repair mechanisms fail, resulting in genomic makeup that is very heterogeneous. It is important to develop robust techniques to distinguish biologically significant mutations from other background mutations.

Address correspondence to Andy Nguyen, MD; Department of Pathology and Laboratory Medicine, University of Texas Health Science Center at Houston, 6431 Fannin, MSB 2.224, Houston, TX 77030, USA; phone: (713) 500 5337; fax (713) 500 0712; e mail: Nghia.D.Nguyen@uth.tmc.edu

CASE	FLT3	NPM1	DNMT3A	IDH2	IDH1	TET2	RUNX1	NRAS	CEBPA	WT1
34	0	0	0	0	0	0	1	0	0	0
42	0	0	0	0	0	0	1	0	0	0
49	0	0	0	1	0	0	1	0	0	0
70	0	0	0	0	0	0	1	0	0	0
76	0	0	0	0	0	0	1	0	0	0
79	0	0	0	1	0	0	1	0	0	0
89	0	0	0	1	0	0	0	0	0	0
97	0	0	0	1	0	0	1	0	0	0
102	0	0	1	1	0	0	0	0	0	0
106	1	0	0	0	0	0	1	0	0	1
119	0	0	0	0	0	0	1	0	0	0
137	0	0	0	1	0	0	0	0	0	0
142	0	0	0	0	0	0	1	0	0	1

Figure 1. *Input data for Neural Network (partial view).*

A pressing issue in treatment of CN-AML is how gene mutation patterns may help physicians guide the management of patients in daily practice. Given the large number of mutations already described and the fact that they often partially overlap, the definition of a standardized and well-accepted prognostic algorithm based on mutation patterns will not be an easy task [9]. Up to 85% of CN-AML patients have mutations in at least one of the following: ASXL1, NPM1, FLT3-ITD, TET2, IDH1/2, and RUNX1. Analysis of mutations in AML is complicated by the difference in types of mutations in de novo (primary) cases versus secondary cases (transformed from myelodysplastic syndrome and myeloproliferative neoplasm). For individual mutations, NPM1, FLT3-ITD, and DNMT3A mutations are found significantly more often in de novo CN-AML than in secondary AML cases [2]. It has been shown that the most frequently mutated genes in secondary cases are ASXL1 and TET2 [2]. NRAS, JAK2, SF3B1, and TP53 mutations are mostly present in secondary AML cases [2]. Furthermore, there is no consistent pattern in mutation acquisition during disease progression in secondary cases. It would be most feasible to limit mutation study of CN-AML to de novo AML cases to extract useful data on prognosis.

Somatic mutations can be categorized into driving mutations, changes that are responsible for disease pathogenesis, and passenger mutations, which are simply the byproduct of the unstable cancer genome and provide no functional role to tumor cells. Distinguishing these two mutation types is important since driving mutations are considered ideal target for therapy whereas treatment that affects passenger mutations is likely to be ineffective. Driving mutations are likely to be seen in a large cohort of samples. Furthermore, it is believed that these mutations should result in changes in protein structure and function with significant effect on cancer pathway. Some passenger mutations may also be seen in high frequency, making it difficult to separate them from the driving mutations with true functional impact.

In this study, we use data from the TCGA database [3] which consists of 200 de novo AML cases and utilize clustering analysis to correlate the presence of significant mutations to prognosis of CN-AML cases. Our results are then compared to those from published data to obtain more insight into this subgroup of de novo CN-AML cases.

Materials and Methods

Data from 200 cases of de novo AML were retrieved from TCGA database (public domain) [3]. Demographic information shows: age 55±16.1, white 89%, black 8%, others 3%, male 54%, female 46%, normal cytogenetics 47%. Molecular testing was performed on multiple platforms: Affymetrix U133 Plus 2, Illumina Infinium Human Methylation 450 BeadChip, and Affymetrix SNP Array 6.0. All karyotypes were analyzed by conventional G-banding in at least 20 metaphases. Results are available for cytogenetics, 260 gene mutations, and survival duration (day to death) for each case [4]. As previously reported in this database, a total of 23 genes were significantly mutated, and another 237 were mutated in two or more samples [4]. Nearly all samples had at least 1 non-synonymous mutation. For meaningful clustering analysis, data for only cases with the following 23 most common mutations (grouped according to categories) were extracted:

