Data Descriptor

Dataset of Two-Dimensional Gel Electrophoresis Images of Acute Myeloid Leukemia Patients before and after Induction Therapy

Juan E. Urrea 1, Luisa F. Restrepo 2, Jeanette Prada-Arismendy 2, Erwing Castillo 3, Manuel M. Goez 4, Maria C. Torres-Madronero 4, Edilson Delgado-Trejos 1 and Sarah Röthlisberger 2.

Abstract: Acute myeloid leukemia (AML) is a malignant disorder of the hematopoietic stem and progenitor cells, which results in the build-up of immature blasts in the bone marrow and eventually in the peripheral blood of affected patients. Accurately assessing a patient’s prognosis is very important for clinical management of the disease, which is why there are several prognostic factors such as age, performance status at diagnosis, platelet count, serum creatinine and albumin that are taken into account by the clinician when deciding the course of treatment. However, proteomic changes related to treatment response in this patient group have not been widely explored. Here, we make available a set of 22 two-dimensional gel electrophoresis (2DGE) images obtained from the peripheral blood samples of 11 patients with AML, taken at the time of diagnosis and after induction therapy (approximately 21–28 days after starting treatment). The same set of 2DGE images is also made available after a preprocessing stage (an additional 22 2DGE pre-processed images), which was performed using algorithms developed in Python, in order to improve the visualization of characteristic spots and facilitate proteomic analysis of this type of images.

Dataset: The dataset will be published as a supplement to this paper, so this field will be filled by the editors of the journal.

Dataset License: CC-BY 4.0

Keywords: acute myeloid leukemia; image preprocessing; proteomics; two-dimensional gel electrophoresis

1. Summary

According to the Global Cancer Observatory (Globocan 2018), each year, 437,033 patients worldwide are diagnosed with some type of leukemia, and 309,006 people die from this disease. Acute myeloid leukemia (AML) is a type of leukemia that mainly occurs in older adults; 42% of Americans diagnosed with AML are over 65 years of age, and their diagnosis is rarely made before 40 years of age, although cases have progressively increased over time [1]. AML is the result of an accumulation of acquired genetic alterations in the...
DNA of hematopoietic progenitor cells, and accurately assessing a patient’s prognosis is very important for clinical management of the disease. The patient’s cytogenetic profile is currently the strongest prognostic factor. For example, a complex karyotype, monosomy 5 or 7, t(6;9), inv(3), or 11q changes, other than t(9;11), have all been associated with a significantly lower response to treatment and overall survival [2]. It is clear that genetic studies are very valuable; however, isolated from a context in which thousands of proteins mediate cellular function, this prognostic model is not complete.

The images of this dataset were obtained by two-dimensional gel electrophoresis (2DGE), a technique that separates proteins according to their isoelectric point and molecular weight [3], followed by protein staining and image capture. Often, 2DGE images include anomalies [4,5] such as vertical lines, horizontal lines, diffuse points, and noise, among others, which make it difficult to identify spots that contain valuable information. Therefore, a preprocessing stage is often necessary in order to discriminate stains and noise from real protein spots [6]. Omitting this stage can affect the interpretation of the data, as noise could be identified as false protein spots [7]. Image preprocessing is responsible for reducing or correcting these irregularities in 2DGE images. The authors have implemented an approach that integrates the techniques of image normalization, noise reduction by non-linear techniques, and background correction [4,8], sequentially applying the following structure: adaptive piecewise histogram equalization for image normalization, a geometric nonlinear diffusion filter (GNDF) for filtering, and multilevel thresholding for background correction, obtaining favorable results [9].

2. Data Description

The database consists of a set of 22 2DGE images obtained from the peripheral blood samples of 11 patients with acute myeloid leukemia. Of these, 11 images correspond to samples taken at the time of diagnosis, and the other 11 correspond to samples taken from the same patients after induction therapy (approximately 21–28 days after starting treatment). Images named with the suffix BEFORE refer to 2DGE images of samples taken at the time of diagnosis (before treatment), while images named with the suffix AFTER correspond to 2DGE images of samples taken after treatment. These 22 images are also made available with the preprocessing stage applied, to which the prefix PREPROC has been applied. Each image in the database is in tagged image file format (TIFF) format with a resolution of 300 dots per inch (DPI). In total, the database, which can be found in the Supplementary Materials, contains 44 images (22 raw 2DGE images and 22 pre-processed 2DGE images). The characteristics corresponding to each image are summarized in Table 1.