-Activated signaling (signal transduction): FLT3-ITD, KIT, KRAS, NRAS, PTPN11

-Myeloid transcription factors (differentiation): NPM1, CEBPA, and RUNX1
-Epigenetic regulation: DNMT3A, TET2, IDH2, IDH1, EZH2, HNRNPK
-Tumor suppressors: TP53, WT1, PHF6
-Spliceosomes: U2AF1
-Cohesins: SMC1A, SMC3, STAG2, RAD21,
-Non-annotated: FAM5C (BRINP3)

Subsequently, only cases with positivity for these 23 mutations and no cytogenetics abnormalities were obtained. Unsupervised neural network analysis with NeuroXL Clusterizer (OLSOFT LLC, Moscow) was performed on these cases to group them into clusters according to their pattern of mutations and survival duration. Neural networks are a proven, widely used technology to solve complex classification problems. Loosely modeled after the human brain, neural networks are interconnected networks of independent processors that can learn the solution to a problem through changing the weights of their connections in a process known as network training. NeuroXL Clusterizer is an add-on program for Microsoft Excel that uses an unsupervised neural network approach (Kohonen) for clustering data [10,11]. NeuroXL Clusterizer implements the Kohonen self-organizing neural network, which performs categorization by learning the trends and relationships within the data through an unsupervised learning process without input of targets by the user. The network consists of an input layer and an output layer. The input layer contains multiple nodes, one for each input parameters (mutation and survival duration). The output layer contains a two-dimensional grid representing the clusters. The nodes in the input layer are connected to the nodes in the output layer with individual strength of connection. Training data (mutation results, survival duration) are fed to from input layer to output layer. The output nodes with connection strength closely matches the input case data would "win" the case and have connection strength adjusted to match the case even more. With all training cases, output nodes with the most "wins" will be saved. The iteration continues until a targeted number of clusters is achieved.

NeuroXL Clusterizer hides the complexity of its advanced neural network-based methods while taking advantage of the analyst's existing knowledge of Microsoft Excel spreadsheets. The user simply supplies the input data and NeuroXL Clusterizer implements a neural network that segments the data according to the user's preferences for display of results. The steps in using NeuroXL Clusterizer in our study are as following: (a) Select NeuroXL Clusterizer from the menu in MS Excel. (b) Specify the range of input data (mutations and prognosis

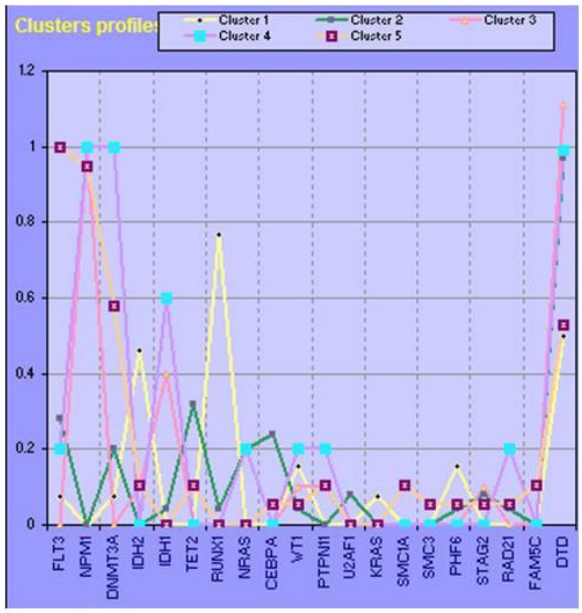


Figure 2. *Incidence average of mutations and prognosis. Scales on vertical axis: -For each mutation: 1 is equivalent to its presence in 100% of cases in a particular cluster; -Prognosis for each cluster: 1 is equivalent to 1,000 day survival after diagnosis.*

in term of survival duration) that one needs to cluster under "Inputs" **(Figure 1)**, and then specify the cell range where one would like to display the clustered data.

Set the number of clusters into which the data are grouped (5 in this study). Specify neural network parameters such as display graph, format, and statistics. In most cases, the default values are applicable. (c) Click the "Clusterize" button. The mutations are now grouped into five distinct categories. The average data values for these categories are displayed, as well as the weight of each cluster, which represents the percentage of items belonging to the cluster. A graphical representation of each cluster can also be created from the data above **(Figure 2)**.

Results

One hundred and twenty-one cases with positivity for the 23 most common mutations were obtained from the original set of 200 AML cases in the TCGA database. Subsequently, 72 cases with no cytogenetics abnormalities (CN-AML) were obtained from these 121 cases. Within the 72 CN-AML cases, the following mutations were not found: TP53, NRAS, KIT, EZH2, and HNRNPK, leaving 18 mutations in this subset of patients.

Table 1. *Associations between mutations and prognosis in initial clustering analysis (n=72).*

Cluster No	Number of cases (%)	Mutations (positive for)	Prognosis (see text for details)
1	13 (18.1)	RUNX1	poor
2	25 (34.7)	Other mutations	intermediate
3	10 (13.9)	NPM1	good
4	5 (6.9)	NPM1/ DNMT3A	intermediate
5	19 (26.4)	FLT3-ITD/NPM1	poor

Table 2. *Summary of important mutations in cluster 2 (group with other mutations).*

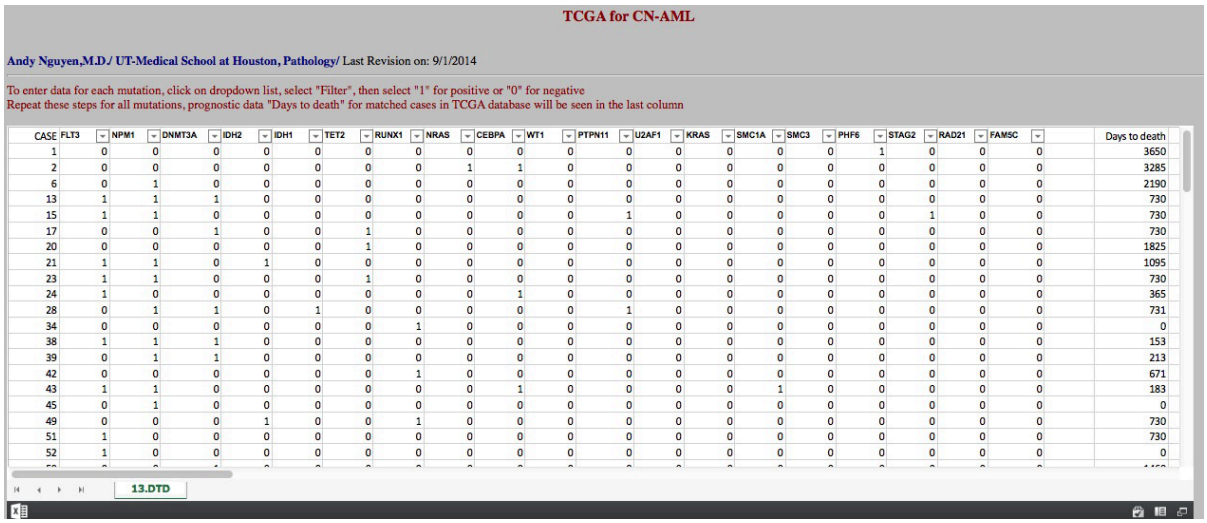
Mutation	Number of cases (%)	Day-to -death, mean	Prognosis (see text for details)
CEBPA	4 (5.5)	1,483	Good
CEBPA/FLT3-ITD	2 (2.8)	182	Poor
FLT3-ITD	4 (5.5)	274	Poor
TET2	4 (5.5)	1,658	Good

Clustering analysis using unsupervised neural network was performed on these 72 cases with 18 most common mutations to group them into 5 clusters according to their pattern of mutations and prognosis. A threshold was set so that a mutation is only listed for a particular cluster if it is present in more than 75% of the cases in that cluster. Using 75% threshold for mutation frequency, initial clustering was found to be based on only 4 mutations: NPM1, FLT3-ITD, RUNX, and DNMT3A. Other mutations do not have significant contribution to attributes of the 5 clusters. The results are shown in **Figure 2** and **Table 1**. The survival duration associated with each cluster is determined by clustering analysis (Poor <550 days, Intermediate = 550-1,000 days, Good >1,000 days).