| 2DGE Image | 2DGE Image Size | 2DGE Pre-Processed Image | 2DGE Pre-Processed Image Size | Width (Pixels) | Height (Pixels) |
|------------|-----------------|--------------------------|-------------------------------|----------------|-----------------|
| HMUA02_BEF | 564 KB          | PREPROC_HMUA02_BEF       | 1.63 MB                       | 965            | 757             |
| HMUA02_AFT | 530 KB          | PREPROC_HMUA02_AFT       | 1.54 MB                       | 1006           | 803             |
| HMUA03_BEF | 607 KB          | PREPROC_HMUA03_BEF       | 1.94 MB                       | 972            | 820             |
| HMUA03_AFT | 554 KB          | PREPROC_HMUA03_AFT       | 1.73 MB                       | 974            | 798             |
| HMUA04_BEF | 589 KB          | PREPROC_HMUA04_BEF       | 1.65 MB                       | 1021           | 842             |
| HMUA04_AFT | 607 KB          | PREPROC_HMUA04_AFT       | 1.88 MB                       | 1035           | 866             |
| HMUA05_BEF | 556 KB          | PREPROC_HMUA05_BEF       | 1.90 MB                       | 1018           | 810             |
| HMUA05_AFT | 558 KB          | PREPROC_HMUA05_AFT       | 1.65 MB                       | 1012           | 788             |
| HMUA010_BEF| 584 KB          | PREPROC_HMUA010_BEF      | 2.21 MB                       | 1021           | 838             |
| HMUA010_AFT| 585 KB          | PREPROC_HMUA010_AFT      | 2.01 MB                       | 1012           | 828             |
| HMUA011_BEF| 527 KB          | PREPROC_HMUA011_BEF      | 1.70 MB                       | 985            | 777             |
| HMUA011_AFT| 506 KB          | PREPROC_HMUA011_AFT      | 1.78 MB                       | 974            | 781             |
| HMUA012_BEF| 603 KB          | PREPROC_HMUA012_BEF      | 2.14 MB                       | 971            | 798             |
| HMUA012_AFT| 430 KB          | PREPROC_HMUA012_AFT      | 1.75 MB                       | 969            | 775             |
| HMUA013_BEF| 498 KB          | PREPROC_HMUA013_BEF      | 1.73 MB                       | 1024           | 803             |
Table 1. Cont.

| 2DGE Image | 2DGE Image Size | 2DGE Pre-Processed Image | 2DGE Pre-Processed Image Size | Width (Pixels) | Height (Pixels) |
|------------|-----------------|--------------------------|-------------------------------|---------------|-----------------|
| HMUA013_AFTER | 561 KB | PREPROC_HMUA013_AFTER | 1.85 MB | 1012 | 820 |
| HMUA015_Before | 552 KB | PREPROC_HMUA015_Before | 1.67 MB | 1054 | 857 |
| HMUA015_After | 576 KB | PREPROC_HMUA015_After | 1.88 MB | 1046 | 870 |
| HMUA017_Before | 501 KB | PREPROC_HMUA017_Before | 1.91 MB | 983 | 757 |
| HMUA017_After | 476 KB | PREPROC_HMUA017_After | 1.84 MB | 1102 | 921 |
| HMUA018_Before | 432 KB | PREPROC_HMUA018_Before | 1.38 MB | 1111 | 858 |
| HMUA018_After | 526 KB | PREPROC_HMUA018_After | 1.79 MB | 1036 | 795 |

3. Methods

3.1. Patients

Peripheral blood was obtained from 11 newly diagnosed patients with de novo AML at Hospital Manuel Uribe Angel in Colombia. Two blood samples were taken from each patient: at the time of diagnosis (before the start of chemotherapy) and once again after completion of the first round of induction therapy, which was typically 2–3 weeks after induction or when neutrophil and platelet recovery was achieved. Relevant clinical information of the patients involved in this study is summarized in Table 2.