Initial clustering analysis shows cluster 2 with mutations other than FLT3-ITD/NPM1, NPM1/ DNMT3A, NPM1, or RUNX1. Cluster 2 with lack of defining mutations needs to be analyzed further for subgrouping of cases. The following significant mutations are seen in at least 20% of the cases in this cluster: CEBPA, FLT3-ITD, and TET2. Further analysis of cases with these mutations reveals the following: (a) 4 cases with CEBPA are associated with good prognosis, (b) 2 cases with CEBPA/FLT3-ITD are associated with poor prognosis, (c) 4 cases with FLT3-ITD are associated with poor prognosis, and (d) 4 cases with TET2 are associated with good prognosis. The data for cases with these mutations are shown in **Table 2**.

Since TP53 and KIT mutations have been found to be clinically significant in previous studies [5-8,12], we further examined the cases with these mutations **(Table 3)**. All 7 cases with TP53 mutation also have complex chromosomal abnormalities, defined as at least 3 or more abnormalities; the survival data show poor prognosis for this group. Among 7 cases with KIT mutation, 5 have core binding factor (CBF) translocations with good prognosis and 2 have complex chromosomal abnormalities with intermediate prognosis. These 14 cases are not included in the clustering analysis for CN-AML cases since they have chromosomal abnormalities. Note that our findings confirmed the association found in existing literature between the following mutations and karyotypes: TP53 mutation and complex chromosomal abnormalities and KIT mutation and CBF translocations [13].

The data in the TCGA database are saved in a cryptic format known as Mutation Alignment Format (MAF) with a tab-delimited flat file structure. Multiple steps are needed to convert the data into readable form such as MS Excel files. For this reason, the data in this database are not easily accessible by general users. To facilitate easy public access to the prognostic data of CN-AML cases in the TCGA database, we compiled the mutation profiles and survival data for all 72 CN-AML cases and uploaded it to Microsoft OneDrive. Its associated html file on OneDrive is linked to our web page (http://HemePathReview.com) for public access.



TCGA for CN-AML

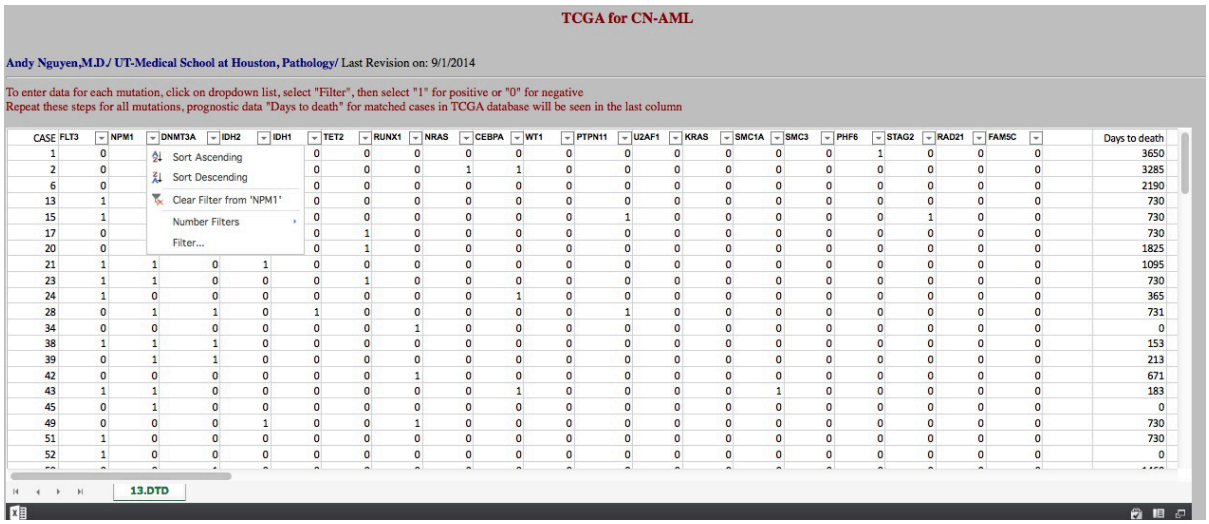
Andy Nguyen,M.D./ UT-Medical School at Houston, Pathology/ Last Revision on: 9/1/2014