Table 2. Clinical Information.

| Patient | Age | Sex | Karyotype | AML Subtype | Blasts (i) | Induction Protocol | Blasts (f) | Response to Induction |
|---------|-----|-----|-----------|-------------|-----------|-------------------|-----------|----------------------|
| HMUA_02 | 35  | M   | Normal    | M3          | 86%       | PETHEMA           | 1%        | CR 4                 |
| HMUA_03 | 56  | M   | Normal    | M4          | 65%       | 7 + 3             | 4.5%      | CR                   |
| HMUA_04 | 67  | M   | Normal    | M4          | 14%       | FLUGA             | 35%       | Resistant            |
| HMUA_05 | 42  | F   | T(8;21)   | M3          | 85%       | PETHEMA           | 1.2%      | CR                   |
| HMUA_10 | 18  | F   | Missing   | M1/M2       | 74%       | 7 + 3             | 0.5%      | CR                   |
| HMUA_11 | 65  | F   | CCR       | M4          | 42%       | 7 + 3             | 11%       | PR 6                 |
| HMUA_12 | 53  | F   | T(8;21)   | M2          | 20%       | 7 + 3             | 1.3%      | CR                   |
| HMUA_13 | 46  | M   | Normal    | M2          | 68%       | 7 + 3             | 35.4%     | Resistant 7           |
| HMUA_15 | 50  | M   | CCR       | M1/M2       | 47%       | 7 + 3             | 8.4%      | PR                   |
| HMUA_17 | 67  | M   | T(8;21)   | M4          | 66%       | 7 + 3             | 0.27%     | CR                   |
| HMUA_18 | 18  | M   | Normal    | M1/M2       | 28%       | 7 + 3             | 0.6%      | CR                   |

1 According to the French–American–British (FAB) classification system. 2 (i): initial blast count, before induction therapy. 3 (f): final blast count, after induction therapy. 4 CR: complete remission, defined as <5% blasts in bone marrow after induction therapy. 5 CCR: complex chromosome rearrangement. 6 PR: partial response, defined as 5–20% blasts in bone marrow after induction therapy. 7 Resistance to therapy, defined by >20% blasts in bone marrow after induction therapy.

3.2. Protein Extraction

Peripheral blood mononuclear cells (PBMCs) were isolated from the blood samples by standard density gradient centrifugation with a Ficoll Histopaque®-1077 (Sigma-Aldrich, St. Louis, MO, USA). In order to extract proteins from the PBMCs, the cells were lysed (0.5% Triton x-100, 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), protease inhibitors) and the proteins precipitated in a 20% (v/v) final concentration of trichloroacetic acid. The protein pellet was resuspended in a rehydration buffer (7 M urea, 2% 3-cholamidopropyl dimethylammonio 1-propanesulfonate (CHAPS), 0.5% carrier ampholytes) and stored at −70 °C.

3.3. Two-Dimensional Gel Electrophoresis

Proteins (50 µg) were loaded by passive rehydration onto 7 cm ZOOM® immobilized pH gradient (IPG) strips with a pH of 3–10 NL (ThermoFisher Scientific, Waltham, MA, USA) at room temperature. Isoelectric focusing was carried out using the following voltage ramp: 100 V for 1 h, 150 V for 1 h, 200 V for 5 min, 450 V for 5 min, 600 V for 5 min, 750 V for 5 min, 950 V for 5 min, 1200 V for 10 min, 1400 V for 10 min, 1600 V for 10 min,
2000 V for 45 min. The IPG strips were then reduced with 100 mM Dithiothreitol (DTT) and alkylated with 2.5% iodoacetamide, according to the manufacturer’s recommended protocol. After this, the IPG strips were loaded onto SDS-PAGE NuPAGE™ Novex™ 4–12% Bis-Tris protein gels 1.5 mm in size (ThermoFisher Scientific) and run at 200 V for 45 min. After electrophoresis, these were stained with SYPRO® Ruby (Invitrogen™, ThermoFisher Scientific), and the gel images were acquired using the ChemiDoc™ MP System (Biorad).