To enter data for each mutation, click on dropdown list, select "Filter", then select "1" for positive or "0" for negative
Repeat these steps for all mutations, prognostic data "Days to death" for matched cases in TCGA database will be seen in the last column

CASE	FLT3	NPM1	DNMT3A	IDH2	IDH1	TET2	RUNX1	NRAS	CEBPA	WT1	PTPN11	U2AF1	KRAS	SMC1A	SMC3	PHF6	STAG2	RAD21	FAM5C	Days to death
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3650
2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	3285
6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2190
13	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	730
15	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	730
17	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	730
20	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1825
21	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1095
23	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	730
24	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	365
28	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	731
34	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
38	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	153
39	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	213
42	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	671
43	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	183
45	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	730
51	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	730
52	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

13.DTD

Figure 3. *Web page for CN-AML cases with mutations and survival duration.*



TCGA for CN-AML

Andy Nguyen,M.D./ UT-Medical School at Houston, Pathology/ Last Revision on: 9/1/2014

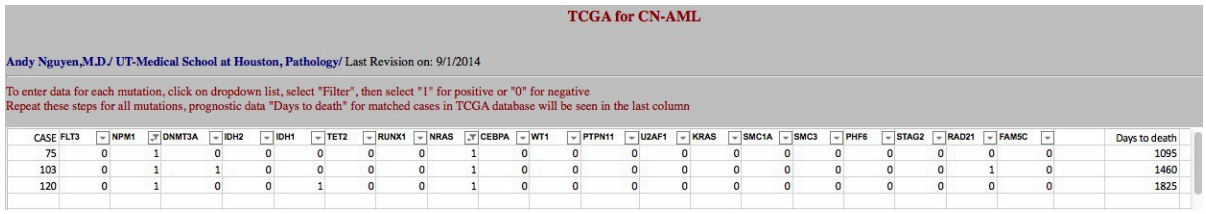
To enter data for each mutation, click on dropdown list, select "Filter", then select "1" for positive or "0" for negative
Repeat these steps for all mutations, prognostic data "Days to death" for matched cases in TCGA database will be seen in the last column

Sort Ascending
Sort Descending
Clear Filter from 'NPM1'
Number Filters
Filter...

CASE	FLT3	NPM1	DNMT3A	IDH2	IDH1	TET2	RUNX1	NRAS	CEBPA	WT1	PTPN11	U2AF1	KRAS	SMC1A	SMC3	PHF6	STAG2	RAD21	FAM5C	Days to death
1	0				0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3650
2	0				0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	3285
6	0				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2190
13	1				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	730
15	1				0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	730
17	0				0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	730
20	0				0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1825
21	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1095
23	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	730
24	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	365
28	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	731
34	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
38	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	153
39	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	213
42	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	671
43	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	183
45	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	730
51	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	730
52	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

13.DTD

Figure 4. *Selecting NPM1 positivity.*



TCGA for CN-AML

Andy Nguyen,M.D./ UT-Medical School at Houston, Pathology/ Last Revision on: 9/1/2014

To enter data for each mutation, click on dropdown list, select "Filter", then select "1" for positive or "0" for negative
Repeat these steps for all mutations, prognostic data "Days to death" for matched cases in TCGA database will be seen in the last column

CASE	FLT3	NPM1	DNMT3A	IDH2	IDH1	TET2	RUNX1	NRAS	CEBPA	WT1	PTPN11	U2AF1	KRAS	SMC1A	SMC3	PHF6	STAG2	RAD21	FAM5C	Days to death
75	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1095
103	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1460
120	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1825

Figure 5. *Subsequent selecting NRAS positivity and display of matching cases.*

Let us assume that we have a CN-AML patient with positivity for NPM1 and NRAS. No specific information is available for such combination in literature and we want to search the compiled data file for any previous patients with this mutation profile for survival duration. A typical session for such a patient is shown in **Figures 3** though **5**. The initial web page **(Figure 3)** shows 18 common mutations together with survival data for all 72 CN-AML cases. NPM1 positivity is selected from the drop-down list in **Figure 4**. NRAS is subsequently selected and the final results **(Figure 5)** show 3 cases with corresponding prognostic data.

Table 3. *Summary of further examination of cases with mutations not shown in CN-AML cases.*

Mutation	Associated cytogenetics	Number of cases	Day-to-death, mean	Prognosis (see text for details)
TP53	Complex*	7	174	Poor
KIT	CBF	5	1,199	Good
KIT	Complex*	2	669	Intermediate

Complex*: complex chromosomal abnormalities (3 or more abnormalities)

Table 4. *Final clustering analysis results and comparison of prognostic groups according to mutation status between ours to other previous studies.*

Prognosis	Our Study	Grossman [6]	Patel [7]	Shen [8]	Dohner [5]
Good	NPM1, CEBPA, TET2	NPM1	NPM1/IDH1, NPM1/IDH2	NPM1, CEBPA	NPM1, CEBPA
Intermediate	NPM1/DNMT3A, other mutations	CEBPA, FLT3-ITD, NPM1/FLT3-ITD, Other mutations	CEBPA, FLT3-ITD, CEBPA/FLT3-ITD, Other mutations	FLT3-ITD, Other mutations	NPM1/FLT3-ITD, FLT3-ITD
Poor	RUNX1, FLT3-ITD/NPM1, FLT3-ITD/CEBPA, FLT3-ITD	MLL-PTD, RUNX1, ASXL1, TP53	PHF6, ASXL1, MLL-PTD, TET2, DNMT3A, FLT3-ITD/TET2, FLT3-ITD/MLL-PTD, FLT3-ITD/ DNMT3A MLL-PTD, DNMT3A, NPM1/DNMT3A MLL-PTD	MLL-PTD, DNMT3A, NPM1/DNMT3A	MLL-PTD

Discussion

Combinations of mutations appear to dictate the clinical behavior of AML in terms of prognosis. The association between mutations and prognosis for clusters 1 through 5 in this study confirms some key findings from various sources of published data as expected, although not entirely reproducing results from each particular study. Results from our clustering analysis on 72 CN-AML cases in the TCGA database show the following prognostic groups: (a) good: NPM1, CEBPA, or TET2, (b) intermediate: NPM1/DNMT3A, or other mutations, (c) poor: RUNX1, FLT3-ITD, FLT3-ITD/NPM1, or FLT3-ITD/CEBPA.

Comparison of our results to previous major studies on AML mutations are shown in **Table 4** [5-8]. Our clustering analysis of 72 CN-AML cases with 23 most common mutations partially reproduced prognostic findings associated with various mutations from previous studies except for two findings: (a) NPM1/DNMT3A appears to be associated with intermediate prognosis in our study but is associated with poor prognosis in another study, (b) FLT3-ITD, FLT3-ITD/NPM1, and FLT3-ITD/CEBPA belong to poor prognostic group in our study whereas they belong to intermediate prognostic group in some other studies. Note that the results from this study are from de-novo AML cases whereas previous studies cover de novo AML and also AML transformed from myeloproliferative neoplasms or myelodysplastic syndrome. The results from this study therefore are more accurate for the selected subgroup of de novo AML. The discrepancy between our results and others is most likely due to difference in patient selection.

There is also apparent discrepancy between findings among various previous studies. This may be due to different ethnic groups and different sets of mutation tests, many of which are not performed in all studies. Since the effect of a given mutation may depend on the presence or absence of other mutations, results of some previous studies may be inaccurate due to incomplete sets of mutations.