3.4. Image Pre-Processing

This step was performed in order to mitigate anomalies due to the acquisition routines and improve spot detection. The approach proposed in [9] was applied, integrating the following techniques for image normalization, noise reduction, and background correction: adaptive piecewise histogram equalization, a geometric nonlinear diffusion filter (GNDF), and multilevel thresholding. The algorithm was executed in Python, which is an open-source programming language, with free access to permanent online support through a considerable number of available libraries, accelerating the creation of multi-stage structure codes with the aim of obtaining consistent, reliable, and potentially integrable results.

Supplementary Materials: The following are available online at https://www.mdpi.com/2306-5729/6/2/20/s1.

Author Contributions: Conceptualization, S.R., J.P.-A., M.C.T.-M. and E.D.-T.; methodology, S.R., M.C.T.-M. and E.D.-T.; software, J.E.U., M.M.G., M.C.T.-M. and E.D.-T.; validation, S.R., M.C.T.-M. and E.D.-T.; formal analysis, M.M.G., M.C.T.-M. and E.D.-T.; investigation, L.F.R.; resources, E.C.; data curation, S.R.; writing—original draft preparation, J.E.U., S.R. and E.D.-T.; writing—review and editing, J.E.U., S.R., J.P.-A., M.C.T.-M. and E.D.-T.; visualization, J.E.U., S.R. and E.D.-T.; supervision, S.R., M.C.T.-M. and E.D.-T.; project administration, S.R.; funding acquisition, S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Instituto Tecnologico Metropolitano ITM, grant number P17215. J.E.U. was recipient of the Jovenes Investigadores e Innovadores ITM 2020 program of the Instituto Tecnologico Metropolitano ITM of Medellin-Colombia.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki, and approved both by the Research Ethics Committee of the INSTITUTO TECNOLOGICO METROPOLITANO (6 June 2014, project code P17215) and the Ethics Committee of the HOSPITAL MANUEL URIBE ANGEL (17 April 2015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available in the Supplementary Material.

Acknowledgments: The authors would like to thank the E.S.E. Hospital Manuel Uribe Angel, the Biomedical Sciences Laboratory, the Smart Machine and Pattern Recognition Laboratory (MIRP Lab), and the Measurement Analysis and Decision Support Laboratory (AMYSOD Lab) of Parque i, Medellin, Colombia.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References
1. Deschler, B.; Lubbert, M. Acute myeloid leukemia: Epidemiology and etiology. *Cancer* 2006, 107, 2099–2107. [CrossRef] [PubMed]
2. Estey, E.H. Acute myeloid leukemia: 2019 update on risk-stratification and management. *Am. J. Hematol.* 2018, 93, 1267–1291. [CrossRef] [PubMed]
3. Beranova-Giorgianni, S. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: Strengths and limitations. *Trac. Trends Anal. Chem.* 2003, 22, 273–281. [CrossRef]
4. Kaczmarek, K.; Walczak, B.; De Jong, S.; Vandeginste, B.G. Preprocessing of two-dimensional gel electrophoresis images. *Proteomics* 2004, 4, 2377–2389. [CrossRef] [PubMed]
5. Nugues, P.M. Two-dimensional electrophoresis image interpretation. *IEEE Trans. Biomed. Eng.* 1993, 40, 760–770. [CrossRef] [PubMed]
6. Villegas-Rivera, G.A.; Torres-Madronero, M.C.; Röthlisberger Booth, S.; Delgado-Trejos, E. Procesamiento de imágenes de electroforesis bidimensional: Una revisión. Sci. Tech. 2019, 24, 76–84. [CrossRef]
7. Tsakanikas, P.; Manolakos, E.S. Improving 2-DE gel image denoising using contourlets. Proteomics 2009, 9, 3877–3888. [CrossRef] [PubMed]
8. Goez, M.M.; Torres-Madronero, M.C.; Rothlisberger, S.; Delgado-Trejos, E. Preprocessing of 2-Dimensional Gel Electrophoresis Images Applied to Proteomic Analysis: A Review. Genom. Proteom. Bioinform. 2018, 16, 63–72. [CrossRef] [PubMed]
9. Goez, M.M.; Torres-Madronero, M.C.; Rothlisberger, S.; Delgado-Trejos, E. Joint pre-processing framework for two-dimensional gel electrophoresis images based on nonlinear filtering, background correction and normalization techniques. BMC Bioinform. 2020, 21, 376. [CrossRef] [PubMed]