Of the adult cancer types that have been extensively sequenced to date, AML has the fewest mutations discovered. The average number of coding mutations per patient is 13, of which only 5 are recurrently mutated in each genome [4]. Data have shown that some mutations that are common in AML (DNMT3A, NPM1, CEPBA, IDH1/2, and RUNX1) are mutually exclusive of the transcription-factor fusions, suggesting that these mutations may have functions in the initiation of CN-AML that are similar to the functions of fusion genes in AML with chromosomal abnormalities [4]. There is supporting evidence that more than one mutation is necessary to trigger AML [12]. This applies not only to cases where several mutations can be documented but also to those that may carry other yet undiscovered mutations. Ley et al [14] found that many samples had mutations in both NPM1 and DNMT3A or in both NPM1 and FLT3-ITD. From this observation, combined with the strong association between samples having concurrent mutations in NPM1, FLT3-ITD, and DNMT3A and distinct clusters in mRNA, miRNA, and DNA methylation, they suggest that samples with mutations in all these three genes represent a novel subtype of AML. If AML requires several genetic hits to develop, then different mutations must play specific roles in the steps of leukemogenesis.

One of the important take-home messages of cancer genome studies is that it is becoming less important to determine the subtypes of cancer cells based on histology or immunophenotypes and more important to treat the underlying driving mutations. The TCGA has been described as producing a "firehose of information" regarding mutations and epigenetic changes. The next phase of cancer research will be to determine the functional importance of these mutations so that we can design more effective targeted therapy that leads to better patient outcomes.

Even though a large number of mutations were tested in each case, the results of this study are limited by the relatively small number of CN-AML cases (n=72). Future studies based on a much larger number of CN-AML patients, and perhaps even more mutation testing, are expected to reveal more detailed stratification in the form of more clusters of cases with different prognosis. The characterization of CN-AML by the presence or absence of mutations in selected genes should allow a division of CN-AML cases into different biological and prognostic groups, as well as facilitating molecular monitoring of CN-AML.

References

1. WHO classification of tumours of haematopoietic and lymphoid tissues. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW, editors. Lyon: IARC; 2008.
2. Fernandez-Mercado M, Yip BH, Pellagatti A, Davies C, Larrayoz MJ, Kondo T, Perez C, Killick S, McDonald EJ, Odero MD and others. Mutation patterns of 16 genes in primary and secondary acute myeloid leukemia (AML) with normal cytogenetics. PLoS One 2012;7(8):e42334.
3. The Cancer Genoma Atlas (TCGA) Data Portal. <https://tcga-data.nci.nih.gov/tcga>.
4. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368(22):2059-74.
5. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA and others. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood 2010;115(3):453-74.
6. Grossmann V, Schnittger S, Kohlmann A, Eder C, Roller A, Dicker F, Schmid C, Wendtner CM, Staib P, Serve H and others. A novel hierarchical prognostic model of AML solely based on molecular mutations. Blood 2012;120(15):2963-72.
7. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, Van Vlierberghe P, Dolgalev I, Thomas S, Aminova O and others. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med 2012;366(12):1079-89.
8. Shen Y, Zhu YM, Fan X, Shi JY, Wang QR, Yan XJ, Gu ZH, Wang YY, Chen B, Jiang CL and others. Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. Blood 2011;118(20):5593-603.
9. Dombret H. Gene mutation and AML pathogenesis. Blood 2011;118(20):5366-7.
10. Kohonen T. Self-organized formation of topologically correct feature maps. Biological Cybernetics 1982;43(1):59-69.
11. Ultsch A, Siemon HP. Kohonen's self organizing feature maps for exploratory data analysis. Proceedings of the International Neural Network Conference. Paris1990. p 305-308.
12. Foran JM. New prognostic markers in acute myeloid leukemia: perspective from the clinic. Hematology Am Soc Hematol Educ Program 2010;2010:47-55.
13. Haferlach T. Molecular genetic pathways as therapeutic targets in acute myeloid leukemia. Hematology Am Soc Hematol Educ Program 2008:400-11.
14. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J and others. DNMT3A mutations in acute myeloid leukemia. N Engl J Med 2010;363(25):2424-33.